Ab Initio Insights into the Electronic Structure of Small Molecular Systems: Characterization of Chemical Trends *via* Coupled Cluster Methods

by

Mark Edward Wolf

(under the direction of Henry F. Schaefer III)

ABSTRACT

One of the advantages of computational quantum chemistry is its ability to leverage theoretical models to extract insights from the study of challenging chemical systems. In particular, the trends that manifest along groups or rows of the periodic table have been of interest to chemists for many years. Comprehensive understanding of these kinds of trends has the potential to evolve our chemical intuition of many systems that are positioned at the cutting edge of experimental research. This work begins with a summary of the foundational theoretical details of the *ab initio* methods employed throughout this body of research. The first application examines atmospherically relevant binary complexes formed between hypohalous acids and water, with close examination of the competition between hydrogen and halogen bonding. Next, motivated by the impressive synthetic work of Cummins, we characterize the $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) tetrahedrane molecules and lay a firm foundation for future experimental work on these systems. Significant attention is given to the electronic structure motifs exhibited by these tetrahedrane species and how they change with increasing pnictogen size. The third project examines the $HNCO + H_2O$ reactions which are of significant importance for many industrial applications. Key stationary points are characterized with energy predictions approaching the CCSDT(Q)/CBS limit. Using *ab initio* composite methods, the changes

in barrier heights for a collection of 24 substituted RNCO species as well as the catalytic effects of additional water and RNCO molecules are determined. The resulting patterns are analyzed with the aim of informing novel material development. The final project concerns the design of an upper-level physical chemistry lab that focuses on the balance between computational cost and accuracy by means of an *ab initio* study of methane combustion energy. The components of the exercise are explained in detail with an emphasis on its intentional design which prioritizes pedagogical flexibility, accessibility, and learning goals that can be generalized to many computational science fields.

INDEX WORDS: coupled cluster theory, *ab initio*, periodic trends, tetrahedrane, isocyanate, hypohalous acid, focal point analysis, halogen bonding

Ab Initio Insights into the Electronic Structure of Small Molecular Systems: Characterization of Chemical Trends *via* Coupled Cluster Methods

by

Mark Edward Wolf

B.S. Calvin College, 2018

B.A. Calvin College, 2018

A Dissertation Submitted to the Graduate Faculty
of the University of Georgia in Partial Fulfillment
of the Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2021

Ab Initio Insights into the Electronic Structure of Small Molecular Systems: Characterization of Chemical Trends *via* Coupled Cluster Methods

by

Mark Edward Wolf

Major Professors: Henry F. Schaefer III

Committee: Gary E. Douberly

Steven E. Wheeler

Electronic version approved:

Ron Walcott

Vice Provost for Graduate Education and

Dean of the Graduate School

University of Georgia

December 2021

ACKNOWLEDGEMENTS

This work was only made possible by the innumerable people who have graciously loved and supported me the past few years. The following is a non-exhaustive list of those to whom I will be forever grateful and deserve far more gratitude than their mention here. To my many excellent Calvin professors who cultivated my love for science, challenged me as an academic, and mentored me on a personal dimension. To the many dear friends I have made at the CCQC. I have cherished the laughter, the sharing of burdens, and the professional collaboration that is ever present at the CCQC. To Boyi who spent hours guiding me on my first project where I would otherwise have been lost. To Dr. Turney who has patiently answered so many questions critical to research and ensured that I always had the resources I needed to succeed. To my committee members (Professor Wheeler and Professor Douberly) who have endowed me with so much knowledge, pushed me to be a better scientist, and carved out significant time to ensure my progression through graduate school. To Professor Schaefer who sought after, molded, supported, equipped, mentored, cared for, and befriended a Michigander who would never be here without your pursuit. To the many godly men and women (Daniel, Joseph, Arron & Deb, Michael & Sarah, Alex, Flynt & Madison, Keith & Amanda, Mitchell & Amber, Preston & Abbey, Jared, Nathan & Nicole, David & Elizabeth, Boyce & Grace, Josh & Lesleigh, Sky & Karen, Brian P., Bud, Brandon, Brian A., et al.) who I have shared many tears and much laughter with over these past few years. Each and every one listed has been an incredible blessing to me, especially in some of my most challenging moments. To the White family who has grafted me in as one of their own and gifted me with the joy of family while away from "home". I am especially thankful for their grace and patience as they most bore my weaknesses and selfishness in the most stressful moments. To my mom, dad, and Laura who have so selflessly supported and loved me my whole life. Without your support and provision in every aspect of life, my trajectory would look drastically different. I am forever grateful for the sacrifices you have made on my behalf and love you all dearly. To my beautiful bride Morgan: The "real" reason I was supposed to be in Georgia. You are my best friend and I cannot type words to express how much you mean to me and how much I love and need you. I would not have finished this program if it were not for you and I am excited to see where we go together from here. To my Lord and Savior Jesus Christ, by whom all things were made and through whom all things are held together. For such a wretched soul as me to be given the gifts, opportunities, and everlasting purpose that I have is beyond my comprehension.

List of Figures

List of Tables

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 *Ab initio* Electronic Structure Theory

Computational quantum chemistry is an increasingly relevant sub-field of chemistry that leverages computational resources to make predictions concerning the physical features of molecules and how they react with one another. This work focuses on *ab initio* electronic structure theory, a specific subset of quantum chemistry that relies upon methods built up from the fundamental mathematics of quantum physics to describe the behavior of electrons in a chemical system. The benefit of this approach is that the methods utilized have mathematically defined approximations (and therefore sources of error) and can, in most cases, be systematically improved to produce sufficiently reliable results, given computational constraints. Unfortunately, the large computational scaling of many of these approximations severely constrain the size of the chemical systems that can be studied. Despite this limitation, high-level *ab initio* methods can provide useful predictions of the physical features of small molecules such as equilibrium geometries, vibrational frequencies (IR), rotational constants (microwave), excited states (UV-VIS), ionization energies, dipole moments, bond orders, etc. These quantitative results, in many cases, can aid experimental chemists in the detection and classification of novel molecular species. Moreover, these predictions can also greatly aid in the qualitative explanation of observed and theorized chemical phenomena. Perhaps the most fundamental application of high-level *ab initio* quantum chemistry is the prediction of reliable energetics and thermodynamics of reaction pathways. These results can help predict novel and experimentally plausible chemical pathways, characterize reaction intermediates which are challenging to observe, identify the most favorable pathway in a chemical reaction, and provide the fundamental quantities (such as reaction barriers within "chemical accuracy" (1 kcal mol^{-1})) necessary to produce state-of-the-art kinetics models. All of these results, if correctly obtained, have the potential to motivate, explain, guide, and occasionally correct the ongoing experimental research of chemists across many different fields. This work will survey the theoretical foundation for these *ab initio* methods, describe some useful tools based on or used in conjunction with these methods, and detail

four research studies where *ab inito* techniques are applied to answer questions spanning multiple chemical sub-disciplines. Many of the theoretical details concerning how these methods are developed and applied to chemical systems are drawn from the excellent reference works of Szabo[1], Cremer[2], Pauling[3], and Levine[4].

1.2 The Schrödinger Equation

Electronic structure theory rests upon the landmark work of Schrödinger in the early 1920s that a many bodied system can be described via a "wave-like" probability distribution known as a wavefunction (Ψ)[5]. This wavefunction is a mathematical representation of n-particles a space at time t and possessing an infinite set of eigenvalues corresponding to any physical observables. Schrödinger discovered the differential equation denoted in Equation 1.1 which can be solved to determine the time independent energy of a molecular system. Given that $\hat{H}$ is the nonrelativistic molecular Hamiltonian (Equation 1.2), the E_i corresponds to eigenstates of the system and thus the quantized energy solutions for a molecular system, the ground state (E_0) generally being the most important. $\hat{H}$ is composed of terms that correspond to classically meaningful electronic interactions and energetic components in a molecular system: the kinetic energy of the nuclei with mass M_N ($-\sum_N \frac{1}{2M_N}\hat{\nabla}_N^2$), the kinetic energy of the electrons ($-\sum_i \frac{1}{2}\hat{\nabla}_i^2$), the nuclear-nuclear repulsion ($\sum_{N>M} \frac{Z_N Z_M}{R_{NM}}$), the electron-nuclear attraction ($-\sum_{Ni} \frac{Z_N}{r_{Ni}}$), and the electron-electron repulsion ($\sum_{i>j} \frac{1}{r_{ij}}$), where N and M are the nuclear indices, and i and j are the electron indices, $\hat{\nabla}$ is the gradient operator, Z is the nuclear charge, R_{MN} the internuclear distance, and r_{ij} the interelectron distance, all in atomic units. The $\hat{\nabla}$ operator is defined in equation 1.3 and consists of partial derivatives with respect to each Cartesian coordinate.

$$\hat{H}\,|\Psi_i\rangle = E_i\,|\Psi_i\rangle \tag{1.1}$$

$$\hat{H} = -\sum_N \frac{1}{2M_N}\hat{\nabla}_N^2 - \sum_i \frac{1}{2}\hat{\nabla}_i^2 + \sum_{N>M} \frac{Z_N Z_M}{R_{NM}} - \sum_{Ni} \frac{Z_N}{r_{Ni}} + \sum_{i>j} \frac{1}{r_{ij}} \tag{1.2}$$

$$\hat{\nabla} = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \tag{1.3}$$

Due to the many coupled interacting terms described in the molecular Hamiltonian, analytic solutions to the Schrödinger equation can only be produced in the simplest of cases such as the hydrogen atom. In practice, approximate solutions are obtained by making careful approximations that introduce minimal or at least well-defined and systematically removable error into the solutions. One of the most foundational and

pragmatic approximations is the Born–Oppenheimer approximation which leverages the drastically different time scales of motion between electrons and the nuclei (a consequence of the fact that a single proton has a mass that is approximately 1836 times greater than an electron and therefore moves much slower). The Born–Oppenheimer approximation assumes that electrons rearrange quickly relative to nuclei and, as a consequence, the nuclei can be considered stationary. This effectively factors out the vibrational and rotational components of the wavefunction and produces adiabatic solutions of the electronic Schrödinger equation which are a function of the nuclear coordinates in three dimensional space. The result of this approximation is a significantly simplified molecular Hamiltonian (Equation 1.4) where the kinetic energy of the nuclei is now zero and the nuclear-nuclear repulsion term is a constant ($\mathrm{V_{NN}} = \sum_{N>M} \frac{Z_N Z_M}{R_{NM}}$).

$$\hat{H} = -\sum_i \frac{1}{2}\hat{\nabla}_i^2 - \sum_{Ni} \frac{Z_N}{r_{Ni}} + \sum_{i>j} \frac{1}{r_{ij}} + V_{\mathrm{NN}} \tag{1.4}$$

The Born–Oppenheimer approximation also has significant practical value for interpreting chemical predictions. Assuming a reliable method is available to estimate solutions of the Schrödinger equation, which will be discussed shortly, we are able to construct a hypersurface known as a potential energy surface or PES. The PES is the mapping of nuclear coordinates, typically to 0 K gas-phase electronic energies. The conceptual nature of a PES has significant benefits for interpreting theoretical *ab initio* results. For example, a point that is a minimum on the PES, with respect to all degrees of freedom, corresponds to a molecular arrangement that is an equilibrium geometry and would be a candidate for spectroscopic observation. Minimum points or transition states (first order saddle points that possess one imaginary vibrational frequency that corresponds to a pathway connecting two minima) are referred to as stationary points and are the most important features of a PES. Therefore, this work primarily focuses on the characterization of key stationary points for key chemical processes. The curvature of the surface near an equilibrium geometry is also important and associates with a molecule's vibrational frequencies which can be used to predict experimental observables. Assuming that the curvature of the PES near these points is quadratic, Harmonic vibrational frequencies may be obtained to both predict vibrational frequencies and verify the nature of each stationary point as a minimum or a transition state. This assumption is generally valid, but can break down in specific systems where molecular vibration is severely anharmonic.

The PES is also useful for determining the reactivity of chemical species. Characterizing features of the PES helps predict the sort of mechanism a chemical reaction will proceed through, including quantification of the the transition state barriers. Reliable predictions of these results are incredibly important for understanding and informing experimental chemical research and can often be useful for determining what modifications might catalyze a reaction. There are certainly situations where the Born-Oppenheimer

approximation itself breaks down, but these cases are not present in this research. Much more relevant to this work is that fact that the PES can be highly dependent on the quality of the methodology used to solve the Schrödinger equation. For example, a molecular isomer may present as a minimum on a low-fidelity PES, yet have no corresponding stationary point on a high-fidelity PES, and is therefore not be a valid experimental target. This means that great care must be taken so that predicted results are both qualitatively and quantitatively correct. The following sections will outline a hierarchy of *ab initio* methods and some accompanying tools that are applied to characterize important stationary points with a high degree of quantitative accuracy and qualitative insight.

1.3 Hartree–Fock Theory

Given the fact that analytic solutions to the Schrödinger equation for multi-electron molecular systems are impossible to obtain, it is necessary to turn to approximate methods that can still produce physically meaningful results. The current foundation for most *ab initio* methods is known as Hartree-Fock (HF) theory. HF theory has proven to be a highly effective method and, though it has many severe limitations, is a necessary first computation intrinsic to the best methods currently in existence. The overarching idea of HF is that it is a procedure to iteratively solve for the best (i.e. lowest energy) set of molecular orbitals (MOs) given the constraint that the MOs are orthonormal to one another. According to the variational theorem, any trial wavefunction will produce an energy that is an upper bound for the true value, thus lower energy predictions are "better" and closer to the exact answer. These MOs are the end goal and provide incredibly valuable chemical insight into a system while also being a necessary precursor for many high-level *ab initio* methods. The following will elaborate on how this is accomplished and provide more technical insight into what a mathematical set of MOs actually is.

The foundation of the HF method is the construction of an approximate wavefunction from a set of indistinguishable one electron wavefunctions (or unoptimized MOs) $\phi_i^j = \phi_i(x_j, y_j, z_j)$ of which the ith orbital depends on the spatial coordinates of the jth electron in a molecular system. It would be nice if we could simply write the total wavefunction as a product of these MOs and proceed to easily solve the Schrödinger equation from there. Unfortunately, these MOs lack the inclusion of important physical properties such as electron spin and the antisymmetry which fermions such as electrons possess. To account for spin and avoid violation of the Pauli principle, we can define a new orbital known as spin orbital $\psi_i^j = \phi_i(x_j, y_j, z_j)\omega_i(s_j)$, where s_j corresponds to either a spin-up ($\omega_i(s_j) = \alpha$) or spin-down ($\omega_i(s_j) = \beta$) state for electron j. These α and β functions simply account for the orthonormality of spin integrals. That is to say the spin integral over a product of two spin functions is one if they are both α or both β and zero otherwise.

We could now write an approximate wavefunction (Φ) that is a product of one electron spin orbitals, but this still fails to properly incorporate antisymmetry into the system. Antisymmetry requires that if the indices of any two electrons are swapped, only the sign of the total wavefunction changes. Antisymmetry can be introduced by leveraging properties of determinants and writing the total wavefunction as a determinant where entries consist of every combination of N electrons in N spin orbitals, represented in Equation 1.5 and known as a Slater Determinant[6]. Each of the N rows correspond to an electron and each of the N columns corresponds to a spin orbital, with N being the number of electrons in the system and $\mathbf{x_i}$ being the spatial and spin coordinates of electron i. It is well-known that swapping a row or column of a determinant will change its sign, thus our wavefunction is now antisymmetric. The $\frac{1}{\sqrt{N!}}$ term is the normalization factor which ensures our total wavefunction remains normalized.

$$\Phi(\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_N}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1^1 & \psi_2^1 & \cdots & \psi_N^1 \\ \psi_1^2 & \psi_2^2 & \cdots & \psi_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1^N & \psi_2^N & \cdots & \psi_N^N \end{vmatrix} \tag{1.5}$$

Using the fact that we can select any set of orthonormal spin orbitals to define the wavefunction (or select orbitals and orthogonalize them), solutions to Equation 1.1 can be obtained for a molecular system and produce an energy result which is presented in Equation 1.6, where $\hat{h}$ and $\hat{g}$ are the one and two electron operators representing pieces of the Hamiltonian, respectively. This energy expression is known as the first Slater-Condon rule and the matrix elements of these operators, known as the one and two electron integrals, are defined by Equation 1.7 and 1.8.

$$E = \langle \Phi | \hat{H} | \Phi \rangle = \sum_i \left\langle \psi_i^i \right| \hat{h}_i \left| \psi_i^i \right\rangle + \sum_{i<j} \left\langle \psi_i^i \psi_j^j \right| \hat{g}(i,j) \left| \psi_i^i \psi_j^j \right\rangle - \sum_{i<j} \left\langle \psi_i^i \psi_j^j \right| \hat{g}(i,j) \left| \psi_j^i \psi_i^j \right\rangle \tag{1.6}$$

$$\langle \psi_p | \hat{h}_i | \psi_q \rangle = \int \psi_p^*(\mathbf{x_i}) \hat{h}_i(\mathbf{x_i}) \psi_q(\mathbf{x_i}) d\mathbf{x_i} \tag{1.7}$$

$$\langle \psi_p \psi_q | \hat{g_{ij}} | \psi_r \psi_s \rangle = \int \psi_p^*(\mathbf{x_i}) \psi_q^*(\mathbf{x_j}) \hat{g}(\mathbf{x_i}, \mathbf{x_j}) \psi_r(\mathbf{x_i}) \psi_s(\mathbf{x_j}) d\mathbf{x_i} d\mathbf{x_j} \tag{1.8}$$

Though we can use Equation 1.6 to compute the energy of any Slater determinant, only the best Slater determinant (i.e. set of MOs), in a variational sense, provides the best approximation of the true wavefunction. The method of Lagrangian multipliers can accomplish this by minimizing the energy with the constrain of orbital orthonormality: $\langle \psi_i | \psi_j \rangle = \delta_{ij}$. Proceeding forward, we consider the Restricted Hartree–Fock (RHF)

case as all species in this study are closed shell. In the RHF case we can integrate out the spin and continue working with a set of doubly occupied spatial orbitals (ϕ).

The equations that satisfy the Lagrangian relationship can be cast into the canonical Hartree-Fock equations (using unitary transformations) which result in a pseudo-eigenvalue problem defined in Equation 1.9, where $\hat{f}$ (the Fock operator) is an effective Hamiltonian composed of one electron operators. The $\hat{J}$ (Coulomb) and $\hat{K}$ (exchange) operators are defined in Equations 1.10 and 1.11, respectively, and introduce what is known as the mean field approximation where the electron-electron repulsion term is approximated by each electron's interaction with the average field of all other electrons rather than its instantaneous interaction with each individually. The set of orbitals that satisfy the variational theorem and minimize the energy will be eigenfunctions of $\hat{f}$ which is itself a function of these orbitals, thus the equations are highly coupled and exact solutions to this pseudo-eigenvalue problem cannot be obtained in all but the simplest cases.

$$\hat{f}\phi_i = [\hat{h}_i + \sum_j (2\hat{J}_j - \hat{K}_j)]\phi_i = \epsilon_i \phi_i \tag{1.9}$$

$$\hat{J}_j \left|\phi_i^i\right\rangle = \left\langle \phi_j^j \right| \hat{g}(\mathbf{x_i}, \mathbf{x_j}) \left| \phi_j^j \phi_j^i \right\rangle \tag{1.10}$$

$$\hat{K}_j \left|\phi_i^i\right\rangle = \left\langle \phi_j^j \right| \hat{g}(\mathbf{x_i}, \mathbf{x_j}) \left| \phi_i^j \phi_i^i \right\rangle \tag{1.11}$$

The Roothaan–Hall equations provide a means to numerically estimate the molecular orbital solutions by employing a basis set approximation. Each molecular orbital is defined as a linear combination of atomic orbitals (Equation 1.12), or basis set. The χ_q's are the atomic orbital functions and the C_{qi}'s are the MO coefficients defining how each χ_q contributes to the MO ϕ_i. The atomic orbitals are described by simple mathematical functions that approximate the well understood physical features of atomic orbitals such as the spherical distribution of a $1s$ orbital. By combining Equations 1.9 and 1.12 and left multiplying by the complex conjugate of the basis function (χ_p*), the integral relationships in Equation 1.13 are obtained for a particular MO which is then extended to the full Roothaan-Hall system of equations in matrix form (Equation 1.14, where $\mathbf{S}$ is the overlap matrix, $\mathbf{F}$ is the Fock matrix, $\mathbf{C}$ contains the MO coefficients, and ϵ is a diagonal matrix composed of the orbital energies), valid for the entire set of MOs.

$$\phi_i = \sum_q \chi_q C_{qi} \tag{1.12}$$

$$\sum \langle \chi_p^* | \hat{f} | \chi_q \rangle C_{qi} = \epsilon_i \sum_q \langle \chi_p^* | \chi_q \rangle C_{qi} \tag{1.13}$$

$$\boldsymbol{FC} = \boldsymbol{SC\epsilon} \tag{1.14}$$

The pseudo-eigenvalue problem can be transformed into a standard eigenvalue problem by means of the symmetric orthogonalization transformation by employing the identity matrix written in terms of the overlap matrix ($\mathbf{I} = \mathbf{S^{-1/2}S^{1/2}}$) to obtain Equation 1.15.

$$\boldsymbol{S^{-1/2}FS^{-1/2}S^{1/2}C} = \boldsymbol{S^{-1/2}SS^{-1/2}S^{1/2}C\epsilon} \tag{1.15}$$

By defining a transformed Fock matrix ($\widetilde{\mathbf{F}} = \mathbf{S^{-1/2}FS^{-1/2}}$) and coefficient matrix ($\widetilde{\mathbf{C}} = \mathbf{S^{1/2}C}$) a standard eigenvalue problem is obtained ($\widetilde{\mathbf{F}}\widetilde{\mathbf{C}} = \widetilde{\mathbf{C}}\epsilon$) where $\widetilde{\mathbf{F}}$ can easily be diagonalized. Final solutions are obtained using an iterative process where initial MO coefficients (i.e. elements of the $\boldsymbol{C}$ matrix) are guessed allowing for an initial energy to be produced as well, $\widetilde{\mathbf{F}}$ is built and diagonalized, the new $\boldsymbol{C}$ matrix is used to produce a new energy and $\widetilde{\boldsymbol{F}}$ matrix, and the procedure is repeated until the change in the energy and the MO coefficients is below some predetermined convergence threshold. The result of this procedure is the best set of molecular orbitals that minimize the energy of the system for the finite basis set utilized in Equation 1.12, under the mean field approximation.

The HF method, though very useful, rests on two primary assumptions that fail to address what is known as the correlation energy of the system. The first assumption is that in HF theory, each electron interacts with the "mean field" of all other electrons which necessarily increases the energy. In reality, the motions of electrons are instantaneously coupled to one another and electrons rearrange to lower the energy of the system. This is known as dynamic correlation. The second assumption HF makes is that the wavefunction can mostly be described by a single Slater determinant. Usually, this is a fair assumption, but falls apart in cases such as unimolecular bond cleavage and species with multiple resonance structures. When two or more Slater Determinants are necessary to describe the electronic structure of a system, it is said to have strong static correlation or be multireference. Multireference systems cannot be reliably described by HF based methods and can be diagnosed using a variety of methods such as inspecting the coupled cluster amplitudes or by obtaining a CASSCF wavefunction and determining the percent weight of each Slater Determinant. Both of these sources of correlation energy can be completely treated by expanding the wavefunction as a linear combination of all possible Slater determinants, within a basis of orbitals as shown in Equation 1.16. The $i, j, k, \cdots$ and $a, b, c, \cdots$ are indices corresponding to the occupied and unoccupied

orbitals, respectively. The $|\Phi\rangle$ terms correspond to Slater determinants with $|\Phi_0\rangle$ being the HF determinant and all others corresponding to excitations from occupied orbitals of the HF determinant into unoccupied orbitals according to the appropriate indicies. The c terms are configuration interaction constants that weight each of the Slater determinants in the expansion. Equation 1.16 is known as the full configuration interaction (FCI) wavefunction and corresponds to mixing between all possible electronic states.

$$|\Psi\rangle = c_0 |\Phi\rangle + \sum_{ia} c_i^a |\Phi_i^a\rangle + \sum_{i<j,a<b} c_{ij}^{ab} \left|\Phi_{ij}^{ab}\right\rangle + \sum_{i<j<k,a<b<c} c_{ijk}^{abc} \left|\Phi_{ijk}^{abc}\right\rangle + \cdots \tag{1.16}$$

The FCI wavefunction described with an infinite basis set is equivalent to the exact molecular wavefunction (Ψ) and produces and energy corresponding to the exact solution of the time-independent, non-relativistic Schrödinger equation. FCI is computationally infeasible in practice for all but the smallest systems as it requires building a matrix of interactions between all possible determinants, a procedure that exhibits factorial scaling with the number of basis functions. The difference between E_{HF} and E_{FCI} is defined as the total correlation energy of the system. Proper treatment of the correlation energy is necessary for reliable quantitative and often qualitative *ab initio* predictions. Thus, in order for our results to be most useful for experimental chemists, post-Hartree-Fock methods need to be employed that can capture as much correlation energy as possible in a cost effective manner.

1.4 Coupled Cluster Theory

As previously stated, the treatment of correlation energy in molecular systems is extremely important and can be critical for producing chemically meaningful results. The FCI energy, by definition, fully treats the correlation energy of a system as it considers contributions from all possible Slater determinants. It is possible to perform truncated CI computations on systems as well, which are far more affordable and can still help describe multireference systems. However, truncated CI approaches suffer from two serious problems: they lack size-extensivity and size-consistency. For a system to be considered size-extensive, the error in the energy should scale proportionally to the number of identical units in the system. For a system to be size-consistent, the sum of the energy of two fragments computed independently should equal the energy of the two fragments computed together at a non-interacting distance. However, it is well know that, due to the means by which truncated CI methods excite to different Slater determinants, neither of these features are maintained and introduce serious errors into the results. Coupled cluster, a technique first applied to quantum chemistry by Čížek and Paldus in the late 1960s and early 1970s[7–9], stands in contrast to the CI methods and provides a systematic, size extensive and size-consistent methodology to affordably approach

the FCI limit. The application of coupled cluster theory is a cornerstone of this research and warrants a brief summary of its mathematical foundation. Crawford provides an excellent review of coupled cluster methods as they apply to quantum chemistry and we summarize the cornerstone theoretical details below.

The fundamental idea behind coupled cluster theory is to approximate solutions to the Schrödinger equation (Equation 1.1) by estimating the true wavefunction (Ψ) as a reference wavefunction (Φ) having an exponential ansatz: $|\Psi\rangle = e^{\hat{T}} |\Phi\rangle$. The reference wavefunction can be anything in principle, but is always chosen to be Slater determinant corresponding to the Hartree–Fock wavefunction. The $\hat{T}$ term is known as the cluster operator and is composed of a sum of excitation cluster operators (Equation 1.17) that correspond to excitations from the reference determinant, the number of which is denoted by the subscript of each operator in the sum.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_N \tag{1.17}$$

Notice that the last term is $\hat{T}_N$ as there cannot be more excitations than electrons in the reference determinant. Each excitation cluster operator is further defined in Equation 1.18.

$$\hat{T}_n = (\frac{1}{n!})^2 \sum_{ij..ab..}^{n} t_{ij...}^{ab...} a_a^\dagger a_b^\dagger \cdots a_j a_i \tag{1.18}$$

The t terms are the coupled cluster amplitudes and are related to the probability of their respective excitations and parameterize the coupled cluster wavefunction. The subscripts i,j,$\cdots$ correspond to indices of occupied orbitals and the subscripts a,b,$\cdots$ correspond to indices of unoccupied orbitals. a_i and $a_a^\dagger$ correspond to an operator that annihilates the i'th orbital in the target determinant and an operator that creates the a'th orbital in the target determinant, respectively. For example, a $\hat{T}_2$ operator will consist of all possible determinants formed by annihilating two reference orbitals from the reference determinant and replacing them with two previously unoccupied orbitals, weighted by their respective cluster amplitudes. Having defined these terms, we can return to the wavefunction written in terms of an exponential ansatz. The ansatz can be expanded via a Taylor series (Equation 1.19).

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \frac{\hat{T}^4}{4!} + \cdots \tag{1.19}$$

If the series is not truncated, it is equivalent to the the FCI result, but this new formulation does not decrease the incredible cost of computing such a result. The coupled cluster method is most advantageous because, unlike a CI truncation, it is size-extensive and size-consistent and can be truncated without intro-

ducing significant error into the system. Truncations usually occur in the cluster operator (Equation 1.17), which is then placed into the infinite series defined in Equation 1.19. A cluster operator that stops with n=2 ($\hat{T} = \hat{T}_1 + \hat{T}_2$) is denoted as producing a CCSD, or "singles and doubles", energy results and likewise a cluster operator that stops with n=3 ($\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$) is denoted as CCSDT, or "singles, doubles, and triples". This truncation can be done up to any order less than or equal to the number of electrons in the system. Generally speaking, higher order terms contribute decreasingly to the final energy result (a simultaneous double excitation is far more probable than a quadruple excitation, e.g.). Increasing the number of excitations also significantly increases the scaling cost of each computation, thus it is generally that case that CCSD is utilized for most species (scaling $O(N^6)$, there N is the number of basis functions) and higher order coupled cluster is only employed in the smallest cases where extreme accuracy is desired. Coupled cluster's favorable convergence is primarily due to the inclusion of cross terms that are not present in the truncated CI expansion. For example, $(\hat{T}_1 + \hat{T}_2)^2 = \hat{T}_1^2 + \hat{T}_1\hat{T}_2 + \hat{T}_2^2$. Terms like $\hat{T}_1\hat{T}_2$ are not present in the truncated CI expansion and their inclusion is why coupled cluster is size-consistent and size-xtensive.

The final coupled cluster energy can be determined by left-projecting the reference determinant onto the Schrödinger equation (Equation 1.20).

$$\langle \Phi | \hat{H} e^{\hat{T}} | \Phi \rangle = E \tag{1.20}$$

Similarly, an expression can be obtained for the amplitudes by left-projecting with an excited Slater determinant instead (Equation 1.21).

$$\left\langle \Phi_{ij\cdots}^{ab\cdots} \right| \hat{H} e^{\hat{T}} | \Phi \rangle = E \left\langle \Phi_{ij\cdots}^{ab\cdots} \right| e^{\hat{T}} | \Phi \rangle \tag{1.21}$$

The problem with Equation 1.21, is that the coupled cluster amplitudes are coupled with the energy. This is generally avoided by utilizing a transformed Hamiltonian ($e^{-\hat{T}} \hat{H} e^{\hat{T}}$) which, when substituted into Equations 1.20 and 1.21, makes the right side of the equation equal to E and 0, respectively, assuming intermediate normalization. This allows for Equation 1.21 to be solved without the energy present on either side of the equation. In practice, the transformed Hamiltonian can be rewritten as a sum of nested commutators between the Hamiltonian and the cluster operators known as the Hausdorff expansion (Equation 1.22).

$$e^{-\hat{T}} \hat{H} e^{\hat{T}} = \hat{H} + [\hat{H}, \hat{T}] + \frac{1}{2!}[[\hat{H}, \hat{T}, \hat{T}] + \frac{1}{3!}[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}] + \cdots \tag{1.22}$$

According to Wick's Theorem, only fully contracted terms (i.e. terms that are connected to the Hamiltonian) in this expression will be nonzero. The fact that the Hamiltonian is composed of one and two electron

operators means that the Hamiltonian can contract with at most four cluster operators and thus terminates at $\hat{T}^4$. Similarly, many additional non-connected terms can be eliminated from the coupled cluster equations. Equations 1.23 and 1.24 are the final coupled cluster equations where subscript "c" denotes only the connected terms following transformation of the Hamiltonian.

$$\langle \Phi | \hat{H}_c e^{\hat{T}} | \Phi \rangle_c = E_c \tag{1.23}$$

$$\left\langle \Phi_{ij\cdots}^{ab\cdots} \right| \hat{H}_c e^{\hat{T}} | \Phi \rangle_c = 0 \tag{1.24}$$

CCSDT results which scale as $O(N^8)$ are far too expensive to compute for most systems. Nevertheless, the triples term has been demonstrated to be necessary for applications that require high accuracy result such as kinetics[10]. A solution to this dilemma is a perturbative correction for the triples term added onto the CCSD result. The perturbative triples treatment is denoted as CCSD(T) and is considered the gold standard of electronic structure theory due to its reliable results and reasonably affordable scaling costs ($O(N^7)$). The primary results in this study are generally obtained using the CCSD(T) method.

1.5 Basis Sets

So far, we have considered the methods that will be used to estimate exact solutions of the Schrödinger equation. Each of the methods previously discussed necessitate the use of some set of mathematical functions that are used to obtain numerical results for the MOs via the Hartree–Fock and the corresponding correlation energy. These functions are systematically designed to model the physical features of atomic orbitals and are collectively referred to as a basis set. It would be ideal to produce results that correspond to an infinite sized basis set, but that is computationally impossible. Instead, basis sets are engineered and selected to estimate this limit. Increasing the number of basis functions will generally approach the complete basis set (CBS) limit, but will also severely increase the computational cost. Therefore, significant work has gone towards the development of basis sets that efficiently approach the CBS limit with as few functions as possible. The correct selection of basis set can be as critical to an accurate description of a chemical system as the choice of theoretical method. For example, a FCI result computed with a minimal basis (i.e. one basis function for each atomic orbital) might be far worse than a CCSD result with a sufficiently large basis set. Therefore, it is worth discussing some of the technical hallmarks of basis set design and, in particular, focus on the Dunning cc-pVXZ family of basis sets[11] that are primarily utilized in this work.

Most modern basis sets are built from individual Gaussian functions which posses cost effective computational properties. Equation 1.25 provides the general form for a normalized Cartesian Gaussian function

where x, y, and z are the Cartesian coordinates, i, j, and k are integers that specify the quanta of angular momentum along each respective Cartesian axis, and α is a parameter that specifies the spread of the function.

$$\phi(x,y,z;\alpha,i,j,k) = (\frac{2\alpha}{\pi})^{\frac{3}{4}} \left[\frac{(8\alpha)^{i+j+k} i!j!k!}{(2i)!(2j)!(2k)!}\right]^{\frac{1}{2}} x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)} \tag{1.25}$$

The sum of the i, j, and k values indicates the type of atomic orbital the Gaussian function is modeling (e.g. a sum of 0,1, or 2 corresponds to s, p, and d, respectively) Gaussian type orbitals have the desirable feature of being analytically integrable which avoids expensive numerical methods to evaluate the molecular necessary for the Hartree–Fock procedure. Unfortunately, Gaussian type orbitals have some significant shortcomings such as their failure to model the nodal structure of atomic orbitals, a decay in electron density that is too fast as distance from the nucleus increases, and failure to represent an undifferentiable cusp at the nuclear center. These problems have been addressed by combining individual Gaussian functions (primitives) as linear combinations to form what is known as a contracted basis function. The exponents and contraction coefficients of each type of basis set are variationally optimized to best agree with experimental data such as the ionization potential of atoms[11,12]. Many additional modifications can be made to a basis set to describe a specific physical feature of a system or to provide optimal results from a particular kind of method. To understand these modifications in the context of this body of research, the following will discuss the primary features and modifications of the Dunning cc-pVXZ family of basis functions which were designed to accurately describe the energy of systems using correlated methods, such as coupled cluster.

The cc-pVXZ basis set nomenclature stands for "correlation consistent polarized valence (*cardinality*) zeta", where *cardinality* = double, triple, quadruple, or pentuple, for X = D, T, Q, 5, respectively. The cardinality controls the basis set size and determines how many basis functions are utilized to describe the valence atomic orbitals, the most important orbitals for describing most chemical phenomena. Generally, the core orbitals of like atoms manifest similar MO coefficients and thus require less flexibility in their description than the valence orbitals. The correlation consistent designation highlights the systematic design of the Dunning basis sets such that an increase in *cardinality* adds basis functions in such a way way that effectuates smooth and predictable convergence towards the CBS limit. This is a necessary requirement for basis set extrapolation which will be discussed further in the next section. Polarization is another important feature of these and many other basis sets that allows for additional mathematical flexibility by adding higher angular momentum basis functions in the description of atomic orbitals. For example, the cc-pVDZ basis set for oxygen has one d function that can provide a better description of the p valence orbitals. As the *cardinality* increases in the Dunning basis set, additional functions are added such that a cc-pVTZ

oxygen atom would have one f and two d functions and a cc-pVQZ basis would have one g, two f, and three d functions, with the trend continuing for higher X. The original cc-pVXZ basis sets were incorrectly parameterized for third row elements when they were first introduced. They have been since updated with the correct inclusion of tight d functions and are now denoted as cc-pV(X+d)Z[13]. This work appropriately utilizes the cc-pV(X+d)Z basis sets to produce any results containing third row elements.

The Dunning basis sets can be further modified with special basis functions to handle unique cases, some of which are present in this work. The "aug" prefix can be added to any cc-pVXZ basis set to denote the inclusion of additional diffuse basis functions for each angular momentum already present in the parent basis set[14]. These diffuse functions are necessary for describing systems where electron density is far from the atomic centers such as anions, noncovalent interactions, and strong dispersion interactions. Slightly more cost effective versions of these basis sets have been developed which remove the highest angular momentum diffuse function for each step as the series progresses ("jul", "jun", "may",$\cdots$).[15] The Dunning basis sets can also be modified to describe core-valence (cc-pwCVXZ[16]) and core-core (cc-pCVXZ[12]) correlation. In most cases, a frozen core approximation (only correlating the valence electrons) and a cc-pVXZ basis is sufficient to describe a system. Certain cases, such as atoms containing "valence-like" d electrons require correlation of outer core electrons to produce reliable results. The cc-pwCVXZ and cc-pCVXZ basis sets and relaxation of the frozen core approximation can produce more rigorous results, but often are not worth it considering the significant increase in their size and therefore computational cost.

One final basis set modification utilized in this work is the inclusion of an effective core potential (ECP) denoted by the "-PP" appended to the basis set name[17]. An ECP replaces some set of the innermost electrons with functions that properly describe the core and possess computationally simple analytic solutions. The use of an ECP is often valid as the inner core of heavier atoms is so much lower in energy, that its contribution to the correlation energy is minimal in most systems. ECPs are advantageous because they decrease the number of basis functions needed to describe large atoms, but are also able to describe the relativistic effects that manifest in the core electrons of heavy atoms. Thus, an ECP can remove the need to treat scalar relativistic effects in most systems without the need for a relativistic Hamiltonian to be employed. Otherwise, treating relativistic effects often requires a large decontracted basis set and is not computationally feasible in most cases.

The Dunning basis sets are powerful and robust for *ab initio* computations as each of the aforementioned modifications (augmentation, core weighting, and ECPs) can be combined and used together while maintaining the correlation consistent features necessary to converge results towards the CBS limit. In addition, full sets of auxiliary basis sets are available for the Dunning family which allows for utilization of the density fitting approximation where appropriate.

1.6 Focal Point Analysis

The previous sections have discussed the foundational technical details necessary to produce solutions of the Schrödinger equation with very high levels of theory. One of the most important applications of these results is to characterize relative energies of stationary points on a PES. Generally a reliable level of theory such as CCSD(T)/cc-pVTZ should be able to produce predictions that are accurate enough to answer some chemical research questions. There are many cases, however, where the *ab initio* predictions must have uncertainty in the electronic energies that is below chemical (1 kcal mol^{-1}) or sub-chemical (1 kj mol^{-1}) accuracy, e.g. combustion kinetics. In addition, we have no way to ensure that a single-point energy is sufficiently converged with respect to finite basis set error or correlation energy. The focal point analysis (FPA) method of Allen and coworkers[18–21] was developed as a tool to produce highly accurate energetic results and observe the convergence towards the FCI/CBS limit as finite basis set error and correlation energy are both systematically treated. FPA is utilized heavily in this work and not only produces results approaching the *ab initio* limit, but diagnoses any underlying problems that might have gone unnoticed if results were obtained using a single level of theory. The rest of this section will explain the mechanics of FPA and how it is used to ensure the prediction of highly converged electronic energy predictions.

In in order for a FPA to produce trustworthy relative energies, it requires product and reactant geometries optimized at a high quality level of theory such as CCSD(T)/cc-pVTZ. Geometries obtained at lower levels of theory risk introducing error due to the geometry optimization that defeats the purpose and misrepresents the results of a FPA. Assuming quality geometries are obtained, single point energies are computed for each method and as many cc-pVXZ basis sets that are computationally feasible. The methods always begin with Hartree–Fock as the baseline and continue to MP2, CCSD, CCSD(T), CCSDT, CCSDT(Q), CCSDTQ, and so on. Due to the increased scaling of higher order coupled cluster methods, the basis set size is increasingly limited and eventually reaches a point where even a computation with a small basis like cc-pVDZ cannot be obtained. Fortunately, higher order coupled cluster corrections are generally decreasingly basis set dependent[22] and can be included as an additive correction using a small basis set. The final results are often presented in an incremental table where the contribution of each method to the predicted FCI/CBS limit is shown (See Table 1.1 for an example). Any values that are in brackets are not explicitly computed and are extrapolated values or additive corrections. The FPA is robust and can be performed with any variation of the Dunning basis sets as long as the modifications are consistent across the entire set of data (e.g. all augmented or core-weighted basis sets).

First considering the rows (i.e. the basis set dimension) of an a FPA table, energies are computed with

Table 1.1: Example incremented FPA table for the relative energy of carbamate relative to water and isocyanaic acid on top of CCSD(T)/cc-pVQZ optimized geometries. Units are kcal mol $^{-1}$.

$\mathbf{HNCO + H_2O \longrightarrow NH_2COOH}$							
	HF	$+\delta$ MP2	$+\delta$ CCSD	$+\delta$ (T)	$+\delta$ T	$+\delta$ (Q)	NET
cc-pVDZ	-26.97	4.82	-2.45	1.55	0.09	0.29	-22.68
cc-pVTZ	-24.27	3.52	-2.67	1.33	[0.09]	[0.29]	[-21.69]
cc-pVQZ	-23.35	3.32	-2.97	1.28	[0.09]	[0.29]	[-21.35]
cc-pV5Z	-23.04	3.30	-3.17	1.28	[0.09]	[0.29]	[-21.25]
cc-pV6Z	-22.95	[3.29]	[-3.26]	[1.28]	[0.09]	[0.29]	[-21.26]
CBS	[-22.93]	[3.28]	[-3.38]	[1.28]	[0.09]	[0.29]	[-21.37]

increasingly higher carnality basis sets as far as possible for each method. The results should demonstrate convergence and provide an estimate of how much error remains due to the use of a finite basis set. If the convergence is too slow, this could indicate an improper basis set selection or the need for larger basis sets to properly describe the system. Schemes developed by Helgaker[23] and Feller[24] are commonly employed to further reduce error due to incomplete basis set via extrapolating to the CBS limit by fitting the three point $E_{\mathrm{HF}} = E_{\mathrm{HF}}^{\infty} + ae^{-bX}$ functional form for the Hartree–Fock energies and the two point $\epsilon_{\mathrm{corr}} = \epsilon_{\mathrm{corr}}^{\infty} + aX^{-3}$ functional form for all correlation energies, where X is the *cardinality* of a Dunning basis set. The CBS estimates for Hartree–Fock and correlated methods can be obtained with energy computations for three or two consecutive basis sets *cardinalities* and using Equations 1.26 and 1.27, respectively. Equation 1.26 is considered in the case of an X = TZ, QZ, 5Z extrapolation and Equation 1.27 defines a X = TZ, QZ extrapolation, as employed in Table 1.1.

$$E_{\infty} = \frac{E_{TZ}E_{5Z} - E_{QZ}^2}{E_{TZ} - 2E_{QZ} + E_{5Z}} \tag{1.26}$$

$$\epsilon_{\infty} = \epsilon_{QZ} - \frac{\epsilon_{QZ} - \epsilon_{TZ}}{4^{-3} - 3^{-3}} \times 4^{-3} \tag{1.27}$$

It should be noted that the extrapolations should not use any *cardinality* less than X=TZ, as smaller basis sets tend to produce poor extrapolation results. The robustness of the final FCI/CBS results can be tested by employing multiple extrapolation schemes and observing if they are in accordance with one another or not.

The second consideration in the focal point table is the columns (i.e. methods) which systematically increase with higher orders of coupled cluster. The FPA table reveals the convergence of the correlation energy by computing increasingly higher order coupled cluster energies, until they become cost prohibited. Most of the time only the X=DZ *cardinality* is used for CCSDT and above due to the intense scaling of

the higher order methods, but in principle, all can be computed and extrapolated if X = TZ and X = QZ values can be obtained. Many studies consider CCSD(T)/CBS results to be the "gold standard" [25–28]. Even if these results are excellent, computing higher order CCSDT or CCSDT(Q) results can confirm the correlation energy estimation of CCSD(T) and provide an idea of how much error remains due to correlation energy. Because of this, a FPA is necessary to properly describe cases that are borderline multireference rather than blindly trusting CCSD(T). Usually, the higher order coupled cluster energies are included as additive corrections which are assumed to have little dependence on basis set [29,30]. Using the representative example of Table 1.1, one would say that this FPA produces a relative energy that targets the CCSDT(Q)/CBS limit and certainly demonstrates convergence within chemical accuracy.

The FPA can be further extended to include additive corrections for various physical effects that are not included in the electronic energy computations. In each case, an additive correction is calculated by first computing some quantity intrinsic to each product and reactant geometry. This might be the zero-point vibrational energy (ZPVE) or some difference computed on the same geometry with and without the inclusion of some treatment to measure its influence (e.g. the frozen core approximation). The sum of the product values are then subtracted from the sum of the reactant values and the correction is added to the predicted FPA relative energy. Some corrections are large and necessary to produce a meaningful prediction while others serve a more diagnostic role and justify assumptions inherent to the geometry optimization and FPA. For example, a ZPVE correction is usually large and converts the predicted relative energy to a 0 K enthalpy which is often needed to compare to experiment. Corrections that predict the influence of the frozen core approximation, scalar relativistic effects, or diagonal Born–Oppenheimer (DBOC) effects are generally small and have little influence on the final result. If one of these were to be large, it might indicate that significant error persists in the final result which may warrant further investigation, or in the case of the DBOC that there is a nearby surface crossing or conical intersection. The specific corrections included and the levels of theory which they are obtained vary case by case depending on the system being studied. This research heavily relies on FPA to robustly predict the energetics of the systems in question and ensure the results are trustworthy for many years to come.

1.7 Natural Bond Orbital Analysis

The previous sections have summarized the methodology used in this research to produce high quality wavefunctions and solutions of the Schrödinger Equation, and therefore, quantitatively meaningful predictions. Though quantitative predictions are in high demand and form the backbone of this work, they are significantly more useful if associated with chemical insights. These insights are often qualitative in nature

but provide great understanding and flexibility for the findings to be generalized to related chemical systems. One of the best tools towards this end is Natural Bond Orbital (NBO) analysis, which is utilized substantially in this work. NBO analysis obtains a set of orbitals uniquely associated with the wavefunction often referred to as a "chemist's basis set" as it consists of orbitals that are transformed to correspond to "Lewis" and "non-Lewis"-type orbitals, rich in chemical intuition. One advantage of this is the ability to produce numerical predictions that can provide insight into the qualitative trends manifest in the high-level quantitative results. The variety of tools and insights that NBO can provide to a study are numerous and the most utilized in this research are discussed with an overview of the theoretical background necessary to understand them. The following discussion is based on the the excellent reviews and seminal papers of Weinhold, Glendening, and coworkers.[31–35]

As shown by Löwdin[36], any wavefunction can be described uniquely by a set of orbitals that are the eigenorbitals of the first-order reduced density operator ($\hat{\Gamma}$, as shown in Equation 1.28).

$$\hat{\Gamma}\Theta_k = p_k\Theta_k \tag{1.28}$$

The eigenvalue (p_k) of each orbital (Θ_k is the population (between zero and two for closed shell systems) and will always correspond to the set of maximum occupancy orbitals. It is often useful to extend this analysis and instead diagonalize localized blocks of the density matrix that correspond to specific atoms and produce what is known as preorthogonalized natural atomic orbitals (PNAOs), denoted as ${}^{p}\Theta_k^A$ for atom A. These orbitals can be orthogonalized between atomic blocks (using the occupancy weighted symmetry orthogonalization procedure[37]) and produce NAOs which are orthonormal with respect to all other orbitals within and outside of their respective atomic block. The populations of the NAOs for a single atom sum to an atomic population which can be summed across atoms to produce the total number of electrons in the system.

The NAO basis is an important intermediate necessary for the determination of Natural Bond Orbitals (NBOs). The NBOs are determined via an algorithm that removes all core orbitals (K_A) from the density matrix via some high occupancy threshold and leaves them unhybridized. Next, the program variationally searches all possible Lewis structure representations of the molecule, selects the representation that maximizes the occupancy of the "Lewis"-type orbitals, and sorts the remaining high occupancy orbitals as localized to single atoms (lone pairs (n_A) or between a pair of atoms (a bonding orbital σ_{AB}). This forms a full set of NBOs, $(K_A)^2)(n_A)^2(\sigma_{AB})^2$, which can be considered the best Lewis structure description of the system. The sum of each NBO's occupation relative to the number of electrons in the system is a metric of how well a particular Lewis structure describes the electronic structure of the system. The NBO analysis prescribed

is inexpensive and is usually qualitatively robust with respect to method and basis set size.[34]

One of the first insights from NBOs into a molecular system is their composition based on natural hybrid orbitals (NHOs). Each bonding NBO is constructed as a linear combination of two NHOs (h_A, h_B for atoms A and B in Equation 1.29) and each lone pair orbital consists of a single h_A. The NHOs are themselves optimized linear combinations of NAOs, represented in Equation 1.30.

$$\sigma_{AB} = a_A h_A + a_B h_B \tag{1.29}$$

$$h_a = \sum_k c_k \Theta_k^A \tag{1.30}$$

The a coefficients defining a σ_{AB} orbital (Equation 1.29) are known as the polarization coefficients and determine the polarization of a bond, i.e. the contribution from each NHO to the bond. The coefficients (c) in Equation 1.30 describe the hybridization of each NHO which can put a numerical estimate to the hybridization which is generally considered conceptual. For example, NBO might predict an $sp^{2.7}$ hybrid directed away from a carbon atom in some bond. This hybridization would slightly change depending on the nature of the carbon bond and the electronic influence of other moieties in the molecule. The antibonding orbitals, which can be important for many chemical analyses, are created from this process as well and constructed as the out of phase analogue of each bonding orbital (Equation 1.31).

$$\sigma_{AB}^* = a_B h_A - a_A h_B \tag{1.31}$$

Simply studying the composition of NBOs in a system and considering how their features (e.g. bond order, hybridization, and polarization) change as a reaction progresses or the system is modified can provide invaluable insights into the chemistry manifest in our sophisticated wavefunction predictions. In many well behaved systems, the NBO Lewis structure representation will provide a reasonable description of the system, that is to say that the core, lone pair, and bonding orbitals contain more than 90 % of the electron density. However, there always persists some electron density that occupies non-Lewis orbitals and, in extreme cases, more than one Lewis structure might be necessary to properly describe the system. Even in the first case, the delocalization of electron density into "non-Lewis"-type orbitals can have a significant impact on the energetics of a system. The specific delocalization motifs can be elucidated via the second-order perturbation analysis of interactions between "Lewis" (denoted by subscript i) and "non-Lewis"-type orbitals (denoted by subscript a) using Equation 1.32.

$$E_{ia}^{(2)} = q_i \frac{F_{ia}^2}{\epsilon_a - \epsilon_i} \tag{1.32}$$

The numerator contains the square of the Fock matrix element between orbital i and a. The denominator is the difference in energy between the unoccupied orbital (ϵ_a) and the occupied orbital (ϵ_i), and the entire fraction is multiplied by the occupancy (q_i) of the ith orbital. The absolute values of these $E^{(2)}$ quantities are not so important per se, but their relative magnitudes are. The NBO program computes all possible combinations of $E^{(2)}$ values, but most will be negligibly small and not printed if below a set tolerance. There will however be some terms that are large and help elucidate where electron delocalization occurs in a molecule and attaches a quantitative estimate to qualitative chemical properties. Thus, given a molecule, we can begin to explain energetic trends in terms of where electrons delocalize to and how modifications to the system might influence these patters and relate to our high-level predictions. The $E^{(2)}$ energies can be leveraged further using the "del" command which zeros out all of a set of particular Fock matrix elements and then performs the same $E^{(2)}$ analysis, allowing for insights into how our true delocalized wavefunction might differ from a perfect Lewis picture.

In order to more fully describe significantly delocalized species, Weinhold and Glendening developed Natural Resonance Theory (NRT)[38–40]. The NRT algorithm uses the results of a $E^{(2)}$ analysis to predict other possible Lewis structures that could be valid descriptions of a specie's electronic density. A full NBO analysis of each of these alternate Lewis structures is performed and the NRT program uses the results to determine a weighted average NRT description. Each Lewis structure is given a weight, which in total sum to 100%. The primary benefit of this analysis is a metric of how valid a single Lewis structure description of the system is. Furthermore, NRT allows the program to produce weighted average bond orders which are more accurate descriptions the bond order obtained from a single Lewis representation. All of the previously discussed NBO methods are utilized heavily in this research and provide invaluable chemical insights into each of the studied systems, particularly for periodic trends.

1.8 Summary

The previous sections provide a foundation for a general understanding of the methods and theories employed in this research. The following chapters will present four research projects that explore a diverse set of chemical species with high-level *ab initio* methods and explain the chemical insights gained from each study. The unifying theme of this dissertation is the exploration of trends (particularly periodic trends) manifest in a set of important chemical systems and how a complete understanding of said trends could be useful

for further research on more complex systems. Chapter two examines the $HOX\cdots H_2O$ (X= F, Cl, Br, I) binary complex which are of significant importance to atmospheric chemistry. In particular, the focus of this study is a quantitative determination of the competition between hydrogen and halogen bonding, as halogen size increases. The next chapter explores the $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) tetrahedrane species which have yet to be isolated. Our findings explore how their electronic structure features change with increasing pnictogen size and how electron delocalization is a primary cause of these changes. Chapter four considers the reaction of substituted isocyanates with water and their ability to form carbamates and imidic acids. Our work predicts how substituents and catalyst molecules influence the barriers to product formation and benchmark all of our findings against highly reliable *ab initio* characterization of the unsubstituted case. Finally, Chapter five presents a novel undergraduate chemistry lab exercise that teaches students about the tension between computation cost and accuracy via *ab initio* study of the combustion of methane. The necessity of the exercise is suggested with respect to modern chemical pedagogy and the student learning goals are summarized in conjunction with detailed explanation of all provided instructional materials. Each chapter of this work provides detailed background information about each system, specifies the theoretical methods employed, and presents the data and findings of each study. The conclusions of this research are unified in their ability to elucidate important periodic trends present in the electronic structures of small systems with the aim of guiding future chemical research on increasingly complex systems.

CHAPTER 2

THE HYPOHALOUS ACID-WATER BINARY COMPLEXES

M. E. Wolf, B. Zhang, J. M. Turney, H. F. Schaefer, *Physical Chemistry Chemical Physics* **2019**, 21 (11), 6160-6170, Reproduced from[41] with permission from the PCCP Owner Societies.

2.1 Abstract

Hypohalous acids (HOX) are a class of molecules that play a key role in the atmospheric seasonal depletion of ozone and have the ability to form both hydrogen and halogen bonds. The interactions between the HOX monomers (X=F, Cl, Br) and water have been studied at the CCSD(T)/aug-cc-pVTZ level of theory with the spin free X2C-1e method to account for scalar relativistic effects. Focal point analysis was used to determine CCSDT(Q)/CBS dissociation energies. The *anti* Hydrogen bonded dimers were found with interaction energies of –5.62 kcal mol^{-1}, –5.56 kcal mol^{-1}, and –4.97 kcal mol^{-1} for X=F, Cl, and Br, respectively. The weaker halogen bonded dimers were found to have interaction energies of –1.71 kcal mol^{-1} and –3.03 kcal mol^{-1} for X=Cl and Br, respectively. Natural bond orbital analysis and symmetry adapted perturbation theory were used to discern the nature of the halogen and hydrogen bonds and trends due to halogen substitution. The halogen bonds were determined to be weaker than the analogous hydrogen bonds in all cases but close enough in energy to be relevant, significantly more so with increasing halogen size.

2.2 Introduction

In recent years, there has been increasing interest in hypohalous acids (HOX, X= F, Cl, Br) due to the critical roles they play in the seasonal depletion of ozone in the atmosphere.[42–50] Hypohalous acids are thought to be "sinks" for halogen radical species, which also directly react with ozone molecules.[51,52] Kinetics simulations have indicated that the reaction between HOCl and HCl is necessary to maintain a usable level of activated chlorine, which is a primary destroyer of ozone.[53] HOX molecules also play a prominent role in biochemistry. Their oxidative nature makes them extremely reactive with biological structures (e.g., NADH,[54] DNA,[55] heme groups,[56] cytochrome c[57]) and often leads to toxic effects in the human body. The biological toxicity of HOCl can conversely be beneficial to the human body by acting as an antibiotic for external wound care.[58] Additionally, hypohalous acids are relevant in many other fields such as molecular biology[59] and immunology.[60]

A unique trait of HOX molecules is their ability to form both hydrogen (HB) and halogen (XB) bonds. The importance of these two noncovalent interactions cannot be overstated, as they are critical to biological systems,[61] organometallic chemistry,[62] crystal structures,[63] and many other applications.[64–66] Moreover, protonated HOCl is a potential transition state for one of the key reactions leading to ozone depletion,[67] the reactivity of which can be intimately affected by existing noncovalent interactions with other atmospheric molecules. Even more pertinent to our present research is that HOX molecules are often in complex with one or more water molecules before reacting in the atmosphere.[68–70] The fact that HOX molecules can partici-

pate in both types of bonds provides a prime opportunity to study the competitive and cooperative nature of these interactions. Both Hydrogen[71–74] and halogen[75–91] bonding have been well studied experimentally and theoretically and we direct the reader to a plethora of literature on both. XB and HB are fundamentally electrostatic processes, however, the importance of dispersion forces are important to correctly model XB.[77,86,92] It is widely accepted than the hydrogen bond is the stronger of the two interactions, however, experimental and theoretical suggest that halogen bonds can be made stronger than hydrogen bonds in a variety of chemical settings.[93–97] Therefore it is beneficial to know the accurate relative strengths of the halogen and hydrogen bonds between HOX and water molecules, as it could affect atmospheric models.

Hypohalous acids are highly reactive and therefore notoriously unstable and difficult to study experimentally.[57] While spectroscopy has confirmed the presence of HOX species in the atmosphere,[98–101] only HOF has been isolated in pure crystalline form.[102] To the best of our knowledge no experimental data is available quantifying the interaction energy of the HOX molecule and water. A theoretical approach is thus quite appropriate. Previous research groups have published computational studies on the intermolecular interactions between HOX species and other molecules.[93,103–112] By far the most common theory employed in the treatment of these complexes is second order Møller–Plessest (MP2) theory, often complemented with density functional theory (DFT). A non-exhaustive list of recent publications includes HOX structural formations with formaldehyde, formamidine,[103] formyl halides,[105] phosphorus ylide,[113]ozone,[107] and sulfoximine.[108] The motivation for most of these studies is to analyze the nature of the hydrogen and halogen bonds that are formed and how the bonding nature changes across the halogen series. In particular, the halogen bond has quite variable strength depending on the halogen atom involved.[78,93] Properly explaining this trend is key to a foundational understanding of the relative strength of these noncovalent interactions in more complex chemical settings. The most comprehensive theoretical study to date of the HOX and water binary complex is that of Panek and Berski that used DFT to study the hetero and homo binary complexes of HOX (X= F, Cl, Br) with water.[114] They determined that DFT increasingly underestimated the interaction energies with a B3LYP functional compared to MP2 (with the same basis set) as the size of the halogen increased. They also applied symmetry adapted perturbation theory (SAPT) to decompose the interaction energy of the noncovalent interactions into physically meaningful components, which varied significantly depending on the halogen involved. Their SAPT results predicted that dispersion interactions were of critical importance to XBs but not HBs agreeing with the research Anderson and coworkers on halogen bonding in general.[77] A similar paper by Zhang and coworkers used MP2/aug-cc-pVTZ to study all of the homobinary complexes of HOCl.[115] They found six structures and determined with natural bond orbital analysis (NBO) that, for most structures, donor-acceptor and charge transfer interactions contributed primarily to complex formation. The NBO analysis also confirmed that dispersion interactions were less important relative to electrostatic

forces in five out of six structures, in disagreement with the findings of Panek and Berski.[114] Multiple theoretical studies have also been published that specifically study binary complexes formed between HOX and water. An early DFT study of the $HOBr\cdots H_2O$ was performed by Ying and Zhao in 1997.[116] Santos and coworkers mapped the MP2 potential energy surface of $HOBr\cdots H_2O$ in 2004,[117] and provides the most accurate structures to date optimized at the CCSD/6-311++G(2d,2p) level of theory. They determined that electron correlation had little effect on interaction energies. In 1995, Dibble and Franciso presented MP2/6-311++G(d,p) geometries along with interaction energies for HOCl and water binary complexes.[68] Qiao and coworkers have published related MP2 work on $HOBr\cdots H_2O$ binary complexes.[118]

Despite the variety of HOX compounds studied in the literature, there is insufficient theoretical rigor in the previously mentioned publications. Most studies use modest (by 2019 standards) levels of theory to optimize geometries and compute energetics. Although DFT functionals may capture interaction energies, there exists no high level *ab initio* study to confirm these results. Thus, we cannot know *a priori* if the functionals used are appropriate. Similarly, many studies use basis sets that may be insufficient. For example, a paper by Zhang and coworkers presents structures for HOCl binary complexes using the aug-cc-pVTZ basis set.[115] However, Dunning and coworkers have published research that establishes the importance of additive tight *d* functions in obtaining accurate descriptions of molecules containing Al-Ar atoms.[13] Likewise, none of the previous studies comment on relativistic effects that may arise from heavier atoms like Br and I. Failing to account for relativity in heavier atoms can seriously skew the true periodic trends of halogen bonding. The combination of these sources of error creates a need for more rigorous study of these systems. Our work seeks to provide a rigorous and accurate theoretical study of the binary complexes formed by HOX molecules and water. Accurate HOX/water structures are computed at the CCSD(T) level of theory, along with a focal point approach to obtain CCSDT(Q)/CBS interaction energies. Proper treatment of scalar relativistic effects is included. An in depth analysis into the nature of HB and XB interactions is provided. Our results can be used as a theoretical benchmark for any future studies pertaining to HOX species.

2.3 Theoretical Methods

Equilibrium geometries were optimized for the HOX monomers (X = F, Cl, Br, I), H_2O, $HOX\cdots OH_2$, and $XOH\cdots OH_2$ structures using coupled cluster theory with single, double, and perturbative triples excitations [CCSD(T)][10,119–121] as implemented in the software package CFOUR.[122,123] Harmonic vibrational frequencies were computed to ensure each structure was a minimum on its potential energy surface. The SCF density, coupled cluster amplitudes, and lambda equations were converged to 10^{-10}. Since we are studying heavy atoms, the inclusion of relativistic effects are needed. The standard one-electron Hamiltonian was aug-

mented with the one-electron variant of the spin-free exact two-component Hamiltonian (SFX2C-1e)[124–126] to include scalar relativistic effects for both geometry optimization and harmonic vibrational frequency computations. The use of the SFX2C-1e method requires properly contracted relativistic basis sets. Since the Dunning basis sets are nonrelativistic basis sets, we used the X2C recontracted basis sets available on the CFOUR website.[127] The aug-cc-pVTZ-X2C basis set[11] was used for all bromine and fluorine structures, and the aug-cc-pV(T+d)Z-X2C basis set[13] was used for all chlorine atoms.[77] All computations were performed with the frozen core approximation. Preliminary tests agreed with previous research claiming that for a proper description of heavy atom halogen bonds, the $3d$ electrons needed to be correlated for all bromine containing structures.[128,129] Therefore, the aug-cc-pCVXZ-X2C basis[130] set was employed for all structures including bromine atoms.

Interaction energies, $E_{\text{int}} = E_{\text{complex}} - (E_{\text{HOX}} + E_{\text{H}_2\text{O}})$, were determined by means of the focal point method developed by Allen and coworkers.[18–21] The three point formula of Feller[24] was used to extrapolate HF energies while the two-point formula of Helgaker[23] was used to extrapolate each post-HF energy. Each extrapolated component is then summed to obtain a CCSDT(Q)/CBS interaction energy. The SFX2C-1e one-electron Hamiltonian was used for our extrapolated energies, along with the X2C recontracted Dunning basis sets. The focal point computations were performed on the CCSD(T)/aug-cc-pVTZ-X2C optimized geometries. The CCSDT(Q)/CBS energies were then corrected to account for approximations used in previous computations. The zero-point vibrational energy correction, ΔE_{ZPVE}, was computed at the CCSD(T)/aug-cc-pVTZ-X2C level of theory. The diagonal Born–Oppenheimer correction, ΔE_{DBOC}, were computed at the Hartree–Fock level of theory with the same basis set to account for the assumption of fixed nuclei.[131,132] A frozen core correction, $\Delta E_{\text{FC}} = E_{\text{AE}} - E_{\text{FC}}$, was also computed to account for core electron correlation. E_{AE} and E_{FC} were computed at the CCSD(T)/aug-cc-pCVTZ-X2C level of theory with all electrons and only the valence electrons correlated, respectively.[12] For bromine atoms, the $3d$ electrons were not included in the frozen core approximation. Thus, the final focal point energy for each structure is given by, $E_{\text{final}} = E_{\text{CCSDT(Q)}-\text{X2C/CBS}} + \Delta E_{\text{ZPVE}} + \Delta E_{\text{FC}} + \Delta E_{\text{DBOC}}$. The corrected energies are used to obtain the interaction energies between HOX monomer and water.

Natural bond order (NBO) analysis was performed to assess the donor and acceptor nature of the bonding orbitals.[34] The NBO6.0 package[35] as interfaced to QChem[133] was used to compute the second-order perturbation energy, $E^{(2)}$ of a hybrid orbital overlap (Equation 3.3).

$$E^{(2)} = q_i \frac{F_{ia}^2}{\epsilon_a - \epsilon_i} \tag{2.1}$$

F_{ia} is defined as the NBO Fock matrix element between donor natural bond orbital i and acceptor

natural bond orbital a. ϵ_i and q_i are defined as the orbital energy and orbital occupation, respectively, of natural bond orbital i. The B3LYP functional[134] with the aug-cc-pVDZ basis set was used for all NBO computations. The orbitals were visualized using the software package Jmol.[135]

Symmetry adapted perturbation theory (SAPT) was used to dissect the interaction energies of the hydrogen and halogen bonds into physically meaningful components: dispersion, induction, exchange, and electrostatic (Equation 2.2).[136,137] A more detailed description of the decomposition scheme is available in the supplementary information. SAPT2+3 computations were performed using the Psi4[138] software package with the standard aug-cc-pVTZ basis (The -PP version is employed for iodine).

$$E_{\text{SAPT2+3}} = E_{\text{Electrostatic}} + E_{\text{Exchange}} + E_{\text{Induction}} + E_{\text{Dispersion}} \tag{2.2}$$

2.4 Results

2.4.1 Equilibrium Geometries

Two types of favorable interactions were found between the HOX monomer and water. One was a binary complex formed by a hydrogen bond between the hydrogen on the hypohalous acid and the oxygen in water. These hydrogen bonded (HB) binary complexes are denoted as $XOH\cdots OH_2$. The second type of interaction was a halogen bond formed between the halogen of the HOX monomer and the oxygen of water. These halogen bonded (XB) binary complexes are denoted as: $HOX\cdots OH_2$. For the $XOH\cdots OH_2$ binary complexes, geometries were found for X=F and Cl with each having both a *syn* and *anti* conformer (Figure 2.1). For X=Br, the *anti* conformer was a minimum on the potential energy surface, but the C_S *syn* conformer was not ($22i$ cm^{-1} vibrational mode). Panek and Berski found this minimum *syn* structure at the MP2/aug-cc-pVTZ level of theory to have a dihderal of 3.2°.[114] Scans were performed to search for a structure that was not C_S but the potential energy surface was found to be too shallow for a minimum structure at our level of theory. The *syn* binary complexes were slightly more favorable energetically by 0.07 and 0.11 kcal mol^{-1} for X=F and Cl, respectively.

For the halogen bonded binary complexes, geometries were found for X = Cl and Br. Both *syn* and *anti* conformers were found for $HOBr\cdots OH_2$, but only the *anti*-$HOCl\cdots OH_2$ structure was found for chlorine (Figure 2.2). The *anti*-HOBr–OH_2 conformer is the more favorable conformer by 0.14 kcal mol^{-1}. A structure for *syn*-HOCl–OH_2 was optimized but was not a minimum on the potential energy surface ($35i$ cm^{-1} vibrational mode). This imaginary mode was due to the torsional vibration about the halogen bond. We agree with previous theoretical research that was unable to find any $HOF\cdots OH_2$ structure. This in part

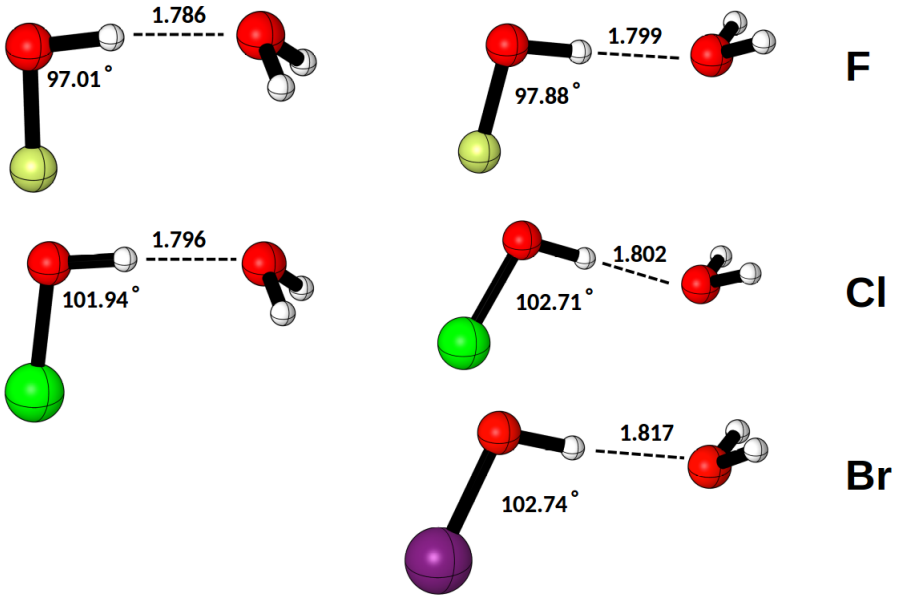

Figure 2.1: CCSD(T)/aug-cc-pVTZ-X2C optimized geometries for the hydrogen bonded binary complexes. The *syn* conformers are shown on the left and the *anti* conformers on the right. The hydrogen bond lengths (in Å) and internal X-O-H angles (in degrees) are shown. Atoms are colored gold, green, and purple for F, Cl, and Br, respectively.

due to the fact that fluorine is more electronegative than oxygen and thus HOF does not have a significant sigma hole on F.

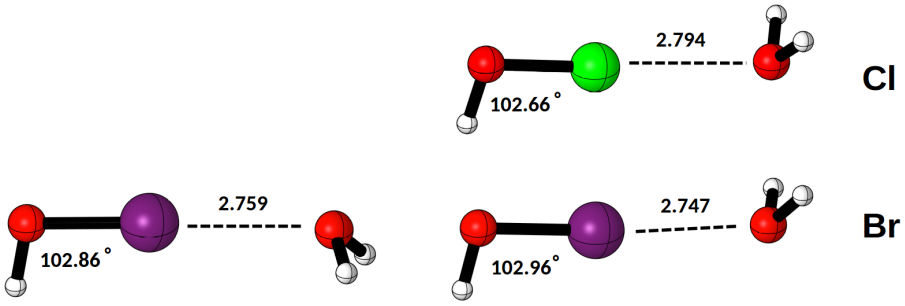

Figure 2.2: CCSD(T)/aug-cc-pVTZ-X2C optimized geometries for the halogen bonded binary complexes. The *syn* conformers are shown on the left and the *anti* conformers on the right. The halogen bond lengths (in Å) and internal H-O-X angles (in degrees) are shown. Atoms are colored green and purple for Cl and Br, respectively.

Optimized geometries were computed for all HOX monomers and water so that accurate interaction energies could be computed. All structures presented have C_S symmetry. We also considered interactions between a hydrogen from water and the oxygen of HOX, but no such structures were found. All geometrical structures reported here were confirmed to be minima on the potential energy surface by computing harmonic vibrational frequencies (Table 6.1). Cartesian coordinates for all structures are provided in the supplementary

information.

The most important geometric parameters for each hydrogen bonded binary complex are presented in (Figure 2.1). We highlight the distance between the hydrogen atom of the HOX and water as well as the internal bond angle of the HOX atoms within the binary complex structure. These parameters are perhaps the most relevant and interesting with respect to understanding the nature of halogen or hydrogen bonding in these species. For the *syn*-XOH$\cdots$OH$_2$ structures, the hydrogen bond lengths (1.786 Å, 1.796 Åfor F ,Cl), increase slightly going down the halogen series. The *anti*-XOH$\cdots$OH$_2$ structures follow a similar trend (1.799 Å, 1.802 Å, 1.817 Å for F, Cl, Br) Panek and Berski computed MP2 structures with larger hydrogen bond distances of 1.806 Å, 1.808 Å, 1.822 Å,[114] respectively. Santos and coworkers, who optimized geometries with CCSD/6-311++G(2d,2p), also overestimated the *syn*-BrOH$\cdots$OH$_2$ hydrogen bond at 1.858 Å.[117] This especially large difference may be explained by our inclusion of the perturbative triples correction, as well as the use of the X2C method to obtain a scalar relativistic corrected geometry. The effect of halogen substitution on the geometries decreases with increasing halogen size. This is most simply due the decrease in the electronegativity of the halogen atom, which lowers the ability of the hydrogen to accept electron density from water. The internal HOX bond angles follow a much less linear trend. For the *anti* isomers, the fluorine binary complex has an angle of 97.88° and the chlorine and bromine binary complexes have angles 102.71° and 102.74°, respectively. The *syn* isomers follow a similar trend but are approximately 1° smaller than the *anti* isomers. The appendix contains the optimized *anti*-IOH$\cdots$OH$_2$ complex which has an interaction distance (1.832 Å) that continues the trend and is longer than the bromine case and exhibits an IOH angle of 104.49°. Interestingly, the same structure was optimized using an effective core potential at the CCSD(T)/aug-cc-pwCVTZ-PP level of theory without X2C and predicts the same hydrogen bond distance to less than one thousandth of an Å.

The HOX$\cdots$OH$_2$ binary complexes have very different geometric parameters compared to the XOH$\cdots$OH$_2$ binary complexes. The energetically favorable *ani* isomers have halogen bond distances of 2.794 Å for the chlorine structure and 2.747 Å for the bromine structure. This is a reversal of the trend for the HB binary complexes down the halogen series. An explanation is that the larger and more diffuse halogens can be induced to form a positive σ hole that accepts electron density from water.[139] We can only compare the *anti* binary complexes to the one *syn*-HOBr$\cdots$OH$_2$ structure found, but we note a slight elongation of the XB to 2.759 in the *syn* isomer. The internal HOX bond angle increases from 102.7° in the chlorine binary complex to 103.0° in the bromine binary complex. There is a slight decrease in the HOX angle from the bromine *anti* to the *syn* isomer. The appendix contains the optimized *anti*-HOI$\cdots$OH$_2$ complex which has an interaction distance (2.823 Å) that is significantly longer then the bromine case and exhibits an IOH angle of 104.97°. The CCSD(T)/aug-cc-pwCVTZ-PP halogen bond distance is 0.003 Åshorter than the value obtained using

X2C, again confirming the accordance of their predictions.

2.4.2 Interaction Energies

The focal-pointed CCSDT(Q)/CBS interaction energies with additive corrections for the XOH$\cdots$OH$_2$ binary complexes are presented in Table 2.1. The additive ZPVE correction was approximately 1 kcal mol^{-1} for XB binary complexes and 2 kcal mol^{-1} for HB binary complexes. The DBOC and frozen core corrections were much smaller, on the order of 0.01 kcal mol^{-1} for the XOH$\cdots$OH$_2$ structures. The interaction energies for the *syn* isomers were –5.62 kcal mol^{-1} and –5.56 kcal mol^{-1} for X=F and Cl, respectively. The *anti* structures had slightly smaller magnitude interaction energies of –5.37 kcal mol^{-1}, –5.28 kcal mol^{-1}, and –4.97 kcal mol^{-1}, respectively. Therefore the substitution of different halogen atoms has little effect on the strength of the hydrogen bonds. This is consistent with the trend seen in the bond lengths. The significance of the CCSDT correction increased as the size of the halogen increased from 0.01 kcal mol^{-1} for fluorine to 0.03 kcal mol^{-1} for bromine. The CCSDT(Q) correction is minimal for all structures indicating convergence with respect to level of theory.

The interaction energies for the halogen bonded binary complexes are much smaller than for the hydrogen bonded binary complexes (Table 2.2). The *anti*-HOCl$\cdots$OH$_2$ interaction energy is only –1.71 kcal mol^{-1}, which is 3.91 kcal mol^{-1} smaller than the strongest hydrogen bound binary complex. Substituting the Cl atom with a bromine increases the magnitude of the interaction energy by 1.32 kcal mol^{-1} to –3.03 kcal mol^{-1}. Bromine is the only binary complex with minima for both the *syn* and *anti* HOBr$\cdots$OH$_2$ structure. The *anti* isomer has a slightly larger interaction energy of –3.03 kcal mol^{-1} compared to –2.84 kcal mol^{-1} for the *syn* isomer. Halogen substitution also has a significant impact on the ZPVE contribution, increasing it from 1.07 kcal mol^{-1} in the *anti*-HOCl$\cdots$OH$_2$ structure to 1.21 kcal mol^{-1} in the *anti*-HOBr$\cdots$OH$_2$. The increase in the ZPVE is primarily do to a 100 cm^{-1} increase in the internal oxygen-halogen bond stretch mode of the HOX portion of the binary complex. This trend is similar in magnitude but opposite in direction for the HB structures.

Panek and Berski reported electronic interaction energies for the same binary complex structures at the MP2/aug-cc-pVTZ level of theory. For the XOH$\cdots$OH$_2$ binary complexes they report interaction energies of –7.45 kcal mol^{-1} for fluorine, –7.37 kcal mol^{-1} for chlorine, and –7.11 kcal mol^{-1} for bromine. These results are consistent with our quantitative trend for hydrogen bonds, but predict interaction energies of smaller magnitude than our corresponding electronic energies for the **syn** complexes which include the ZPVE. Likewise, they also report interaction energies for the HOX$\cdots$OH$_2$ binary complexes of –2.68 kcal mol^{-1} for chlorine and –4.00 kcal mol^{-1} for bromine. Again, their interaction energies overestimate the magnitude

Table 2.1: Focal-point analysis of the interaction energies of the hydrogen bonded $XOH\cdots OH_2$ structures in kcal mol^{-1}. The CCSD(T)/aug-cc-pVTZ-X2C equilibrium geometries were used for all computations here. Bracketed values indicate extrapolated energies or corrections. δ indicates an incremental change in the energy from the previous level of theory. The CCSDT(Q)/X2C-CBS interaction energies and corrections are shown below in accordance to the formula: $E_{\text{final}} = E_{\text{CCSDT(Q)}-\text{X2C/CBS}} + \Delta E_{\text{ZPVE}} + \Delta E_{\text{FC}} + \Delta E_{\text{DBOC}}$

syn-FOH$\cdots$OH$_2$							
	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
aug-cc-pVDZ	–6.76	–1.52	+0.28	–0.46	+0.01	+0.00	–8.44
aug-cc-pVTZ	–6.39	–1.69	+0.25	–0.43	+0.01	[+0.00]	[–8.25]
aug-cc-pVQZ	–6.28	–1.68	+0.29	–0.43	[+0.01]	[+0.00]	[–8.08]
aug-cc-pV5Z	–6.22	–1.64	+0.31	–0.42	[+0.01]	[+0.00]	[–7.96]
CBS LIMIT	[–6.18]	[–1.60]	[+0.32]	[–0.42]	[+0.01]	[+0.00]	[–7.87]
E_{int} = –7.87+2.17+0.06+0.02 = **–5.62**							
anti-FOH$\cdots$OH$_2$							
	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
aug-cc-pVDZ	–6.53	–1.29	+0.28	–0.37	+0.01	+0.00	–7.89
aug-cc-pVTZ	–6.24	–1.41	+0.25	–0.33	+0.01	[+0.00]	[–7.72]
aug-cc-pVQZ	–6.16	–1.40	+0.29	–0.33	[+0.01]	[+0.00]	[–7.59]
aug-cc-pV5Z	–6.11	–1.36	+0.30	–0.32	[+0.01]	[+0.00]	[–7.48]
CBS LIMIT	[–6.07]	[–1.31]	[+0.31]	[–0.32]	[+0.01]	[+0.00]	[–7.39]
E_{int} =–7.39+1.94+0.06+0.02 = **–5.37**							
syn-ClOH$\cdots$OH$_2$							
	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
aug-cc-pV(D+d)Z	–5.79	–2.27	+0.53	–0.48	+0.03	–0.01	–7.99
aug-cc-pV(T+d)Z	–5.52	–2.54	+0.54	–0.49	+0.03	[–0.01]	[–8.00]
aug-cc-pV(Q+d)Z	–5.43	–2.53	+0.57	–0.49	[+0.03]	[–0.01]	[–7.87]
aug-cc-pV(5+d)Z	–5.39	–2.51	+0.58	–0.49	[+0.03]	[–0.01]	[–7.8]
CBS LIMIT	[–5.38]	[–2.49]	[+0.60]	[–0.49]	[+0.03]	[–0.01]	[–7.74]
E_{int} = –7.74+2.09+0.07+0.02 = **–5.56**							
anti-ClOH$\cdots$OH$_2$							
	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
aug-cc-pV(D+d)Z	–5.73	–1.98	+0.48	–0.41	+0.03	–0.01	–7.63
aug-cc-pV(T+d)Z	–5.54	–2.17	+0.48	–0.40	+0.03	[–0.01]	[–7.61]
aug-cc-pV(Q+d)Z	–5.47	–2.13	+0.50	–0.40	[+0.03]	[–0.01]	[–7.48]
aug-cc-pV(5+d)Z	–5.43	–2.11	+0.52	–0.40	[+0.03]	[–0.01]	[–7.41]
CBS LIMIT	[–5.41]	[–2.09]	[+0.53]	[–0.40]	[+0.03]	[–0.01]	[–7.35]
E_{int} =–7.35+1.97+0.08+0.02 = **–5.28**							
anti-BrOH$\cdots$OH$_2$							
	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
aug-cc-pVDZ	–5.28	–2.17	+0.48	–0.43	+0.03	–0.01	–7.38
aug-cc-pVTZ	–5.09	–2.26	+0.49	–0.41	+0.03	[–0.01]	[–7.25]
aug-cc-pVQZ	–5.03	–2.24	+0.52	–0.41	[+0.03]	[–0.01]	[–7.14]
aug-cc-pV5Z	–4.99	–2.21	+0.52	–0.41	[+0.03]	[–0.01]	[–7.07]
CBS LIMIT	[–4.97]	[–2.18]	[+0.53]	[–0.41]	[+0.03]	[–0.01]	[–7.01]
E_{int} = –7.01+1.95+0.07+0.02 = **–4.97**							

Table 2.2: Focal-point analysis of the interaction energy of the halogen bonded $HOX\cdots OH_2$ structures in kcal mol^{-1}. The CCSD(T)/aug-cc-pVTZ-X2C equilibrium geometries were used for all computations here. Bracketed values indicate extrapolated energies or corrections. δ indicates an incremental change in energy from the previous level of theory. The CCSDT(Q)/X2C-CBS interaction energies and corrections are shown below in accordance to the formula: $E_{\text{final}} = E_{\text{CCSDT(Q)-X2C/CBS}} + \Delta E_{\text{ZPVE}} + \Delta E_{\text{FC}} + \Delta E_{\text{DBOC}}$

anti-$HOCl\cdots OH_2$							
	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
aug-cc-pV(D+d)Z	–1.17	–2.10	+0.52	–0.35	+0.02	–0.02	–3.10
aug-cc-pV(T+d)Z	–0.76	–2.35	+0.56	–0.38	+0.02	[–0.02]	[–2.92]
aug-cc-pV(Q+d)Z	–0.69	–2.36	+0.58	–0.39	[+0.02]	[–0.02]	[–2.86]
aug-cc-pV(5+d)Z	–0.68	–2.35	+0.59	–0.40	[+0.02]	[–0.02]	[–2.84]
CBS LIMIT	[–0.67]	[–2.34]	[+0.59]	[–0.40]	[+0.02]	[–0.02]	[–2.82]
E_{int} = –2.82+1.07+0.04+0.00= **–1.71**							
anti-$HOBr\cdots OH_2$							
	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
aug-cc-pVDZ	–2.50	–2.67	+0.71	–0.41	+0.02	–0.02	–4.88
aug-cc-pVTZ	–1.82	–2.93	+0.75	–0.46	+0.03	[–0.02]	[–4.46]
aug-cc-pVQZ	–1.79	–2.90	+0.75	–0.48	[+0.03]	[–0.02]	[–4.42]
aug-cc-pV5Z	–1.78	–2.87	+0.75	–0.49	[+0.03]	[–0.02]	[–4.39]
CBS LIMIT	[–1.77]	[–2.84]	[+0.74]	[–0.49]	[+0.03]	[–0.02]	[–4.36]
E_{int} = –4.36+1.21+0.14-0.02 = **–3.03**							
syn-$HOBr\cdots OH_2$							
	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
aug-cc-pVDZ	–2.27	–2.62	+0.69	–0.41	+0.02	–0.02	–4.60
aug-cc-pVTZ	–1.60	–2.88	+0.73	–0.46	+0.03	[–0.02]	[–4.20]
aug-cc-pVQZ	–1.57	–2.85	+0.73	–0.47	[+0.03]	[–0.02]	[–4.15]
aug-cc-pV5Z	–1.56	–2.81	+0.73	–0.48	[+0.03]	[–0.02]	[–4.12]
CBS LIMIT	[–1.55]	[–2.78]	[+0.72]	[–0.48]	[+0.03]	[–0.02]	[–4.08]
E_{int} =–4.08+1.10+0.14-0.00 = **–2.84**							

of our focal point energies. Dibble and Francisco[68] report an interaction energy of –5.9 kcal mol^{-1} for *syn*-$ClOH\cdots OH_2$ at the MP4/6-311++G(3df,3pd)//MP2/6-311++G(d,p) level of theory. This energy is in reasonable agreement with our focal point energy of –5.56 kcal mol^{-1} but a bit larger in magnitude. For the bromine binary complexes, Santos and coworkers[117] report interaction energies of –4.2 kcal mol^{-1} for *anti*-$BrOH\cdots OH_2$ and –2.1 kcal mol^{-1} for *anti*-$HOBr\cdots OH_2$ at the CCSD(T)/6-311++G(3df,3dp//MP2/6-311++G(2d,2p)) level of theory (including a basis set superposition correction). The interaction energies from Santos are both lower in magnitude than the focal point predictions we present here with the halogen bond significantly showing a deviation of nearly 1 kcal mol^{-1}, indicating the necessity of our focal point analysis.

The iodine complexes were not analyzed using FPA due to their cost and the availability of recontracted X2C basis sets for 5Z cardinality, however, CCSD(T)/aug-cc-pCVTZ-X2C interactions were obtained. The $IOH\cdots OH_2$ complex exhibits a ZPVE corrected interaction energy of –5.05 kcal mol^{-1} while the $HOI\cdots OH_2$ is slightly weaker at –4.38 kcal mol^{-1}, with the lack of additive corrections and FPA likely resulting in slight overestimates of the interaction energies. It is worth noting that both the XB and HB interaction energy are in accordance with the trends discussed thus far. We predict a gap of only 0.67 kcal mol^{-1} between the HB and XB iodine complexes and therefore expect both to be important for any associated chemical

processes. The corresponding CCSD(T)/aug-cc-pwCVTZ-PP interaction energy predictions were within 0.06 kcal mol^{-1} of the CCSD(T)/aug-cc-pCVTZ-X2C results again demonstrating that the use of an ECP or X2C produces virtually identical results for this system, as Iodine should be an upper bound for the extent of relativistic effects in the hypohalous acid series.

2.4.3 Natural Bond Orbital Analysis

NBO results are presented in Figure 2.3 and show the dominant orbital pair interactions. The primary interacting orbitals are similar for the *syn* and *anti* isomers in all structures; thus only the lowest energy isomer for each hydrogen and halogen bonded binary complex are shown. For each binary complex, the lone pair electrons from the oxygen of water donate into acceptor orbitals on the HOX monomer. The NBO scheme[34] produces an interaction energy (designated $E^{(2)}$ here) which is a ratio of the Fock matrix element between orbitals i and j and the difference between orbital energies $(\epsilon_j - \epsilon_i)$. Thus a large $E^{(2)}$ energy corresponds to a large interaction between orbitals i and j. Figure 2.3 also presents the $E^{(2)}$ values between the donating and accepting orbitals. For the HB binary complexes, the interaction is dominated by the donation of the lone pair on the oxygen atom in water into a sigma bonding orbital in the HOX monomer. The energy of this interaction for the most favorable isomer is 18.6 kcal mol^{-1} for fluorine, 17.8 kcal mol^{-1} for chlorine, and 15.3 kcal mol^{-1} for bromine.

For the halogen bonded binary complexes, the primary orbital interaction is again the donation of the oxygen lone pair of water into a sigma bonding orbital of the HOX monomer, for which chlorine has a value of 3.75 kcal mol^{-1} and bromine 6.92 kcal mol^{-1}. This is consistent with the work of Oliveira and Kraka, which found for various halogen bonded systems that the primary NBO overlap is from the lone pair of an electron donor into the bonding orbital of the halogen atom and some electron withdrawing substituent at the CCSD(T)/aug-cc-pVTZ level of theory.[83] The increasing trend of the $E^{(2)}$ energy with increasing halogen size is the opposite of the trend observed with halogen bonded structures. Halogen bonds are thought to form via the donation of electron density into a positive σ hole unoccupied orbital on the halogen[139], as seen in Figure 2.3. Based on our results, halogens of increasing size form larger σ holes that increase electron accepting ability. Halogen substitution has a much larger effect for the HOX$\cdots$OH$_2$ than the XOH$\cdots$OH$_2$ structures. For XB and HB binary complexes, respectively, the $\epsilon_j - \epsilon_i$ component of $E^{(2)}$ is essentially the same, and the Fock energy component is the primary contributor to changes in $E^{(2)}$, especially for the XB binary complexes.

Research by Zhang and coworkers[115] agrees qualitatively with our general trend that hydrogen bonds have greater $E^{(2)}$ energies than the the orbital overlap of halogen bonds. Zhang reports the primary orbital

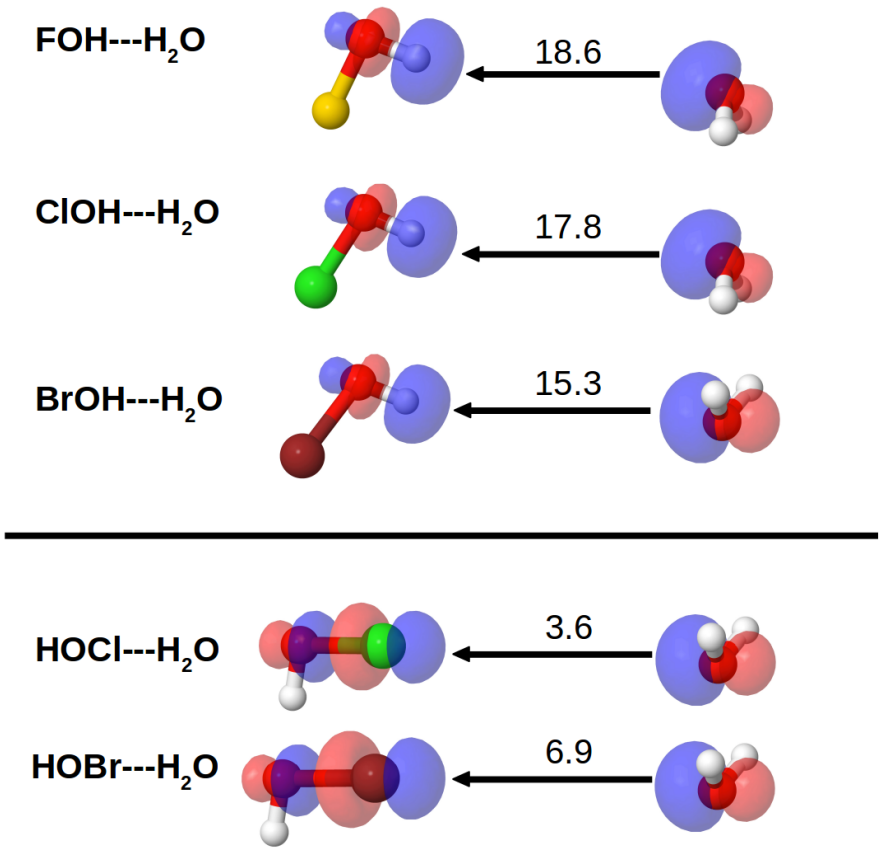

Figure 2.3: Primary natural bond orbital overlap interactions for the most favorable conformer of each binary complex. The three hydrogen bonded structures appear first, followed by the halogen bonded structures. The black arrow signifies the donor acceptor relationship with the $E^{(2)}$ energy printed above in kcal mol^{-1}. The colors yellow, green, and purple represent F, Cl, and Br atoms, respectively.

interaction energy for the ClOH$\cdots$OH$_2$ binary complex's hydrogen-oxygen sigma bonding orbital overlap as 29.4 kcal mol^{-1}, compared to 10.3 kcal mol^{-1} for the oxygen-chlorine σ bond overlap of the HOCl$\cdots$OH$_2$ binary complex for the same interaction. Other theoretical papers that studied other small organic compounds with hydrogen and halogen bonding saw a decrease in the $E^{(2)}$ energy for hydrogen bonds and an increase for halogen bonds when the halogen atom increases in size.[105,106,110,112,113]

2.4.4 SAPT Analysis

The interaction SAPT2+3 energy decomposition[137,140] for each binary complex type is shown graphically in Figure 2.4. The lower energy *syn* conformers are shown for the HB binary complexes and the *anti* conformers for the XB binary complexes. All SAPT2+3 data is presented in detail in the supplementary information. For all nine binary complexes, the SAPT2+3 interaction energies, despite underestimating focal point interaction energies, were excellent qualitative estimates of the relative focal point interaction

energies. However, the SAPT2+3 interaction energies were usually closer to the true interaction energies than a simpler scheme such as SAPT0 energies.[137,140] To determine the effect of geometry on the SAPT results, we ran our level of SAPT theory with the MP2 structures found by Panek and Berski.[114] The results from the MP2 structures and our CCSD(T) structures show differences of less than one kcal mol^{-1}.

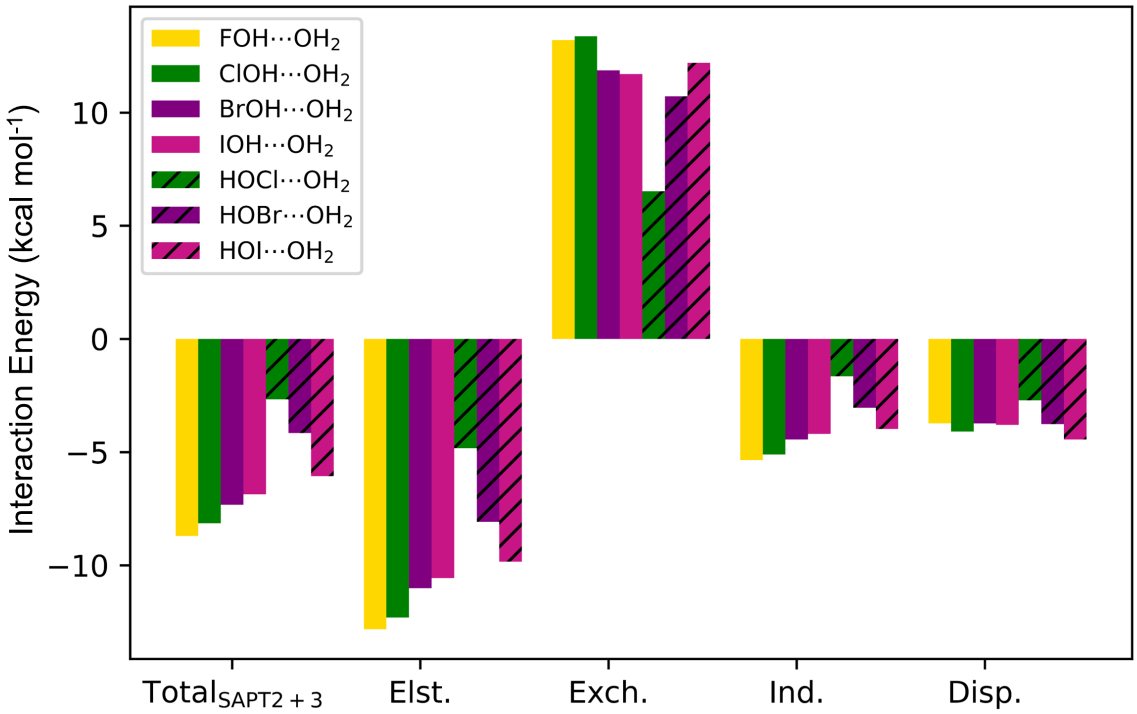

Figure 2.4: SAPT2+3/aug-cc-pVTZ decompositions are shown for the most energetically favorable conformer of each binary complex. The decomposition is broken into favorable electrostatic (Elst), induction (Ind), and dispersion (Disp) terms, as well as unfavorable exchange (Exch) interaction.

It is helpful to examine the trends in the SAPT2+3 components across the periodic trends with increasing halogen size (Figure 2.4). The electrostatic component (labeled "elec") steadily decreases from fluorine to iodine in the HB structures. This correlates with the decreasing electronegativity of the halogen atom. The same effect lowers the coulombic attraction between the donating lone pair and the acceptor hydrogen. The trend is reversed for the XB binary complexes. Iodine forms a "more positive" σ hole than chlorine, increasing the electrostatic attraction from 4.8 kcal mol^{-1} to 9.8 kcal mol^{-1}.[78]

The exchange component (labeled "exch") is perhaps the most interesting, as it does not seem to follow a periodic trend for the $XOH\cdots OH_2$ binary complexes. The exchange energy increases from fluorine to chlorine but then decreases from chlorine to bromine. This might be due to the competition between atom size and spatial orientation with respect to the hydrogen bond. Increasing the size of the halogen atom will increase the orbital overlap of the halogen atom with the lone pair on the donating oxygen. This repulsion is alleviated by increasing the internal HOX bond angle as well as the $HO\cdots X$ bond length, so much so that

bromine is far enough away from the hydrogen bond that it does not contribute as much to the exchange component. In the halogen bond structures, the exchange is greatly dependent on the identity of the halogen. The exchange component nearly doubles from 6.5 kcal mol^{-1} for the chlorine structure to 10.7 kcal mol^{-1} when bromine is substituted.

Induction (ind) and dispersion (disp) contribute far less to the total SAPT2+3 interaction energies but are relevant because the exchange and electrostatic terms are opposite and nearly equal. This cancellation allows for induction and dispersion to impact the final interaction energies significantly.[77,141–143] For the HB binary complexes, induction decreases with substitution of larger halogen atoms. This is due to the decreasing electronegativities of larger halogens, which yield a smaller dipole. The opposite is true for the XB binary complexes where the σ hole on the halogen becomes more easily induced with a less electronegative atom like bromine. As expected, the dispersion term increases for all structures when the size of the halogen increases. It is important to note that in the HB binary complexes, changes in induction are larger than those for dispersion across the halogen series. For XB binary complexes, induction and dispersion work in tandem to increase the interaction energy.

Another way to analyze the SAPT decomposition is to consider the sum of the three favorable interactions (electrostatics, induction and dispersion) and their respective percent contributions. Most noteworthy, is the fact that the electrostatic component composes between 57% and 59% of this sum for all HB complexes, while the percentage decreases to between 52% and 54% for the XB structures, providing further support for the importance of dispersion and induction in the XB complexes.[144] Another noteworthy difference is that the induction percent contribution decreases and the dispersion percent increases in the HB series with increasing halogen size. The opposite is true in the XB case.

2.4.5 Comparison of XB and HB

The structures presented in this study highlight the differences between hydrogen and halogen bonds, as modeled with high level *ab initio* theory. The most obvious result is the that the halogen bonds are generally weaker than the hydrogen bonds in all of our HOX-water binary complexes. Our results highlight two reasons why this is the case:

1. The halogen atoms are much larger compared to the hydrogen atom and therefore will form longer bonds. This larger distance greatly reduces the electrostatic attraction between atoms; this attraction is the primary driver for the bond formation based on the SAPT2+3 scheme[136].

2. The oxygen lone pair on water is better able to donate electron density to a hydrogen than a halogen atom on the HOX molecule. HOX compounds have a partial positive charge on their proton which

easily accepts electron density from an oxygen lone pair. The dipole decreases with larger and less electronegative halogen atoms. The halogen bond is not caused because the halogen atom inherently has a partial positive charge, but because a positive sigma hole is induced on the halogen atom.[78,139] The larger halogens have greater polarizabilities which result in more positive sigma hole to accept the oxygen lone pair.

The dissociation energies of the $XOH\cdots OH_2$ binary complexes are much greater than those of the $HOX\cdots OH_2$ binary complexes, but close enough that both types of interactions may occur. The difference between our weakest halogen bond ($HOCl\cdots OH_2$= –1.71 kcal mol^{-1}) and our strongest hydrogen bond ($FOH\cdots OH_2$= –5.62 kcal mol^{-1}) is 3.91 kcal mol^{-1}. Increasing halogen size lessens the gap between hydrogen and halogen bonds so much so that the difference between the focal point interaction energies of $BrOH\cdots OH_2$ (–4.97 kcal mol^{-1}) and $HOBr\cdots OH_2$ (–3.03 kcal mol^{-1}) is only 1.94 kcal mol^{-1}. This is relevant to atmospheric chemistry where HOBr molecules often bind to water.[42] Our results predicts that both hydrogen and halogen bonds may be present under atmospheric conditions. This may affect the mechanisms that lead to ozone depletion from HOX molecules.

2.5 Conclusion

Our work provides the most reliable theoretical study of hypohalous acid and water binary complexes to date. We have included extensive electron correlation with the CCSD(T) level of theory, used proper $+d$ basis sets for chlorine computations, and included relativistic effects with the SFX2C-1e method. Our structures were found to have different geometric parameters than previous studies, most noteworthy being our shorter intermolecular bond lengths. Interaction energies were extrapolated to the complete basis set limit and CCSDT(Q) level of correlation. The interaction energies decrease in magnitude for the hydrogen bond binary complexes and increase for the halogen bond binary complexes when the size of the substituted halogen atom increases.

NBO and SAPT analyses were performed to compare the nature of each type of intermolecular interaction. NBO analysis displays the trends in electron donation from water to either the H or X of the HOX species. This donation decreases with increasing halogen size for HBs and increased for XBs. SAPT2+3 analysis showed that halogen substitution does impact both the interaction energies and the individual components, especially for halogen bound binary complexes. The largest change was due to the electrostatic interaction term.

Perhaps the most important result from this research is the large difference in interaction energies between a halogen bonded and hydrogen bonded binary complex with the same atoms. For the chlorine and

bromine binary complexes, the difference in interaction energies were around 4 kcal mol^{-1} and 2 kcal mol^{-1}, respectively. The difference shrinks to less than 1 kcal mol^{-1} for iodine at the slightly lower CCSD(T)/aug-cc-pCVTZ-X2C level of theory. This confirms that even though hydrogen bonds are generally stronger than halogen bonds, they can become close enough in energy, with increasing halogen size that both interactions must be considered.[93–97] This may have consequences for atmospheric scientists studying the role of HOBr and HOCl in the depletion of ozone.

CHAPTER 3

THE HIGHLY STRAINED $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) TETRAHEDRANES: THEORETICAL CHARACTERIZATION

M. E. Wolf, E. A. Doty, J. M. Turney, H. F. Schaefer, *The Journal of Physical Chemistry A* **2021**, 125 (12), 2612-2621, Reprinted with permission from [145]. Copyright 2021 American Chemical Society.

3.1 Abstract

Recent experimental research by Cummins and coworkers has established the existence of a tetrahedrane molecule with one CH moiety replaced by phosphorus. We present here the first theoretical studies of the $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) class of molecules. Geometries are obtained at the highly reliable CCSD(T)/aug-cc-pwCVTZ(-PP) level of theory. Harmonic vibrational frequencies are determined and analyzed to confirm the nature of each stationary point and provide helpful findings that may aid in the detection of each species. Most notable is the result that the geometric parameters associated with the $(CH)_3$ moiety in the tetrahedrane exhibit little change under pnictogen substitution, while the Pn-C bonds and C-Pn-C bond angles greatly increase and decrease, respectively. Strain energies are predicted and range from 110.4 kcal mol^{-1} ($N(CH)_3$) to 92.6 kcal mol^{-1} ($Bi(CH)_3$) at the DF-CCSD(T)//B3LYP-D3/aug-cc-pV(T+d)Z(-PP) level of theory. The obtained geometries are further analyzed with Natural Bond Orbital (NBO) methods to understand the bonding and electronic structure of each species. We also provide insight into how different substituents can help make the tetrahedrane structure more energetically favorable due to electron delocalization into substituent antibonding orbitals. The effect of additional delocalization also weakens the Pn-C bond, especially for the heavier pnictogens. This work concludes with a list of considerations that summarizes our key findings and motivates future work aimed at producing novel pnictogen substituted tetrahedrane molecules.

3.2 Introduction

The study of strained molecules continues to be of great interest to both theoretical and synthetic chemists; see for example references .[146–151] Strained molecules come in many varieties but are defined by the IUPAC as "exhibiting an energetic enhancement due to unfavorable bond lengths, bond angles, or dihedral angles relative to an appropriate standard."[152] One way this can manifest is via unusually acute angles at carbon centers which often results in high energy density materials.[153,154] Perhaps one of the most notorious strained molecules is the tetrahedrane isomer of (CH_4).[155] Each carbon atom has a single bond to the other three carbons and a fourth single bond to a hydrogen atom oriented away from the cage resulting in T_d symmetry. Despite the synthesizing difficulties, tetrahedrane is still predicted to be a possible molecule to isolate due to its kinetic stability.[156]

Chemists have made significant progress in characterizing diverse tetrahedrane containing species by means of numerous theoretical studies. A common strategy has been to substitute the carbon atoms with other realistic possibilities such as silicon[157], boron[158], nitrogen[159,160], germanium[161]phosphorus[159,162],

tin[163], or lead.[163] Likewise, experimental progress has also been made using this strategy such as the synthesis of AsP_3 by Cummins and coworkers[164] in 2009 and the isolation of Di-*tert*-butyldiphosphatetrahedrane by Wolf and coworkers just recently.[165] A second far more flexible strategy has been to replace the hydrogens on tetrahedrane with different substituents that may make the tetrahedral geometry more favorable. In 1978, Maier and coworkers isolated the first alkyl substituted tetrahedrane derivative by substituting the hydrogen atoms with *t*-Bu groups.[166] The steric hindrance of these bulky ligands resisted opening of the tetrahedrane core in what was coined the "corset effect". This approach has been quite successful since the seminal work of Maier and coworkers as tetrahedrane derivatives have been isolated featuring substituents such as: trimethylsilyls[167], sulfonyls[168], and perfluoroaryls.[169] These impressive synthetic accomplishments have progressed in tandem with continued theoretical study of tetrahedranes with various substituents.[158,159,161,170–173]

The phosphorus substituted versions of tetrahedrane have been a particular research thrust. In 2019, Wolf and coworkers synthesized the tetrahedrane containing $P_2(Ct\text{-Bu})_2$ molecule followed shortly by the work of Cummins and coworkers in 2020 who synthesized the $P(Ct\text{-Bu})_3$ tetrahedrane molecule.[153] Cummins and coworkers provide two possible explanations for why this structure could be isolated. First, they argued that the *t*-Bu groups form an attractive network of 9 hydrogen-hydrogen interactions that "forms the basis for the corset effect."[174] Second, they briefly mentioned that electron density on the carbon atoms could be delocalized into the substituent orbitals and produce a more energetically favorable structure. Just as theoretical work has reinforced the claim that tetrahedrane is a reasonable synthetic target[156] due to its significant kinetic stability, phosphatetrahedrane has also been considered as a potential synthetic target.[175] However, this would be a challenging task as Ivanov and coworkers predicted that the phosphatetrahedrane isomer is significantly higher in energy than the global minimum (i.e. vinylacetylene with the H-C≡C- group replaced by a P≡C- group.) by 34.6 kcal mol^{-1} at the CCSD(T)/CBS level of theory.[175] Related work by Rayne and Forrest predicted that the heat of formation for phosphatetrahedrane is 28 kcal mol^{-1} smaller than the heat of formation of tetrahedrane computed at the G4MP2/G4 level of theory[159] which implies that it might be a thermodynamically reasonable target. The transition state barriers computed by Schreiner and coworkers predicting that tetrahedrane is kinetically stable, implies that phosphatetrahedrane might be as well if it undergoes similar decomposition mechanisms.[156] Despite this, there have apparently been no successful attempts to synthesize the unsubstituted phosphatetrahedrane molecule.

With the synthesis of phosphatetrahedrane a plausible possibility in the near future, it would also be of great benefit for chemists to understand the periodic trends of the entire $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) series of molecules. It is well known that the standard form of phosphorus, though certainly not the most thermodynamically favorable, is a tetrahedron and recent theoretical work has also characterized

the Bi_4 tetrahedral structure.[176] Therefore, there is precedent for small tetrahedral species containing both light and heavy pnictogens. Cummins and coworkers have made previous progress along these lines in 2009 when they proposed a simple synthesis of AsP_3 in a tetrahedral configuration.[164] Following this recent research, it is reasonable to consider the possibility of isolating any of the $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) molecules. In this work, we optimize the geometries each of the $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) species and determine whether they exist as minima on their reliable CCSD(T) potential energy surfaces. We further the understanding of this series by characterizing how pnictogen substitution affects the geometry, strain energies, natural charges, dipole moments, harmonic vibration frequencies, and electronic structure of the $Pn(CH)_3$ molecule. Additionally, we use density functional theory and density fitted Møller–Plesset second order perturbation theory (DF-MP2) to analyze the effects of substituting various groups in place of each hydrogen. We further examine the claims made by Cummins and coworkers of how the substituents influence the electronic structure of the tetrahedrane center. Natural Bond Orbital (NBO) analysis plays a central role in this discussion and allows us to characterize the electronic effects that are manifest upon substituent changes. Our results provide a series of key observations that could aid in understanding features of novel $Pn(CH)_3$ species and hopefully motivate future work on this system.

3.3 Theoretical Methods

All geometries were optimized via coupled cluster theory with single, double, and perturbative triples excitations [CCSD(T)][10,119–121], correlating all electrons, in the software package CFOUR 2.0.[123] The Dunning-type aug-cc-pwCVTZ basis set was utilized to properly describe the correlation of all electrons[16,177,178] and was downloaded from the Basis Set Exchange.[179–181] The additional augmented functions are necessary for the basis set to have enough flexibility to properly describe such a strained system. Effective core potentials (ECPs) were employed to model the 10, 28, and 60 innermost core electrons of As, Sb, and Bi, respectively, and account for scalar relativistic effects. All computations involving atoms with an ECP utilized the aug-cc-pwCVTZ-PP basis set[16,182,183]. The Hartree-Fock energy, coupled cluster amplitudes, and Lambda equations were all tightly converged to less than 10^{-8}. Harmonic vibrational frequencies were also computed at the same level of theory to confirm that each optimized geometry is properly characterized as a minimum.

The strain energy[152,157,184] for tetrahedrane was estimated by determining the reaction enthalpy for the isodesmic reaction[185] defined in Equation 3.1.[157]

$$(CH)_4 + 6(CH_3)_2 \longrightarrow 4(CH_3)_3CH \tag{3.1}$$

Equation 3.2, based on the method of Tam and coworkers[186], is utilized to determine a qualitative estimate of the strain energy for each $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) species relative to $(CH)_4$, using a homodesmotic reaction.

$$Pn(CH)_3 + (CH_3)_3CH \longrightarrow (CH)_4 + Pn(CH_3)_3 \tag{3.2}$$

Assuming absence of strain in the $(CH_3)_3CH$ and $Pn(CH_3)_3$ species, the enthalpy of this reaction estimates the difference in strain energies between $(CH)_4$ and $Pn(CH)_3$, with a positive value indicating $(CH)_4$ is more strained. The geometry and zero-point vibrational energy (ZPVE) for each species was obtained at the B3LYP-D3/aug-cc-pV(T+d)Z(-PP) level of theory.[177,178,187–189] Density fitted CCSD(T)/aug-cc-pV(T+d)Z(-PP) single point energies were obtained for each geometry, corrected with the ZPVE values, and utilized to obtain the strain energy estimates.

To study the geometric changes upon the substitution of three *t*-Bu groups on the phosphatetrahedrane parent molecule, it was necessary to employ a lower level of theory. A mixed basis set was chosen to balance the cost and accuracy or these larger computations. All four atoms of the tetrahedrane center as well as the most proximal substituent carbon atoms were described with an aug-cc-pV(T+d)Z(-PP) basis set. The rest of the carbon atoms and hydrogen atoms were described with the cc-pVDZ and aug-cc-pVDZ basis sets, respectively. The geometric changes due to the addition of the *t*-Bu groups was predicted by a set of better density functionals which have analytic gradients[190] (B3LYP[189], M05-2X[191], BP86[192], MN15[193], and ωB97X[194]) as well as density fitted Møller–Plesset second-order perturbation theory[195] (DF-MP2) as implemented in Psi4.[196] All other smaller $P(CR)_3$ structures discussed in the substituents section were optimized with DF-MP2 and the aug-cc-pV(T+d) basis set.

Natural Bond Orbital (NBO) analysis was employed to study the electronic structure characteristics of the five pnictogen CCSD(T)/aug-cc-pwCVTZ tetrahedrane structures and their larger substituted derivatives.[34] The NBO7.0 module[35] was interfaced with GAMESS[197] using a def2-TZVP basis set[198,199] along with the B3LYP[189] functional to model the exchange. One of the most important NBO results is the second order perturbation energy ($E^{(2)}$) predicting the importance of different instances of delocalization from Lewis to non-Lewis type orbitals (Equation 3.3).

$$E^{(2)} = q_i \frac{F_{ia}}{\epsilon_a - \epsilon_i} \tag{3.3}$$

The F_{ia} term is the Fock matrix element between Lewis orbital (i) and non-Lewis type orbital (a). The

q and ϵ terms correspond to orbital occupation number and the orbital energy, respectively.

3.4 Results

3.4.1 Geometries

All equilibrium geometries were obtained at the CCSD(T)/aug-cc-pwCVTZ (-PP for elements containing pseudopotentials) level of theory, and the geometric parameters are presented in Figure3.1. All the Cartesian coordinates can be found in the Supplementary Information. Each structure exhibited C_{3v} symmetry with the pnictogen atom centered along the C_3 axis. Also included in Figure 3.1, for reference, is the $(CH)_4$ tetrahedrane molecule which has T_d symmetry. Our results are the first to describe the periodic trends that manifest when the CH group of tetrahedrane is substituted with heavier pnictogens.

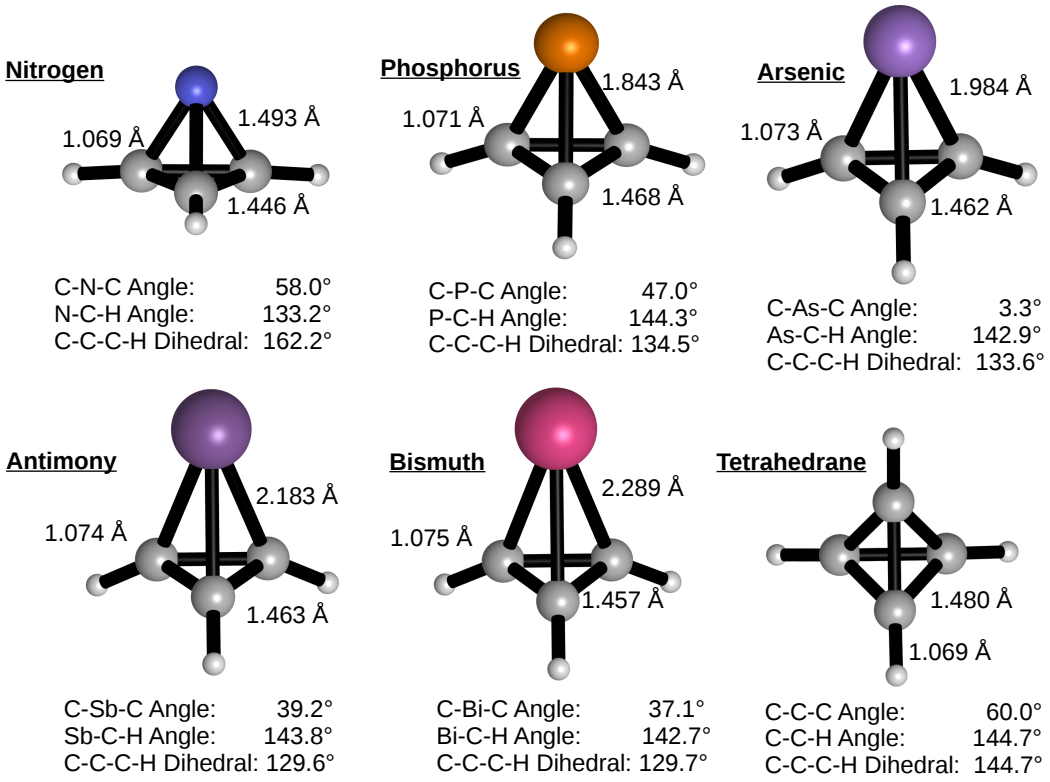

Figure 3.1: Equilibrium geometries optimized at the CCSD(T)/aug-cc-pwCVTZ(-PP) level of theory for $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) and $(CH)_4$. The structures exhibit necessary C_{3v} and T_d symmetries

Pnictogen substitution of tetrahedrane ($Pn(CH)_3$ with Pn = N, P, As, Sb, Bi) results in a set of clear and interesting trends. First, simply from visual inspection, the nitrogen containing $N(CH)_3$ molecule looks different as the hydrogen atoms are close to being in plane with the three carbon atoms. This agrees with the findings of Rayne and Forrest who predicted the geometries of $N(CH)_3$ and $P(CH)_3$ at the G4MP2 composite

level of theory.[159] The small atomic radius and large electronegativity of nitrogen is a partial explanation for why the $N(CH)_3$ deviates from the heavier $Pn(CH)_3$ analogues. Unsurprisingly, the geometric parameters that varied most upon pnictogen substitution were those most associated with the pnictogen atom. The C-Pn bond expands significantly with heavier pnictogen substitution, from a length of 1.493 Å in $N(CH)_3$ to 2.289 Å in $Bi(CH)_3$. (Our computed P-C bond length is 0.014 Å less than the MP2/aug-cc-pVTZ value computed in a recent study by Elguero and coworkers[200] and 0.008 Å less than the G4MP2 bond length predicted by Rayne and Forrest) The increasing Pn-C distances are related to the fact that the heavier pnictogens have much larger atomic radii[201] and diffuse valence orbitals which necessarily require a larger internuclear distance to form a covalent bond with the the carbon atoms. This suggests that the atomic radius of the pnictogen is the most important factor for predicting tetrahedrane geometry even in a strained system. The $(CH)_4$ molecule has a C-C bond length of 1.492 Å which is very close to the N-C bond length in $N(CH)_3$ (1.493 Å) but quite smaller than the the P-C bond length in $P(CH)_3$ (1.843 Å), verifying that the $N(CH)_3$ core geometrically resembles tetrahedrane more than the heavier pnictogen substituted analogues. The $(CH)_3$ moiety is rather invariant under pnictogen substitution as evidenced by the minimal changes in the C-C and C-H bond lengths.

Another insightful geometric parameter is the C-Pn-C angle. In $(CH)_4$, this angle is 60° as required of a molecule with tetrahedral symmetry. Our predicted angle measurements can be used to quantify deviation from a perfect tetrahedron center. For all structures that we optimized, the C-Pn-C angles decreased as pnictogen size increased. The C-N-C angle was again the closest to a true tetrahedron at 58.0°. The C-Pn-C angle decreases by 11.0° to 47.0° when the nitrogen is substituted by a phosphorus atom and continues to decrease to 37.0° for the C-Bi-C angle. The large change in the C-Pn-C angle reinforces the fact that the $Pn(CH)_3$ is primarily perturbed by the position of the pnictogen atom rather than a significant expansion and contraction of the three carbon atoms.

One aspect of the $(CH)_3$ moiety that does depend on the identity of the fourth vertex of the tetrahedrane center is how the hydrogen atoms bend out of the C-C-C plane. The best measure of this is the angle formed between the midpoint of a C-C bond, the opposite carbon atom, and the corresponding attached hydrogen. An angle of (0°) would correspond to the hydrogen atoms perfectly in plane with the carbon atoms. Our geometries predict that the hydrogens deviate from the plane by only 9.1° for $N(CH)_3$ which significantly increases for the rest of the pnictogen series (26.9°, 27.7°, 31.0°, and 31.1°, for Pn = P-Bi, respectively). For tetrahedrane, this angle is 19.5, indicating a character between the $N(CH)_3$ and $P(CH)_3$ structures. This trend is in agreement with the work of Schleyer and Jemmis as well as Hoffmann and Elian and coworkers who determined that there is a correlation between the diffuse nature of the "cap" moiety on an aromatic system and the degree to which the C-H bonds bend out of the plane of the aromatic molecule.[202,203]

3.4.2 Dipole Moments and Natural Charges

The CCSD(T)/aug-cc-pwCVTZ dipole moment of $N(CH)_3$ is 1.95 Debye oriented along the C_3 axis towards the more electronegative nitrogen. The decreasing electronegativity of the heavier pnictogen atoms in conjunction with their increased distance from the carbon atoms causes the dipole moment to invert and orient away from the pnictogen atom with decreasing magnitudes of 1.25, 1.06, 0.73, and 0.25 Debye for P, As, Sb, and Bi, respectively. The NBO natural atomic charges were computed and also help elucidate this dipole moment trend. For each structure, the carbon atoms carry a partial negative charge. In the case of $N(CH)_3$, the carbons exhibit only a slight negative charge of –0.14 but this greatly increases to between –0.43 and –0.39 for the heavier pnictogens. The nitrogen atom in $N(CH)_3$ carries a significantly larger negative charge of –0.31, resulting in a dipole moment oriented towards the pnictogen. The heavier pnictogens have a much smaller electronegativity than carbon which results in them all carrying significant positive charges (0.52, 0.49, 0.51, 0.45 for Pn = P, As, Sb, and Bi, respectively). Cummins and coworkers note that the significant negative charge carried by the carbon atoms could produce a more energetically favorable structure if it could be delocalized into the substituents of the $Pn(CH)_3$ molecule, which we discuss in detail in the substituents section.

3.4.3 Strain Energies

The enthalpy of the reaction defined in Equation 3.1 is used to estimate the strain energy of the $(CH)_4$ molecule. The equation is balanced with the reactants and products side each containing the same number of C-C or C-H bonds. Therefore, any nonzero reaction enthalpy is assumed to be primarily a consequence of the strain possessed by tetrahedrane. Obviously, strain is not the only factor influencing the enthalpy of Equation 3.1. Nevertheless, this is reasonable enough to provide an excellent qualitative prediction of the strain possessed by the $(CH)_4$ molecule. Our predicted CCSD(T)//B3LYP-D3/aug-cc-pV(T+d)Z(-PP) enthalpy of the reaction in Equation 3.1 is highly exothermic with a magnitude of 134.8 kcal mol^{-1}. Our predicted strain energy is in reasonable agreement with previous estimates that have been produced at lower levels of theory.[157,204,205]

The model of Tam and coworkers was followed to produce Equation 3.2.[186] We assume that that the carbon reference species, $(CH_3)_3CH$, and the pnictogen reference species, $(CH_3)_3Pn$, posses no strain. Therefore, the enthalpy of the reaction defined in Equation 3.2 will be a qualitative estimate of the difference between the strain energy of $(CH)_4$ and $Pn(CH)_3$, with positive values indicating that $(CH)_4$ possesses more strain. We predict Equation 2 reaction enthalpies of 24.3, 34.5, 38.8, 39.6, and 42.1 kcal mol^{-1} for Pn = N, P, As, Sb, and Bi, respectively. Each of the $Pn(CH)_3$ species is significantly less strained than $(CH)_4$. As the

pnictogen size increases, the strain energy significantly decreases, likely due to the diffuse nature of the larger pnictogens. Increasing pnictogen size beyond arsenic only marginally relieves strain as there is less than a 5 kcal mol^{-1} difference in strain energy between the Bi(CH)$_3$ and As(CH)$_3$ species. Despite this decrease, Bi(CH)$_3$ still exhibits significant strain (92.6 kcal mol^{-1}) and would likely remain a challenging molecule to isolate.

3.4.4 Vibrational Frequencies

Harmonic vibrational frequencies for the pnictogen substituted tetrahedranes were produced at the same level of theory as the geometry optimizations and are presented in Table 3.1. The harmonic frequency data may be used to aid in the detection of these structures.

Table 3.1: The harmonic vibrational frequencies of the Pn(CH)$_3$ molecules at the CCSD(T)/aug-cc-pwCVTZ(-PP) level of theory, in cm^{-1}.

Normal Mode	Physical Description	Nitrogen	Phosphorus	Arsenic	Antimony	Bismuth
$\omega_1(a_1)$	Pn-C Symmetric Stretch	758	666	514	442	395
$\omega_2(a_1)$	C-H Umbrella Motion	1085	799	779	761	739
$\omega_3(a_1)$	C-C Symmetric Stretch	1470	1348	1348	1328	1334
$\omega_4(a_1)$	C-H Symmetric Stretch	3378	3339	3325	3302	3294
$\omega_5(a_2)$	C-C-C Twist	919	940	954	964	969
$\omega_6(e)$	Pn-C Asymmetric Stretch	655	570	516	471	442
$\omega_7(e)$	Pn-C-H Asymmetric Bend	825	746	724	708	687
$\omega_8(e)$	C-C Asymmetric Stretch	881	824	821	813	813
$\omega_9(e)$	C-C-C Scissoring Motion	1146	1074	1071	1067	1072
$\omega_{10}(e)$	C-H Asymmetric Stretch	3344	3314	3300	3278	3269

Vibrational mode $\omega_1(a_1)$ describes the stretching of the Pn-C bonds for the Pn(CH)$_3$ (Pn = P, As, Sb, Bi) species but exhibits significant C-H umbrella-like motion of the C-H bonds for the N(CH)$_3$ species. This mode (ω_1) is quite sensitive to pnictogen substitution with a frequency of 758 cm^{-1} for Pn = N shrinking to a frequency of 395 cm^{-1} for Pn = Bi. ω_1 even changes enough to become the lowest vibrational frequency for Pn = As, Sb, and Bi. $\omega_2(a_1)$ is similar to ω_1 and exhibits umbrella motion of the hydrogens coupled with some Pn motion. The uniform translation of the heavy atoms explains why ω_2 is not nearly as sensitive to pnictogen substitution as ω_1. The N(CH)$_3$ structure is again unique with a ω_2 frequency of 1085 cm^{-1} which drops significantly to between 799 and 739 cm^{-1} for the heavier pnictogen analogues. The remaining a_1 and a_2 modes have frequencies that differ slightly for N(CH)$_3$ and then exhibit almost no sensitivity among the heavier pnictogens.

The rest of the vibrational modes (ω_6–ω_{10}) are all doubly degenerate with e symmetry. ω_6 features contraction of a Pn-C bond while the other two Pn-C bonds expand and remain fixed, respectively. The harmonic frequency for ω_6 decreases from 655 cm^{-1} for $N(CH)_3$ to 442 cm^{-1} for $Bi(CH)_3$ with the largest change occurring between the lighter pnictogens. For the case of Pn = N and P, ω_6 is the lowest frequency vibration but is overtaken by ω_1 for the heavier elements. Similar to the higher frequency a_1 and a_2 modes, the rest of the e modes are composed of little pnictogen atom motion and therefore have little sensitivity to pnictogen substitution.

3.4.5 NBO Analysis

To further understand the characteristics of the $Pn(CH)_3$ molecules, it is helpful to closely examine some of the prominent features of the electronic structure via an NBO analysis. In particular, we present three results of the NBO analyses: Natural Resonance Theory (NRT) weights, bond orders, and the hybridization of the bonding and lone pair orbitals, which provide complementary insights towards a better description of the $Pn(CH)_3$ electronic structure. It is important to emphasize that the purpose of the NBO results is to produce a quantifiable description of the qualitative trends predicted by the sophisticated coupled cluster wavefunction results and connect them to chemical intuition.

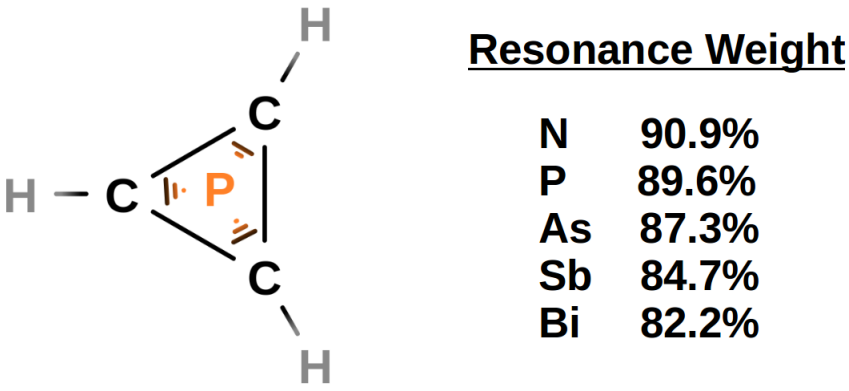

Figure 3.2: Natural Resonance Theory weights for the dominant Lewis structure representations.

The NRT resonance weights indicate how well each $Pn(CH)_3$ structure is represented by a single Lewis structure. The larger the leading resonance weight is, the less electron delocalization occurs in a given system. Each geometry presented is well described by three Pn-C, C-C, and C-H single bonds and a single lone pair on the pnictogen as depicted in Figure 3.2. The resonance weight on this dominant Lewis structure decreases from the nitrogen version (90.9 %) to bismuth (82.2 %), indicating that pnictogen size is positively correlated to delocalization in the $Pn(CH)_3$ molecules. The $Bi(CH)_3$ structure has the smallest leading

resonance weight, but the next largest are a triad of degenerate resonance structures corresponding to the breaking of two of the Bi-C bonds and the formation of a second bond between those two carbons. The weight for each of these is still minimal though at only 2.9 %. The NRT results confirm the validity of describing these molecules with a single Lewis structure but also highlight the increasing contribution from other electronic arrangements, further elucidated by the NRT weighted average bond orders.

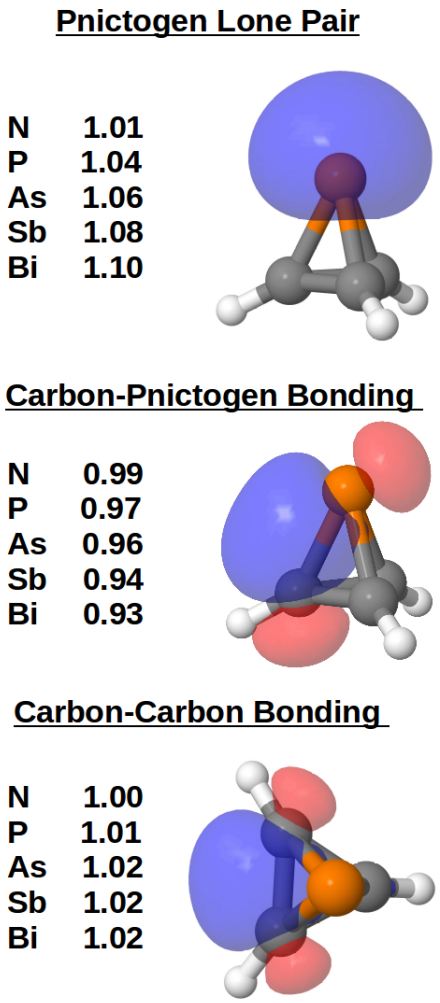

Figure 3.3: NBO orbital bond orders for key orbitals in the $Pn(CH)_3$ structures. The $P(CH)_3$ molecule is used as a representative to visualize the lone pair, C-Pn bonding, and C-C bonding orbitals.

Figure 3.3 presents the NBO bond orders for the pnictogen lone pair, the Pn-C bonding orbital and the C-C bonding orbital, which are all derived from a weighted average of the representative NRT descriptions. A bond order of 1.00 describes a formal single bond or lone pair in a traditional Lewis sense. Cummins and coworkers[153] perform a similar analysis for the $P(CH)_3$ molecule and produce results that are in excellent agreement with our $P(CH)_3$ predictions, indicating that the NBO results for this system seem to be insensitive to the level of theory used to obtain the geometry. Our predictions highlight a noticeable change in the Pn-

C bonding NBO which steadily decreases from N (0.99) to Bi (0.93), indicating that electron density is removed from the Pn-C bond as pnictogen size increases. This result may be an additional reason for the significant increase in the C-Pn distance, up to 2.289 Å in $Bi(CH)_3$. The elongated bond is primarily due to the increasing atomic radius; however the CCSD(T) bond length we predict may also increase due to electron delocalization away from a formal Bi-C single bond. The electron density from the Pn-C bond must go somewhere else and clearly accumulates on the Pn lone pair. The pnictogen lone pair bond orbital population significantly increases from nitrogen (1.01) to bismuth (1.10). The C-C bond order is quite consistent across all structures and only exhibits a slight increase from N (1.00) to Bi (1.02). The larger C-C bond order is likely a small portion of the electron density from the Pn-C bond that did not go to the pnictogen lone pair.

The NBO orbital hybridization also provides insight into the bonding and lone pair orbitals for each tetrahedrane complex (Figure 3.4). First, it is helpful to consider the unsubstituted $(CH)_4$ tetrahedrane. Each of the four C-C bonding orbital is composed of a 50/50 contribution from each carbon atom, as expected. The *s*-orbital character of each C component is 19.8% and the *p*-orbital character is 80.1%. The composition of the Pn-C bonding orbitals is perturbed from a 50/50 mix as pnictogen atoms replace one of the CH moieties.

Figure 3.4 represents the percent breakdown in the hybridization of the Pn-C bonding orbitals as well as the hybridization of the pnictogen lone pair orbital. The N-C bond orbital is mostly composed of the nitrogen hybrid orbitals, which is not the case for the other Pn-C bonding orbitals. The high electronegativity and compact orbitals of nitrogen make this difference unsurprising. For the Pn-C bonding orbitals, there is an obvious trend as the % *s*-orbital character decreases with larger pnictogen size. This is likely due to the need for more *p*-orbital character in the increasingly longer Pn-C bonds. Cummins and coworkers note in their NBO analysis that all of the central P-C bonds are composed of significant *p*-orbital character (especially for the phosphorus hybrids) which is in agreement with our predictions for the entire pnictogen series.[153] They also highlight that the carbon hybrid orbitals directed away from the tetrahedrane center exhibit 40% *s*-orbital character which they connect to Wiberg's discussion[204] on the origin of strain energy in $(CH)_4$. We find the carbon *s*-orbital character directed away from tetrahedrane center remains invariant with increasing pnictogen size. However, our NBO results predict that the pnictogen lone pair orbitals significantly increase in *s*-orbital character with larger, less electronegative pnictogen substitution which is in agreement with Bent's Rule that predicts the tendency of *s*-orbital character to manifest in orbitals directed towards electropositive groups.

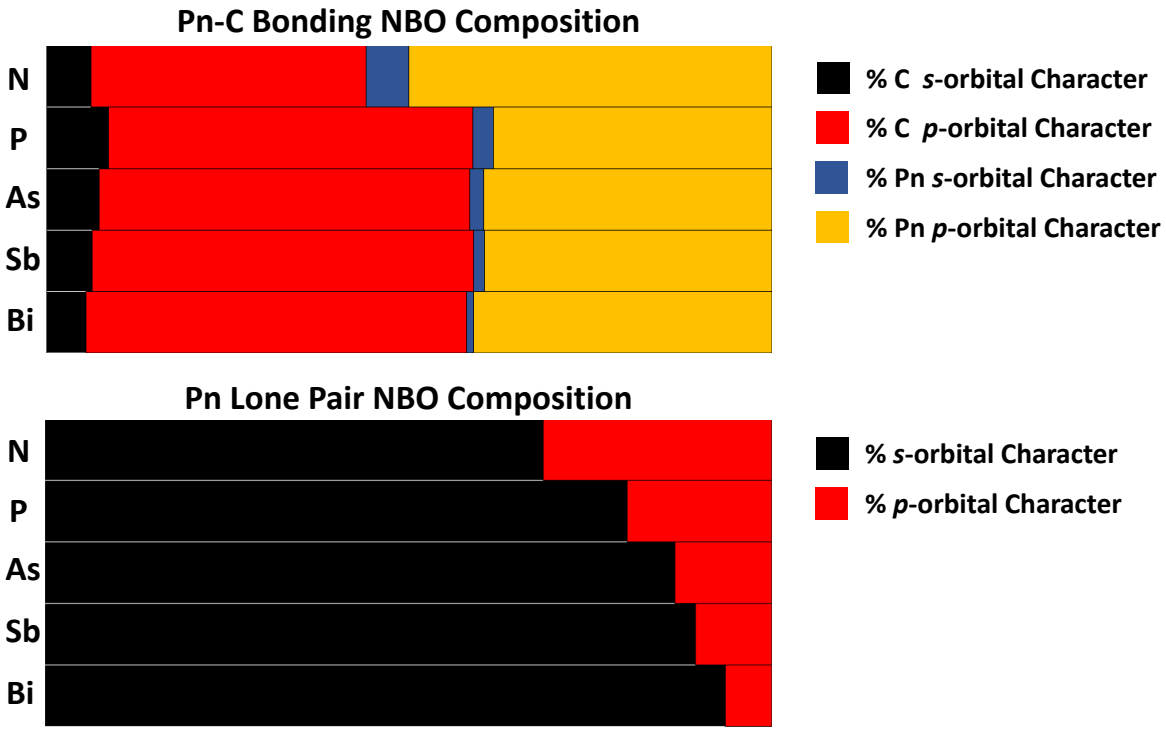

Figure 3.4: The composition of the Pn-C bonding NBO and the Pn lone pair NBO. The Pn-C bonding NBO is divided into contributions from the carbon atom (black and red for s-orbital and p-orbital character, respectively) and contribution from the pnictogen atom (blue and gold for s-orbital and p-orbital character, respectively.) The Pn lone pair NBO is divided into s-orbital (black) and p-orbital (red) character.

3.5 Substituents

The recent work of Cummins and coworkers[153] described their impressive ability to synthesize a long sought after phosphatetrahedrane by replacing the hydrogens with *t*-Bu groups. This success motivates an analysis of how the Pn(C*t*-Bu)$_3$ structures differ from the Pn(CH)$_3$ species and more generally how substituents influence the favorability of these Pn(CH)$_3$ structures. To investigate the geometric effects of the *t*-Bu groups on the parent Pn(CH)$_3$ structures, we obtained the unsubstituted and *t*-Bu substituted structures with DF-MP2 and a variety of density functionals and the results are presented in the SI. For the optimized Pn(C*t*-Bu)$_3$ structures, our predicted P-C bond lengths agree quite well with the P-C distance measured by Cummins and coworkers via X-ray crystallography[153], exhibiting a at most a deviation of 0.03 Å. The C-C bond predictions unsurprisingly follow suit with at most a deviation of 0.02 Å. However, agreement with the C-C-P angles is slightly worse with our predictions overestimating the measured angle with a maximum deviation of about 1°, also implying that our methods underestimate the C-P-C angle. Given the cost limitations to utilizing higher levels of theory and the influence of experimental conditions, our predictions are certainly trustworthy enough to describe the qualitative trends manifest in the Pn(CR)$_3$ species.

The results from each method predict that the N(C*t*-Bu)$_3$ structure changes very little upon *t*-Bu substitution where the heavier pnictogen containing geometries are increasingly sensitive to the *t*-Bu groups. All methods are in agreement that the *t*-Bu groups are associated with a significant increase in the P-C bond length, a significant decrease in the C-Pn-C angle, and little change to the C-C bond lengths. The changes in the P(CH)$_3$ structure following *t*-Bu$_3$ substitution do not provide much new information on their own. However, they do provide an opportunity to understand how the changes in geometry relate to the electronic structure, which might provide more insight into why the P(C*t*-Bu)$_3$ molecule could be isolated but the parent P(CH)$_3$ molecule remains elusive. Cummins and coworkers briefly proposed two features of the P(C*t*-Bu)$_3$ system that could elucidate what makes it a more favorable structure. The first is a network of nine hydrogen-hydrogen attractive interactions between the *t*-Bu groups that "serves as a basis for the so called corset effect". This phenomenon is well described by Fokin and coworkers who estimate the magnitude of the corset effect in the (C*t*-Bu)$_4$ molecule (3.1 kcal mol^{-1} at the B3LYP/6-31G(d,p) level of theory) using an isodesmic reaction.[174]

The second reason noted by Cummins and coworkers is that substituting a CH moiety from tetrahedrane with a phosphorus atom increases the partial negative charge on the carbon atoms.[153] Our previous natural charge analysis agrees with this claim and further predicts a significant negative charge on the carbons of each Pn(CH)$_3$ molecule. They go on to claim that because of this, substituents that are able to delocalize

the negative charge should result in more energetically favorable structures. In order to investigate this claim, further analysis of the delocalization within the unsubstituted $P(CH)_3$ molecule was performed using the NBO second order perturbation energy results.[35] The $E^{(2)}$ results of the $P(CH)_3$ molecule revealed two significant donations from Lewis type orbitals to non-Lewis orbitals. They correspond to delocalization from the C-C bonding orbital to the opposite C-H* antibonding orbital (9.3 kcal mol^{-1}) and donation of the C-P bonding orbital into the two adjacent C-H* antibonding orbitals (5.4 kcal mol^{-1} each), and are both depicted in Figure 3.5. We will denote bonding orbitals as X-X, antibonding orbitals as X-X*, and a donation between a Lewis and non-Lewis orbital as X-X $\rightarrow$ X-X*. All other $E^{(2)}$ values were much smaller, and though they certainly play a role in the complete electronic structure of the $P(CH)_3$ molecule, are dwarfed in importance by these two dominant interactions. To elucidate some of the specific effects of these interactions, the NBO deletion command was used. This option allows the user to delete specific Fock matrix elements, recompute one SCF step with those elements deleted, and then observe the effect on the SCF energy and the new electron density. From this, we determined what features of the electron delocalization are a result of the specific deleted interactions.

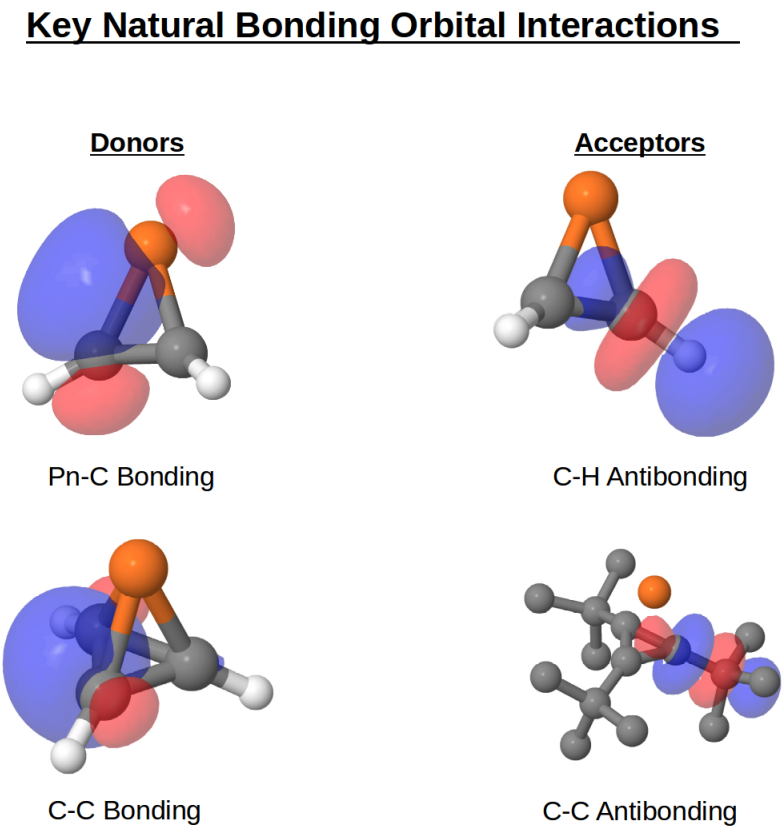

Figure 3.5: Representation of the primary donor and acceptor NBO orbitals in $P(CH)_3$ and $P(Ct\text{-Bu})_3$ molecules. The C-C* visualization excludes hydrogens for clarity.

For $P(CH)_3$, we deleted all three C-C→C-H* and six P-C→C-H* interactions. Blocking these delocalization elements increases the SCF energy by 55.2 kcal mol^{-1} which is not meaningful by itself, but is useful for later comparisons. In the $P(CH)_3$ system, the nine interactions significantly contribute to an electron density decrease in the P-C bonding orbitals but have little effect on the C-C bonds. All other changes in electron density (aside from the obvious decrease in the C-H* orbital occupations) were at least an order of magnitude smaller than the change in the P-C orbital, indicating that the P-C bond strength and distance is most closely related to donation of electron density from the C-P and C-C orbitals into the C-H* orbitals. Similar interactions to these were also noted by Zhou and coworkers[206] in nitro substituted versions of tetrahedrane. They found that the as more nitro groups were added, greater C-C→C-N* donation occurred which drastically increased the C-C bond lengths to the point where they could be easily cleaved and the structure fall apart.

Having established the effect of these delocalizations on the parent molecule, its is prudent to analyze how these features change with different substituents on the $P(CR)_3$ molecule. For the $P(Ct\text{-Bu})_3$ structure, the nine aforementioned interactions are still the dominant Lewis to non-Lewis donations and have larger $E^{(2)}$ values of 15.4 and 6.3 kcal mol^{-1} for C-C→C-C$_{t-\text{Bu}}$* and P-C→C-C$_{t-\text{Bu}}$*, respectively. The shift in the energy upon deletion is significantly higher as well at 71.2 kcal mol^{-1}, indicating that the nine interactions provide about 16 kcal mol^{-1} more favorable interactions to the tetrahedrane structure due to *t*-Bu substitution. Just as in the unsubstituted phosphatetrahedrane, another result of these nine interactions is a slight decrease in the P-C orbital occupations. The $P(Ct\text{-Bu})_3$ geometry has three fold symmetry about the tetrahedron core within the precision of all geometric parameters we present, despite the overall geometry lacking symmetry. There is overall little change to the total C-C orbital occupations which, again, is in agreement with the negligible change in the C-C distances, as predicted by our geometry optimizations. The large change in the deletion energy going from R=H to R=*t*-Bu implies an increased proclivity for delocalization into the substituent orbitals and associates well with the P-C bond lengths significantly increasing and the C-P-C angles significantly decreasing. To further confirm this association, a small set of $P(CR)_3$ molecules (R = H, *t*-Bu, F, Cl, CH_3, CCH, NH_2, CF_3, and CCF) were optimized and the same analysis performed. The SCF deletion energy of the principle delocalizations into the C-R* orbitals varied significantly from the smallest (R = H, 55.2 kcal mol^{-1}) to the largest (R = F, 84.3 kcal mol^{-1}). The trade-off is that this increase in delocalization is associated with weaker P-C bonds, which may facilitate the tetrahedrane center opening, despite the energetic advantage. This is further evidenced by a scatter plot (Figure 6.2 in the SI) between the total deletion energy of all C-C→C-H* and P-C→C-H* interactions and the predicted C-P bond distance. There is a very strong association between the two ($R^2 = 0.96$), providing further evidence that substituent selection has a clear relationship to the P-C bond in the $P(CR)_3$ molecules. This highlights the importance

of selecting substituents which facilitate delocalization into the C-R* orbitals, which is in agreement with the statements made by Cummins and coworkers.[153] However, caution should be had in selecting -R groups that decrease the proclivity for the molecule to form strong enough Pn-C bonds to hold together.

A final insight from this analysis is how delocalization changes with heavier pnictogens substituted into the $Pn(CH)_3$ molecule. The most significant change in the delocalization of the heavier pnictogen containing molecules is increased delocalization into the Pn-C* orbital. A similar deletion analysis was performed on the $Pn(CH)_3$ structures, again deleting any Fock Matrix element connected to the three C-H* orbitals this time with a second computation deleting any matrix elements related to the three Pn-C* orbitals. The deletion energy of delocalization into the C-H* orbitals slightly decreases across the pnictogen series from 49.9 kcal mol^{-1} in $N(CH)_3$ to 43.7 kcal mol^{-1} in $Bi(CH)_3$. The insensitivity of this deletion energy is likely due to two opposing trends. The decreasing proximity of the larger pnictogens to the C-H* orbital lowers the deletion energy, while the diffuse nature of the larger pnictogens makes delocalization more facile. Delocalization into the Pn-C* orbitals is less than 22 kcal mol^{-1} for Pn = N and Pn = P, but increases rapidly to 46.2 kcal mol^{-1} in $Bi(CH)_3$. Thus, for all molecules in the series, there is consistent delocalization out of the Pn-C orbitals but increasing delocalization into the Pn-C* orbitals. This means that caution in substituent selection with respect to weakening the Pn-C bond becomes increasingly important as pnictogen size increases.

3.6 Summary of $Pn(CH)_3$ Theoretical Insights

Our results have affirmed and extended the claim made by Cummins and coworkers that electron delocalization from the tetrahedrane core is an important feature of the $Pn(CH)_3$ system. We also find that the electronic structure effects of the substituents play an important role in determining the favorability of each tetrahedrane molecule in addition to any other features such as the "corset effect". Our reliable CCSD(T)/aug-cc-pwCVTZ(-PP) minima provide a good starting point to suggest that $Pn(CH)_3$ species should at least be considered as synthetic targets in the near future. We combined this foundation with our substituent analysis to develop and reinforce our chemical intuition concerning the $Pn(CH)_3$ class of molecules. The following is a summary of several general considerations from our analyses that characterize the features of the $Pn(CH)_3$ tetrahedrane derivatives.

1. The strain energy of the $Pn(CH)_3$ molecules significantly decreases with larger pnictogen size and all exhibit much smaller strain energies than the $(CH)_4$ parent tetrahedrane. This implies that the more diffuse atoms are able to lessen the energetic penalty of the tetrahedrane configuration. $Bi(CH)_3$ is the least strained species in the series, but still possesses a large strain energy (92.6 kcal mol^{-1}).

2. Significant delocalization occurs within the $P(CR)_3$ structure with the most dominant cases involving the P-C→C-R* and the C-C→C-R* antibonding orbitals. The NBO deletion computations revealed that these interactions disproportionately pull electron density from the P-C bond orbitals and have little effect on the C-C bonds.

3. Replacing the substituents can significantly increase delocalization into the C-R* orbitals and the magnitude of which is strongly associated with the P-C bond length. Thus, varying the substituents can provide greater opportunity for electron delocalization, but caution must be taken such that too much electron density is not removed from the P-C bonds and the tetrahedrane center is able to open up.

4. As pnictogen size increases, delocalization into the C-R* orbitals is minimally affected, but there is a significant increase in electron delocalization into the Pn-C* orbitals. This is associated with lower Pn-C bond orders and allows for less liberty in selecting substituents that might further weaken the Pn-C bonds for heavier pnictogens.

We acknowledge that these considerations make no comment on the kinetic or thermodynamic stability of each structure, which are some of the primary challenges that chemists must overcome to isolate these elusive molecules. The tetrahedrane configuration is not close to the global minimum so any formation pathway would need to energetically bias away from it[175], possibly limiting the synthetic strategies that can be employed. Nevertheless, our work provides a firm theoretical foundation for any future study of related $Pn(CR)_3$ species and hopefully motivates the importance of further research on this system.

3.7 Conclusion

This research was motivated by the remarkable recent (2020) synthesis of $P(Ct\text{-}Bu)_3$ and presents the most reliable and complete theoretical study of the $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) molecules to date. We report geometries and harmonic vibrational frequencies at the CCSD(T)/aug-cc-pwCVTZ(-PP) level of theory, confirming the existence of a theoretical tetrahedrane minimum for each pnictogen considered. The optimized geometries predict that the C-C bond lengths are rather invariant to pnictogen substitution while the Pn-C lengths and C-Pn-C angles significantly increase and decrease, respectively, with larger pnictogen size. Predicted strain energies indicate that all $Pn(CH)_3$ species are significantly less strained than tetrahedrane. The strain energy predictions significantly decrease from 110.4 kcal mol^{-1} in $N(CH)_3$ to 92.6 kcal mol^{-1} in $Bi(CH)_3$. Extensive NBO and NRT analyses also elucidated important features of the electronic structure of the $Pn(CH)_3$ molecules and their trends across the pnictogen series.

Further investigations were performed to understand the effects of *t*-Bu substitution on each structure. Our computations provide evidence that the *t*-Bu substitution increases the Pn-C bond length, has little effect on the C-C length, and decreases the C-Pn-C angle. Considering a wider array of substituents and building on the observations of Cummins and coworkers[153], we predict significant delocalization of the negative charge on the carbon atoms in the $Pn(CH)_3$ molecules into the C-R* substituent orbitals. The NBO deletion analysis provides insight into how the delocalization changes upon changing substituents and with substitution of heavier pnictogen atoms. There is a strong association between delocalization into the C-R* orbital and the length of the Pn-C bond. Larger pnictogens also weaken the Pn-C bond by donating electron density into the Pn-C* orbitals. Our results provide insights into the $Pn(CH)_3$ class of molecules and will hopefully motivate and guide future work on this interesting and elusive subset of tetrahedrane molecules.

CHAPTER 4

CATALYZED REACTION OF ISOCYANATES (RNCO) WITH WATER

M. E. Wolf, J. E. Vandezande, H. F. Schaefer, *Physical Chemistry Chemical Physics* **2021**, 23, 18535-18546, Reproduced from[207] with permission from the PCCP Owner Societies.

4.1 Abstract

he reactions between substituted isocyanates (RNCO) and other small molecules (e.g. water, alcohols, and amines) are of significant industrial importance, particularly for the development of novel polyurethanes and other useful polymers. We present very high-level *ab initio* computations on the HNCO + H_2O reaction, with results targeting the CCSDT(Q)/CBS//CCSD(T)/cc-pVQZ level of theory. Our results affirm that hydrolysis can occur across both the N=C and C=O bonds of HNCO via concerted mechanisms to form carbamate or imidic acid with ΔH_{0K} barrier heights of 38.5 and 47.5 kcal mol^{-1}. A total of 24 substituted RNCO + H_2O reactions were studied. Geometries obtained with a composite method and refined with CCSD(T)/CBS single point energies determine that substituted RNCO species have a significant influence on these barrier heights, with an extreme case like fluorine lowering both barriers by close to 20 kcal mol^{-1} and most common alkyl substituents lowering both by approximately 4 kcal mol^{-1}. Natural Bond Orbital (NBO) analysis provides evidence that the predicted barrier heights are strongly associated with the occupation of the in-plane C−O* orbital of the RNCO reactant. Key autocatalytic mechanisms are considered in the presence of excess water and RNCO species. Additional waters (one or two) are predicted to lower both barriers significantly at the CCSD(T)/aug-cc-pV(T+d)Z level of theory with strongly electron withdrawing RNCO substituents also increasing these effects, similar to the uncatalyzed case. The 298 K Gibbs energies are only marginally lowered by a second catalyst water molecule, indicating that the decreasing ΔH_{0K} barriers are offset by loss of translational entropy with more than one catalyst water. Two-step 2 RNCO + H_2O mechanisms are characterized for the formation of carbamate and imidic acid. The second step of these two pathways exhibits the largest barrier and presents no clear pattern with respect to substituent choice. Our results indicate that an additional RNCO molecule might catalyze imidic acid formation but have less influence on the efficiency of carbamate formation. We expect that these results lay a firm foundation for the experimental study of substituted isocyanates and their relationship to the energetic pathways of related systems.

4.2 Introduction

In recent years, there has been significant experimental research progress on substituted isocyanate molecules including substituents such as, -CN[208], -$ClSO_2$[209], -CCH[210,211], -CH_3[212,213], -Ph[214,215], -CH_2CH_2Cl[216], -*t*-Bu[217], and many more.[218–221] Isocyanic acid (HNCO) is the lowest energy isomer of the HNCO system and the simplest isocyanate[222,223], an important class of molecules for many diverse chemical contexts. HNCO is an atmospheric pollutant formed via the burning of fossil fuels, cigarette use, and other

secondary sources.[224–228] Many combustion processes are also related to the HNCO species and its derivatives.[18,214,229,230] Isocyanate molecules are also important interstellar species[211–213,231] that are implicated as critical intermediates for prebiotic precursors such as formamide[232–234], thyamine[235], and urea.[236] One of the reasons isocyanates are so ubiquitous is their diverse and facile reactivity with many organic molecules[218,224,237–239,239], thus motivating further research into understanding the important features of reactions involving isocyanates.

Many industrial applications such as coatings, adhesives, sealants, elastomers, and insulation rely on reactions involving isocyanates to form useful polymer networks.[208,240–248] A key reaction involved in polymer formation is alcoholysis via the HNCO species to produce carbamates[248,249]. The rate of this reaction can be significantly modified in many ways by changing substituents on the RNCO or ROH reactants[250–252], autocatalysis of the reactants[253–258], solvent selection[256], or via other more efficiently designed catalysts[259–262] such as organotin.[263,264]

It is well established that HNCO will react with water across either the N=C or C=O bonds, producing carbamate and imidic acid, respectively.[255] Raspoet and coworkers predicted with QCISD(T)/6-31G** that reaction across the N=C bond to form carbamate is favored with a reaction barrier of 37 kcal mol^{-1}, 9 kcal mol^{-1} lower than the barrier to reaction across the C=O bond. Perhaps the most rigorous determination of the barrier to carbamate formation is from Nicolle and coworkers in 2016 who predicted a barrier height of 37 kcal mol^{-1}, relative to the prereactive complex, at the M06-2X/6-311++G(d,p) level of theory.[254] Both of these studies also note that additional catalytic water molecules significantly reduce the enthalpy of reaction across both bonds.

Additional theoretical research has helped further understanding of substituted isocyanate species. In 1994, McAllister and Tidwell performed comprehensive structural and isodesmic reaction analyses on a large selection of substituted isocyanates, with the goal of theoretically motivating the discovery of novel isocyanate derivatives.[265] Their work highlights many insightful features of a diverse range of RNCO species, but is not connected to any particular reaction mechanisms. In the past decade, a series of studies[251,266–273] by Konovalov and coworkers predicted the quantum mechanical features of the $R_1NCO + R_2OH$ (R_1, R_2=Ph, CH_3, H) reactions in various conditions, including solvent effects and cooperative catalysis with the B3LYP/6-311++G(df,p) method. However, their small sample size precludes a deep understanding of how substituents generally influence the reaction energetics. Our preliminary computations also indicate water + HNCO reaction energies differs by more than 1 kcal mol^{-1} between the CCSD(T)/cc-pVQZ and B3LYP/cc-pVQZ levels of theory computed on the same geometries. This discrepancy between DFT and CCSD(T) in the simplest of cases exemplifies the need for rigorous *ab initio* methods to draw meaningful conclusions from theoretical study of more complex $H_2O + RNCO$ reactions.

In 2013 Wagner and coworkers predicted the barrier heights to the formation of carbamates given a small

variety of substituted isocyanates (RNCO, R=CH_3, CH_2F, C_6H_5, SiH_3, SiH_3CH_2) with the B3LYP/6-31G* method.[250] They determined that greater electron withdrawing substituents on the RNCO molecule lowered the barrier to carbamate formation. However, Wagner and coworkers only consider the reaction across the N=C double bond and exclude any mention of how substituents influence reaction across the C=O double bond. They also neglect to explain why electron withdrawing effects influence the reaction barriers. The results of Wagner and coworkers also demonstrate significant discrepancies based on predictions made at the MP2/6-31G* and B3LYP/6-311++G(d,p) levels of theory employed in their study (e.g. a 5 kcal mol^{-1} disagreement for the $CH_3NCO + CH_3OH$ transition state). More recently, Zhao and Suppes predicted the enthalpy of reaction for various isocyanates reacting with increasingly larger alcohols to form carbamates with the B3LYP/6-31G(d,p) method[274]. Their work elucidates trends concerning an array of aryl isocyanates and predicts that increasing the size of the reacting alcohol has little influence on the enthalpy of the urethane product relative to the reactants. In addition to these works, there are plenty of examples in the literature of specific $R_1NCO + R_2OH$ reactions studied at various levels of theory.[249,256,257,275,276] All of this previous research emphasizes the importance of clearly understanding how substituted RNCO species influence the electronic structure and energetic landscape of the $RNCO + H_2O$ reactions. Despite its significant importance to industrial chemistry, the literature lacks a comprehensive and reliable theoretical benchmark for this system and a detailed analysis of the relationship between the isocyanate substituents and the electronic structure features of these reactions.

Our research builds on this body of research and improves previous characterization of the $RNCO + H_2O$ reactions in four important ways. First, we thoroughly consider the H_2O reaction across the isocyanate C=O bond which is often ignored and assumed to be high enough in energy to be unimportant. Even if that is generally the case, it would be helpful to test this assumption for many substituents and understand what conditions might result in exceptions. It is a possibility that imidic acid formation could be favored over carbamate depending on the energetic barriers of each process. Second, our work rigorously characterizes the energetic landscape of the parent $HNCO + H_2O$ reaction and uses these results to benchmark the levels of theory used to analyze the substituent trends. Previous studies used differing theoretical methods, making it difficult to compare results across the literature and determine consistent trends. Our reliable *ab initio* results will lay a firm and consistent foundation for any future work on isocyanate alcoholysis reactions. Third, we systematically study how substituents influence the electronic structure of important cooperative catalytic pathways (i.e. multiple water or RNCO molecules) that previous research[256–258] has suggested as very important for these systems. Finally, our sophisticated *ab initio* predictions are analyzed with Natural Bond Order(NBO) analyses to provide detailed understanding of the electronic structure manifest in each process in order to characterize relationships that could generalize to larger or more complex isocyanate

reactions.

The RNCO + H_2O reactions are characterized considering substituents grouped into two categories: the first group is a fundamental collection of carbon groups, pnictogens, chalcogens, and halogens that demonstrate periodic trends (R=CH_3, SiH_3, GeH_3; NH_2, PH_2, AsH_2; OH, SH, SeH; F, Cl, Br) while (R=CH_2CH_3, $CH(CH_3)_2$, $CH_2CH_3CH_3$, $C(CH_3)_3$, C_6H_5, $CHCH_2$, COH,CF_3, COOH, SO_2Cl, and CN) consists of larger substituents that are more relevant to industrial applications.[210,214,215] Three important autocatalyitc mechanisms (RNCO + 2 H_2O, RNCO + 3 H_2O, and H_2O + 2 RNCO) are characterized considering a smaller subset of the aforementioned substituents. The influence of substituents on the efficiency of these catalyzed reactions is systematically studied for the first time. Our work should motivate and ground future study of these and related systems and, our findings can be extended in conjunction with other advanced analyses (e.g. kinetic models, solvent effects, etc.).

4.3 Theoretical Methods

The geometries for the benchmark H_2O + HNCO reaction are obtained using the CCSD(T)[10,119–121] method via the CFOUR 2.0[123] software package. The Dunning cc-pVQZ basis set is utilized for all stationary points except for the van der Waals complexes which are optimized with the aug-cc-pVTZ basis set to properly describe long range interactions.[188,277,278] Harmonic vibrational frequencies are obtained for each stationary point using the same level of theory as the geometry optimization and verify each geometry as a minimum or first-order saddle point (transition state). The connectivity of each stationary point is verified by performing qualitative intrinsic reaction coordinate (IRC) scans. Relative energies of each stationary point are further refined utilizing the focal point method of Allen and coworkers.[18–21] The SCF/cc-pVXZ (X=Q,5,6) energy is extrapolated using the three point extrapolation formula of Feller[22] ($E_{\mathrm{HF}}(X) = E_{\mathrm{HF}}^{\infty} + ae^{-bX}$) and the correlation energy up to CCSD(T)/cc-pVXZ (X=Q,5) is extrapolated using the two point formula of Helgaker[279] ($E_{\mathrm{corr}}(X) = E_{\mathrm{corr}}^{\infty} + aX^{-3}$) towards the complete basis set (CBS) limit .

A series of additive corrections are obtained to further refine the energy predictions and justify the approximations used in the geometry optimization. The zero-point vibrational energies (ZPVE) are obtained from the harmonic vibrational frequency computations producing 0 Kelvin enthalpies (H_{0K}). Higher order CCSDT and CCSDT(Q) additive corrections[280–283] are computed with the cc-pVDZ basis set to capture as much electron correlation as possible and demonstrate convergence towards the FCI limit. A frozen core correction (Δ_{FC}) is obtained using the cc-pCVQZ basis set[12] to account for the difference in the CCSD(T) energy correlating all electrons and the CCSD(T) energy freezing the core. To account for scalar relativistic effects, the spin-free exact two-component one-electron (SF-X2c-1e) method along with the decontracted cc-

pCVTZ basis set is used to compute a scalar relativistic correction (Δ_{rel}).[126,284–288] The correction accounts for the difference in energy with and without the (SF-X2c-1e) method turned on, while correlating all electrons. Finally, a diagonal Born-Oppenheimer correction (Δ_{DBOC}) is included as a diagnostic at the Hartree–Fock/cc-pVQZ level of theory to ensure that each stationary point is not influenced by any nearby conical intersection.[131,132] For the van der Waals complexes, corrections were computed with the appropriate augmented basis functions, which are detailed in Table S14. All corrections are obtained using the following software packages: CFOUR 2.0[123], Molpro 2010[289], and Psi4.[196]

Geometries and harmonic frequencies for all species in the H_2O + RNCO reactions are obtained using the MP2[TZ,QZ] + ΔCCSD(T)/DZ composite method recently highlighted[27] by Sherrill and coworkers and implemented in Psi4. The MP2[TZ,QZ] term refers to a density fitted second-order Møller–Plesset (MP2) gradient extrapolated to the CBS limit using the cc-pV(X+d)Z (X=T,Q) basis sets and the two-point extrapolation formula of Helgaker. The ΔCCSD(T)/DZ term corrects this extrapolated gradient with an additive density fitted CCSD(T)/cc-pV(D+d)Z gradient, which showed strong correlation with CCSD(T)/cc-pVQZ results, detailed in the appendix. The energy of each stationary point is further refined with CCSD(T)/cc-pV(X+d)Z (X=T,Q) single points extrapolated to the CBS limit. For the analysis considering the influence of multiple catalytic H_2O or RNCO molecules, the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV[TZ,QZ]Z + ΔCCSD(T)/6-31+G** level of theory[15,290–295] is utilized to properly describe the noncovalent interactions involved in catalysis of the transition state with augmented basis functions while maintaining a feasible computational cost.

Natural Bond Orbital (NBO) analyses[34,35] are performed where appropriate to elucidate the electronic structure features associated with our predicted results. The NBO 6.0 program is interfaced with QCHEM[296] using a def2-QZVP basis set[297] and the B3LYP functional[189] to describe the exchange.

4.4 Results

4.4.1 High-Level HNCO + H_2O Reaction

The energetic corrections included in the focal point analysis (Tables 6.14-6.16) behave uniformly across all stationary points and indicate no anomalous features of this system. The ZPVE corrections are the largest and are no greater than 5.5 kcal mol^{-1} with transition state ZPVE corrections generally closer to 1.5 kcal mol^{-1}. The frozen core and scalar relativistic corrections are both small and never larger than 0.24 kcal mol^{-1}. The DBOC corrections are negligible and indicate that none of the stationary points is in the vicinity of a conical intersection or surface crossing. In all cases, excellent convergence is exhibited with

respect to basis set and higher order coupled cluster terms. These results affirm that our predictions are well within the bounds of chemical accuracy (i.e. one kcal mol^{-1}) with respect to our electronic energies. It should be noted that our enthalpy results correspond to gas phase results at 0 K. Any Gibbs Free Energy results correspond to water vapor at 298 K and 1 atm of pressure which overestimates the entropic effects manifest in the liquid phase.

Water and HNCO can react via two different concerted mechanisms, with the water O–H bond breaking across the N=C double bond to form carbamate or across the C=O double bond to form imidic acid (Figure 4.1). Both mechanisms proceed through transition states (**TS1** and **TS2**, respectively) that form four-membered rings between the two reactants. The predicted **TS1** barrier (38.5 kcal mol^{-1}) is much lower than the **TS2** barrier (47.5 kcal mol^{-1}). This is qualitatively in agreement with previous research[252,254,258] but we predict **TS1** to be somewhat higher than the recent work of Nicolle and coworkers who predicted a relative **TS1** ΔH_{0K} of 35.3 kcal mol^{-1} at the M06-2X/6-311++G(d,p) level of theory.[254] IRC computations confirm the concerted nature of these pathways and indicate that the reactants begin separated and form the depicted products. The possible van der Waal complexes formed between HNCO and water (**VDW1** and **VDW2**) are included in Figure 4.1, but are not definitively part of the reaction pathway and have small enough binding energies (1.7 and 4.3 kcal mol^{-1}, respectively) that they are mostly excluded from our discussion and would certainly be decreasingly relevant at high temperatures. Figure 6.11 depicts the 298 K Gibbs free energy surface with the loss of translational entropy leading to an average increase of relative energies by about 9 kcal mol^{-1} for all stationary points and confirms our assumptions concerning the van der Waal complexes.

The **TS1** stationary point exhibits an imaginary mode of $1726i$ as one of the water hydrogens begins to form a bond with the nitrogen atom, and the water oxygen begins to bond with the carbon. Natural Resonance Theory (NRT) predicts the bond order between the nitrogen and the water hydrogen to be 0.49 and the bond order of the nitrogen and carbon atoms to be 1.48 at **TS1**. Likewise, the carbon and water oxygen interaction exhibits a bond order of 0.56. The NBO second-order perturbation ($E^{(2)}$) analysis indicates that the most significant delocalization corresponds to electron density donated from the water lone pair into one of the isocyanate C–O* orbitals ($E^{(2)}$ =144.2 kcal mol^{-1}), which is depicted in Figure 4.2. This transition state leads to the highly exothermic formation (–15.6 kcal mol^{-1}) of carbamate (**M1**). Nicolle and coworkers predict **M1** to posess a reaction energy of –23.5 kcal mol^{-1} (at the CCSD(T)//M06-2X/6-311++G(d,p) level of theory) relative to water and HNCO, again supporting the need for a rigorous description of the correlation energy.[254] **M1** is nearly planar, with the NH_2 moiety slightly displaced from the plane of the molecule. The NBO $E^{(2)}$ analysis additionally predicts that significant delocalization of the carbonyl lone pair electron density into the N–C* and O–C* orbitals might be one of the contributing factors to the favorable product

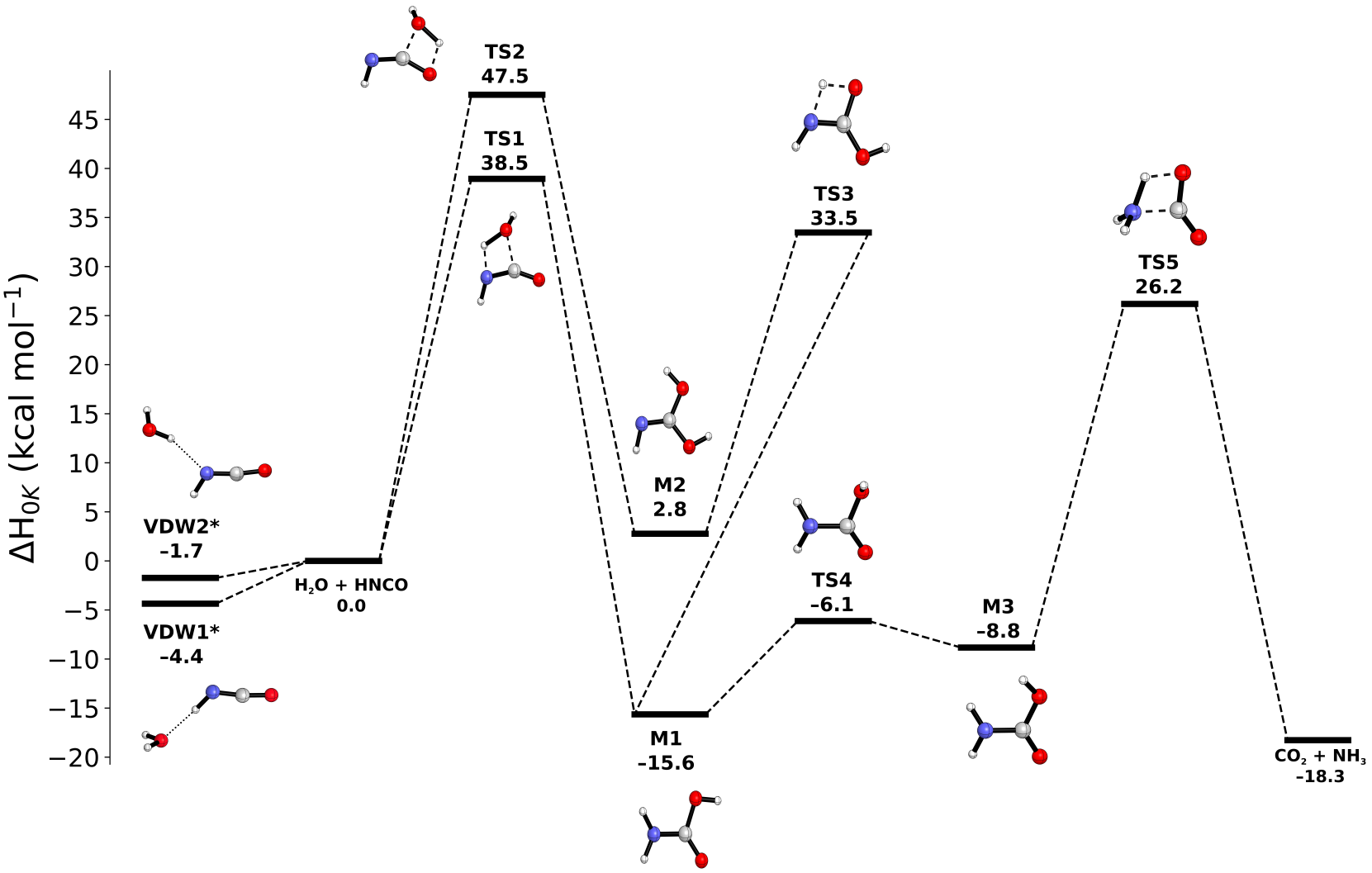

Figure 4.1: Predicted 0 K enthalpies for stationary points on the parent HNCO + H_2O potential energy surface. All values are in kcal mol^{-1} and target the CCSDT(Q)/CBS limit. Structures denoted with an asterisk are computed using augmented basis functions.

energy, affirming the findings of Bharatam and coworkers concerning the importance of delocalization in ureas and related species.[298]

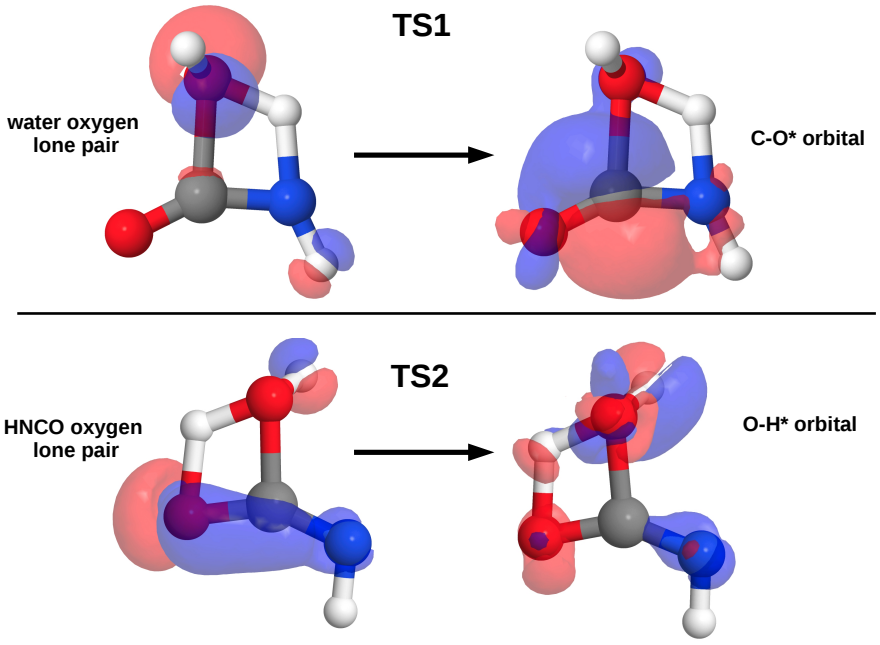

Figure 4.2: The dominant predicted NBO $E^{(2)}$ relationships in the **TS1** (Top) and **TS2** (Bottom) structures. The electron donating NBO is on the left and the acceptor is on the right.

M1 can proceed through **TS4** with a small barrier of 9.5 kcal mol^{-1} corresponding to the rotation of the hydroxyl group towards the amino group. The resulting product, **M3**, is moderately higher in energy than **M1** but is a necessary intermediate for **TS5** and the ultimate dissociation into NH_3 and CO_2. The reaction barrier for this process is quite high at 35.0 kcal mol^{-1} and is characterized by an imaginary mode of $1799i$. Dissociation to CO_2 and NH_3 is an exothermic process (9.4 kcal mol^{-1} lower than **M3**) and results in the most energetically favorable stationary point on the surface. One noteworthy feature of **TS5** is that the IRC path in the direction of **M3** terminates at a new first-order saddle point. This is because **TS5** possesses C_s symmetry and the IRC is unable to break symmetry. Slightly projecting in either direction along this mode and following the resulting IRC leads to **M3** and confirms the connectivity of **M4** and $NH_3 + CO_2$.

Water and HNCO can alternatively react through **TS2** which exhibits a $1718i$ imaginary mode as the water oxygen begins to form a bond with the isocyanate carbon and one of the water hydrogens begins to bond with the isocyanate oxygen. The new O_{HNCO}-H_{water} bond is predicted to have a NRT bond order of 0.32 while the carbon and water oxygen nearly exhibit a formal single bond order (0.90). The dominant $E^{(2)}$ term in the **TS2** structure (90.2 kcal mol^{-1}) corresponds to delocalization of the isocyanate oxygen lone pair electron density into the most proximal water O-H* orbital as represented in Figure 4.2. The magnitude of

this interaction is much smaller than the delocalization predicted in the **TS1** structure and the preclusion of delocalization could be a partial explanation for why **TS2** is much higher in energy. The product **M2** of reaction through **TS2** is the planar hydroxylated imidic acid molecule which lies only 2.8 kcal mol^{-1} higher in energy than water + HNCO. NBO computations indicate that **M2** also exhibits significant delocalization as electron density from both hydroxyl groups is donated into the out of plane C−N* orbital. **M2** also has two other conformers which correspond to rotations of the hydroxyl groups and are predicted to have CCSDT(Q)/CBS energies of 2.0 and 5.1 kcal mol^{-1} relative to **M2**.

M2 is connected to **M1** via **TS3**, which corresponds to a hydrogen transfer from one of the hydroxyl groups to the nitrogen with an associated imaginary mode of $2012i$. The barrier for this process is 30.7 kcal mol^{-1}. Unsurprisingly, **M2** lies 18.4 kcal mol^{-1} higher in energy than **M1**, confirming that the reaction through **TS1** is both kinetically and thermodynamically favored. The formation of **M1** via the step-wise procession through **TS2** and **TS3** has been debated in the literature[257] and our results seem to indicate that formation of **M1** directly though **TS1** is the most likely option absent of any other effects such as catalysts.

4.4.2 RNCO Substituent Analysis

In most industrial applications of isocyanates, the parent HNCO species is often substituted with some much larger RNCO species. The most common -R groups are generally large aromatic rings or long polymers which can both be modified to possess a diverse array of electronic structure features. Therefore, it is necessary to understand how different substituents influence the electronic structure and the resulting energetics of the RNCO + H_2O reactions. The ΔH_{0K} barrier heights and relative product enthalpies are predicted for a diverse set of 24 different substituted RNCO molecules proceeding through **TS1** and **TS2**. In all cases, the isocyanate substituents are far enough away from the active site of each reaction mechanism to ensure that the reaction pathway remains qualitatively invariant to choice of -R group. All possible conformers of each RNCO species were searched for and considered in this analysis. They are denoted alphabetically with a subscript A corresponding to the lowest energy conformer if applicable. Conformers other than the lowest are only considered in the analysis if either barrier height is lowered by more than 1 kcal mol^{-1} relative to conformer A. We admit that many of the substituents chosen might be unrealistic for industrial applications but justify the selections for two reasons. 1) The selected subset of substituents (listed in Figure 4.3) exhibits a wide array of electronic effects that allows for greater confidence that our predicted trends are robust and sufficiently span the effects normally exhibited by more standard substituents. 2) The chosen substituents are small enough to be well described by high-level methods and eliminate as much uncertainty as possible

due to compromises in level of theory. The goal of our results is that the electronic structure trends are informative and reliable enough to guide the intuition of more practical isocyanate applications.

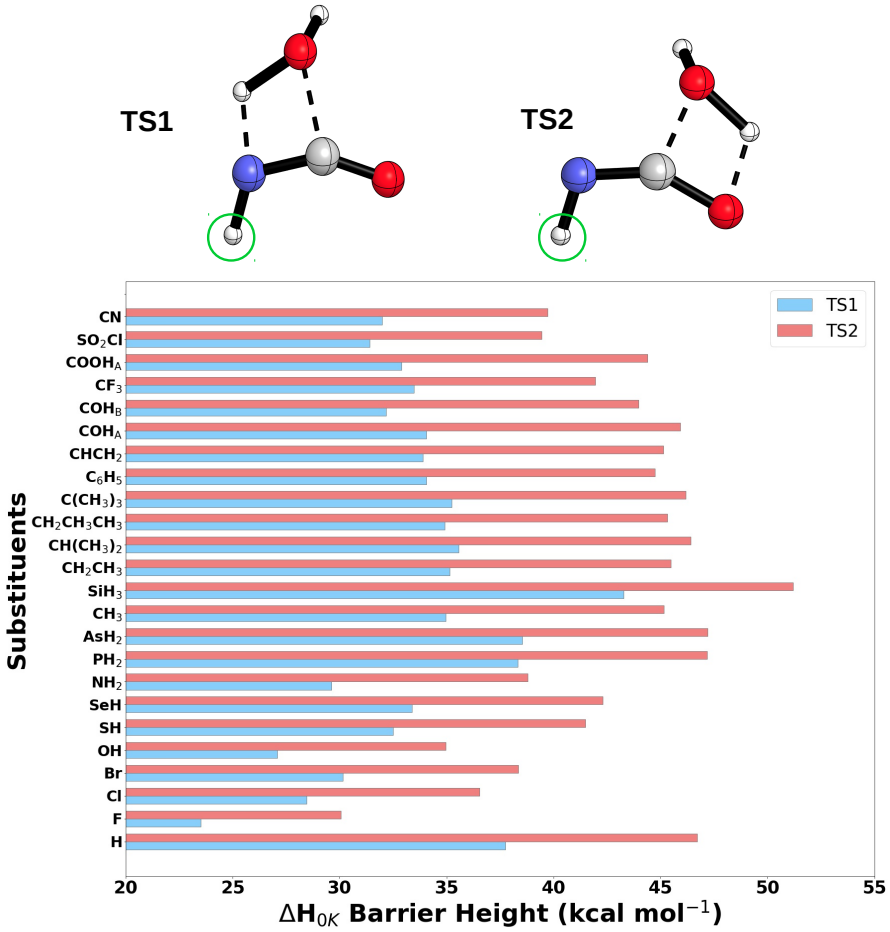

Figure 4.3: Predicted **TS1** and **TS2** ΔH_{0K} barrier heights at the CCSD(T)/CBS//MP2[TZ,QZ] + ΔCCSD(T)/DZ level of theory. The green circles indicate the location of the substituents.

Figure 4.3 summarizes the predicted CCSD(T)/CBS//MP2[TZ,QZ] + ΔCCSD(T)/DZ barrier heights for reaction across both the N=C (**TS1**) and C=O (**TS2**) bonds of each substituted isocyanate. The parent case, R=H, has a predicted barrier of 37.8 kcal mol^{-1} which is within the range of chemical accuracy of our benchmark value targeting the CCSDT(Q)/CBS limit and further reinforces the reliability of this analysis. The barriers for **TS1** are quite sensitive to substituent selection, spanning a range of 19.8 kcal mol^{-1}. The lowest barrier is predicted for R=F at 23.5 kcal mol^{-1} and the highest barrier corresponds to R=SiH_3 at 43.3 kcal mol^{-1}, significantly higher than the R=H parent case. The small main group substituents unsurprisingly have the largest effect on the barrier height, due to their extreme electron donating and withdrawing abilities. Generally speaking, most of the more standard hydrocarbon substituents lower the

barrier to carbamate formation by about 3–4 kcal mol^{-1}. This tight range seems to indicate that the size of the substituent has only minor influence on the reaction barrier, which is not surprising as Zhao and Suppes came to similar conclusions considering the size of the reacting alcohol.[274] In **TS1** and **TS2**, the substituents are opposite from the active site of the reaction and have no proximity to any moieties that might result in steric hindrance. The one exception to this might be the R=$COH_{\mathbf{B}}$ case, as the substituent hydroxyl group is oriented towards the isocyanate oxygen, resulting in a favorable intramolecular interaction.

The predicted barrier heights for **TS2** manifest trends that are qualitatively similar to the **TS1** results, with the **TS2** process always significantly less favorable. The **TS2** barrier predictions are slightly more sensitive than the **TS1** predictions and span a range from 21.2 kcal mol^{-1} with the lowest barrier predicted for R=F (30.1 kcal mol^{-1}) and the highest R=SiH_3 (51.2 kcal mol^{-1}). The most competitive case, R=F, still predicts **TS1** to be the more favorable pathway by 6.5 kcal mol^{-1}. Our results confirm that it is highly unlikely that the **TS2** barrier could be made competitive with the **TS1** barrier by substituent selection alone in the uncatalyzed RNCO + H_2O case, also noting that this conclusion is agnostic to other factors such as catalysts, solvent effects, etc. Figure 6.5 presents a scatterplot between the substituted **TS1** and **TS2** barrier heights, which correlate excellently, indicating that both barriers exhibit similar relationships to substituent identity. The supplementary information (Tables S1-S6) also contains the 298 K Gibbs free energy correction for each reaction which is generally insensitive to substituent choice and is close to 11 kcal mol^{-1} for all species.

One of the primary goals of this analysis is to elucidate any features of the RNCO molecule that could predict *a priori* the barrier heights and provide a deeper understanding of the key electronic structure features of each mechanism. NBO predictions were utilized to generate results that might be related to the transition state barriers such as atom charges, orbital occupations, and NRT bond orders. We determined that one of the best predictors of barrier heights for the RNCO molecules is the occupation of the in-plane C–O* orbital. Figure 4.4 depicts a scatter-plot between the **TS1** barrier and the C–O* orbital occupation, as well as a line of best fit that has a high R^2 value of 0.88. (The R=GeH_3 case is excluded as an outlier because it exhibits a C–O triple bond at the transition state and is not comparable to the other species.) The C–O* orbital occupation is an excellent proxy for the electron withdrawing ability of the substituents. Our results are in qualitative accord with the work of Wagner and coworkers[250] who determine, with little explanation, that electron withdrawing groups lower the barrier to reaction of alcohols across the N=C double bond. Utilizing the C–O* orbital occupation as a predictive quantity is also appropriate, due to its direct relationship with the previously discussed dominant $E^{(2)}$ interaction characteristic of the **TS1** structure. The association is so strong because less electron density present in the C–O* orbital allows for more facile delocalization of the water oxygen lone pair. Thus, the extent to which the C–O* orbital is filled in the isolated RNCO molecule

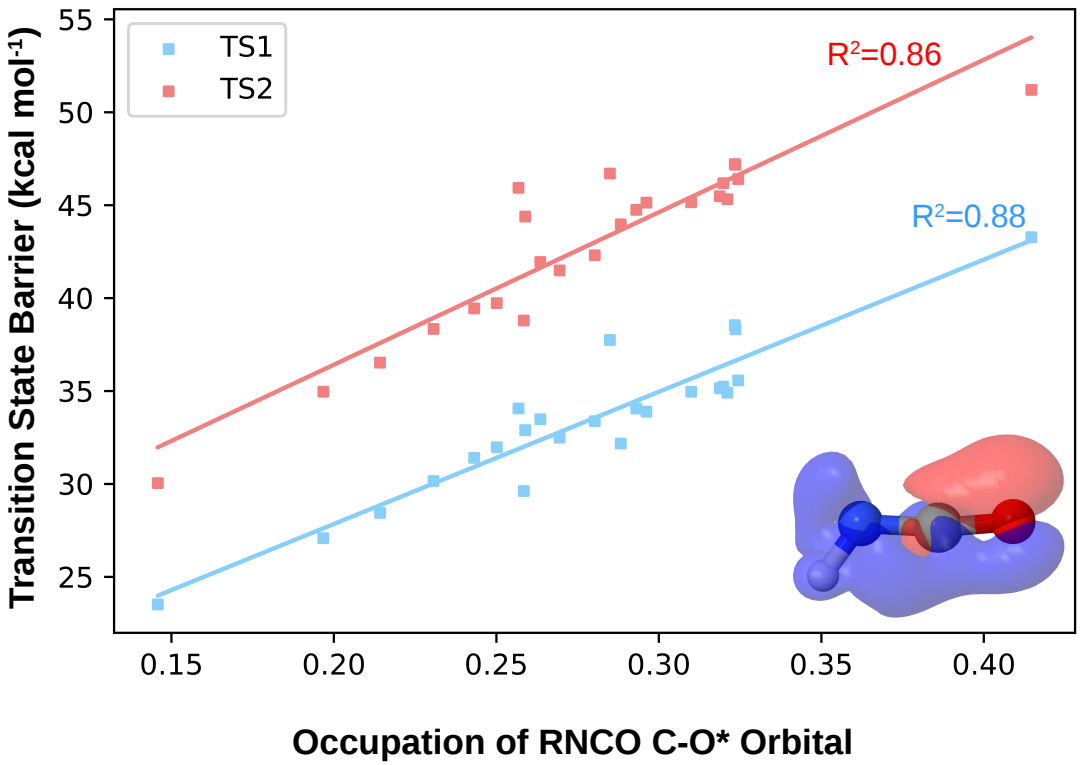

Figure 4.4: Relationship between the **TS1** (blue) and **TS2** (red) ΔH_{0K} barrier heights, respectively, and the occupation of the in plane C−O* orbital (depicted for HNCO in the bottom right of the figure) of the isolated RNCO molecule. Barrier heights are predicted at the CCSD(T)/CBS//MP2[TZ,QZ] + ΔCCSD(T)/DZ level of theory.

will directly relate to energetic barrier of carbamate formation.

Our computations predict that a similar relationship also holds for the **TS2** barriers. The line of best fit for these barriers is red in Figure 4.4 and also possesses an excellent R^2 of 0.86. In the unsubstituted **TS2** structure, the dominant $E^{(2)}$ interaction involves the delocalization of the isocyanate oxygen lone pair into the closest water O−H* orbital. As electron density is removed from the C−O* orbital via substituent effects, the C−O bond shortens and the isocyanate oxygen lone pair can delocalize more easily. To summarize, electron withdrawing substituents on the isocyanate species lower the barrier to reaction via the pathways through both **TS1** and **TS2**. This principle can almost certainly be generalized to RNCO reactions with different R-OH species and even related reactants such as R_2N-H, because the reaction motif will remain rather consistent. However, more complicated molecules reacting with RNCO might convolute these well-behaved trends, introducing new factors such as sterics, dispersion interactions, and exotic electronic structure features. Nevertheless, we expect our trends to hold given similar reaction motifs and the fact that the isocyanate substituents are not involved in the active site of the reaction. It is also key to note that the same correlations hold and maintain their fidelity when using the 298 K Gibbs energies, which usually shift all estimated values by approximately 11 kcal mol^{-1} regardless of the substituent. Thus, our substituent analysis seems robust to higher temperature conditions.

The relative enthalpies of the products of each pathway are presented in Figure 4.5. The carbamate relative energies (shown in blue) are always exothermic and somewhat sensitive to substituent substitution. They range from –28.9 (R=F) to –9.7 kcal mol^{-1} (R=SiH_3). Highly electronegative substituents such as R=F, OH, and NH_2 make the carbamate more favorable than the reactants relative to the R=H case. This is likely associated with the aforementioned proclivity of the carbonyl to donate electron density into the neighboring C−N* and C−O* orbitals. The COH_B carbamate product also stands out as being particularly favorable (see Figure 6.6 for scatterplot between substituted **M1** and **M2** relative enthalpies), but this is partially due to an intramolecular interaction originating from the substituent hydroxyl group. The imidic acid products are much more sensitive to substituent selection, which is unsurprising as NBO analysis of the parent imidic acid predicts significantly more delocalization into the C−N* orbital compared to the carbamate. The imidic acid enthalpies relative to the reactants range from –18.3 (R=F) to 9.2 (R=SiH_3) kcal mol^{-1} and many of the alkyl substituted imidic acid products are predicted to be slightly less thermodynamically favored than the reactants. There is no case where the imidic acid arrangement is lower in energy than the carbamate analogue. The closest the two isomers get in energy is 10.1 kcal mol^{-1} (R=F), indicating that they are unlikely to be thermodynamically competitive in any case. Assuming no catalyst, our data supports that the H_2O + RNCO reaction would favor proceeding through the **TS1** pathway independent of substituent.

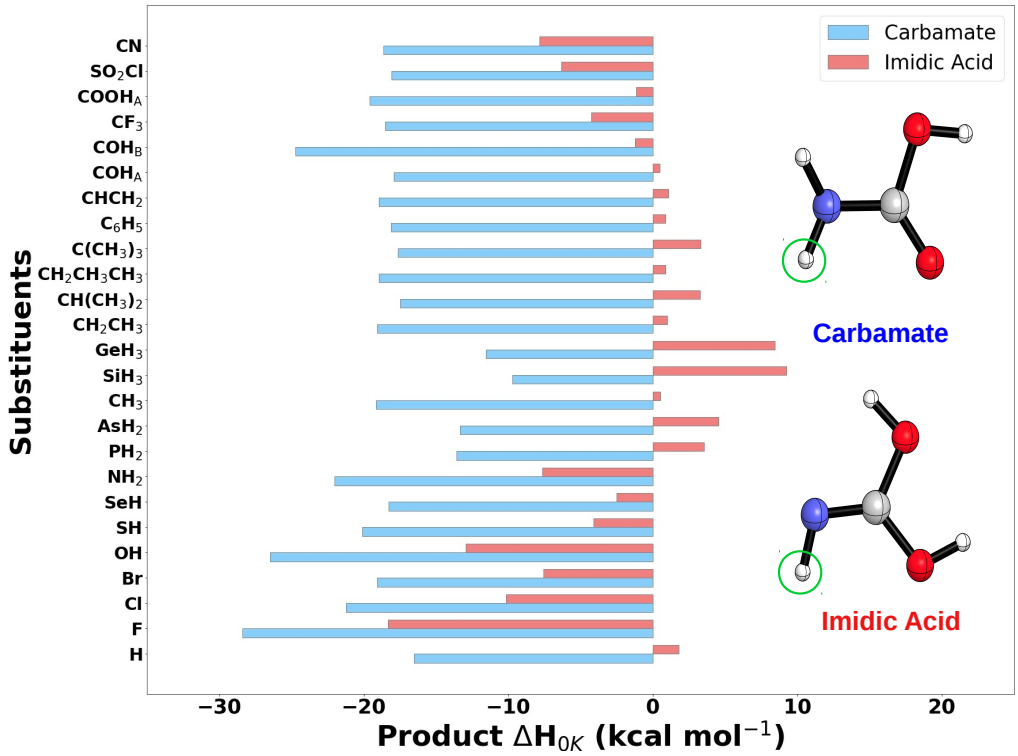

Figure 4.5: Predicted **M1** and **M2** ΔH_{0K} relative enthalpies at the CCSD(T)/CBS//MP2[TZ,QZ] + Δ CCSD(T)/DZ level of theory. The green circles indicate the locations of the substituents.

4.4.3 Multimolecular Mechanisms

Recent research has highlighted the importance of cooperative mechanisms leading to the same carbamate and imidic acid products in the presence of excess water or isocyanate molecules.[256–258] Wei and coworkers predicted at the MP2/6-311++G** level of theory that additional water molecules lower the barrier to reaction across both the N=C or C=O bonds.[258] We present an analysis of how substituents influence the barrier to carbamate and imidic acid formation considering one catalyst water molecule, two catalyst water molecules, and one catalyst RNCO molecule. Our CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([T,Q]+d)Z + ΔCCSD(T)/6-31+G** results are a significant improvement over previous computations and allows for confident determination of the trends that manifest in the autocatalyzed energetics due to substituent effects.

Catalyst Waters

The H_2O + HNCO products, carbamate and imidic acid, can be formed via a concerted reaction mechanism in the presence of one or two catalyst waters. In these cases, the catalyst waters form six or eight atom ring transition states, respectively including either the N=C or C=O bonds. The IRCs from these processes connect to pre and post reactive van der Waals complexes which are reasonably strong relative to the separated molecules (between 5-10 kcal mol^{-1} for the single water catalyzed case at the ωB97x-D3/def2-QZVPPD level of theory). These complexes are excluded from our analysis as they are decreasingly relevant at higher temperatures, and preliminary computations indicated that substituents have far less influence on the energy of the van der Waals complexes than on the transition state energies. In certain cases, the determined enthalpies of the catalyzed transition states are negative relative to the separated products, but are still referred to as "barrier heights". This should be understood as a consequence of selecting the separated reactants as a reference, rather than selecting the water trimer and HNCO molecule as a reference, for example. Figure 4.6 presents the ΔH_{0K} barrier heights for each process relative to the separated reactants. **TSN** and **TSO** refer to reaction across the N=C and C=O bonds and are depicted with blue and red bars, respectively. The number following this designation indicates the number of water molecules involved in the reaction.

In all cases, additional water molecules significantly lower the enthalpy of the transition state, with the first catalyst water having the largest effect and lowering the transition state barrier by at least 20 kcal mol^{-1}. This is unsurprising and NBO $E^{(2)}$ results indicate significant delocalization between the additional water molecules, lowering the energy of the transition state ring structures. Catalytic water molecules have a greater impact on the **TSO** barriers, which results in the **TSO3** and **TSN3** barriers being reasonably competitive in

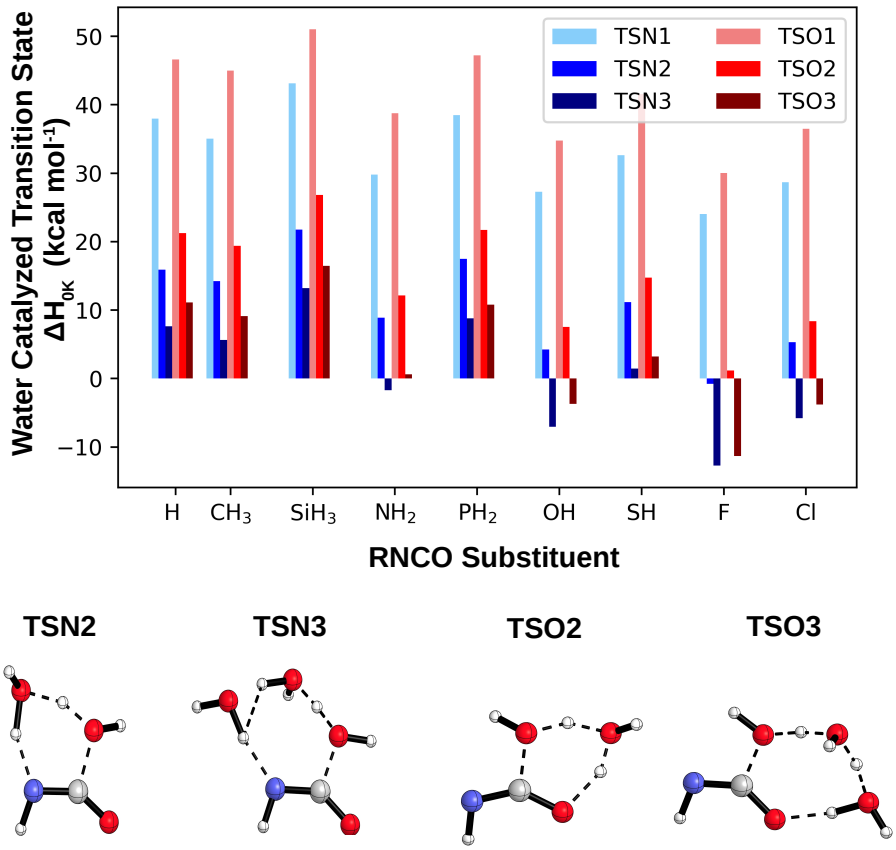

Figure 4.6: The ΔH_{0K} barrier heights for the formation of carbamate (blue) and imidic acid (red) at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV[TZ,QZ]Z + ΔCCSD(T)/6-31+G** level of theory. The lighter shades of blue and red indicate the number of water molecules involved in the reaction. The water catalyzed transition state structures are pictured and labeled in the lowest portion of this picture.

some cases with a difference of less than 4 kcal mol^{-1} across all substituents. Figure 6.8 presents 298 K Gibbs energies for each transition state. One catalyst water molecule still significantly lowers the ΔG_{298K} barriers for each mechanism (**TSN1** by at least 11 kcal mol^{-1} and **TSO1** by at least 15 kcal mol^{-1}), regardless of substituent . However, the energetic benefit of a second catalyst water is negligible as it is overcome by the loss of translational entropy. This is in qualitative agreement with the work of Wei and coworkers[258] who demonstrate convergence in the Gibbs energy of the reaction barriers with more than two catalyst waters.

The data in Figure 4.6 clearly indicate that the substituent selection plays a significant role in the energetics of the water catalyzed transition states. The previous relationship between substituent and barrier height in the H_2O + RNCO case no longer holds with these different mechanisms, at least not for the same reasons. In the case of **TSN2** and **TSN3**, there is a weaker correlation between the RNCO C−O* occupation and the barrier height, as the dominant $E^{(2)}$ interactions have changed with the new mechanism. We observe that in **TSN2** and **TSN3**, the NBO method predicts a transition state where the nitrogen exhibits two lone pairs. The first lone pair is of almost entirely *p*-orbital character, while the other, from the broken N=C bond, is significantly hybridized. The *p*-orbital character of this second lone pair correlates quite strongly with barrier heights (R^2=0.93 and 0.92 for **TSN2** and **TSN3**, respectively. See Figure 6.13.) and can be explained in terms of Bent's rules. Strongly electronegative substituents direct more *p*-orbital character from the nitrogen in the R-N bond, which increases the *p*-orbital character of the second nitrogen transition state lone pair and increases the proclivity of a N=C double bond breakage. In the **TSO2** and **TSO3** cases, we cannot make the same argument, but we do note that the RNCO C−O* occupation correlates extremely well with the barrier heights (R^2 = 0.84 and 0.85 for **TSO2** and **TSO3**. See Figure 6.14.). There is not a clear single explanation for this relationship, but it does confirm that the electron withdrawing ability of the RNCO substituent is a reliable predictor for how much one or two water molecules catalyze imidic acid formation.

Catalyst RNCO

An additional RNCO molecule can also catalyze the reaction between water and RNCO to form carbamate or imidic acid through two-step mechanisms, as previously described by Cheikh and coworkers[256] (see Figure 6.12). The mechanisms towards severing the N=C and C=O bonds both begin with six-membered ring transition states, with the catalyst RNCO nitrogen abstracting a water hydrogen and its carbon attacking the other RNCO nitrogen ($\mathbf{TSN}_A$) or oxygen ($\mathbf{TSN}_A$), respectively. The IRCs for these processes again trace to initial van der Waals complexes which need not be explicitly considered. Proceeding through these transition states forms intermediate species (**INT-N** and **INT-O**, respectively) which are energetically favored relative to the separate reactants. The final step proceeds through $\mathbf{TSN}_B$ and $\mathbf{TSO}_B$, breaking a

C−N or C−O bond of the intermediate as one of the amino hydrogens is transferred from the leaving RNCO molecule to the nitrogen or oxygen of the desired product. There are two noteworthy features concerning the energetic landscape of these multi-step reactions. First, the **INT-N** stationary point is significantly lower in energy compared to **INT-O**. This is probably due to the central linking moiety as either C−NH−C or C−O−C, respectively. The latter likely suffers from steric hindrance of the carbons which impede a favorable C−O−C bond angle. Second, in both cases, the second set of barriers ($\mathbf{TSN}_\mathrm{B}$ and $\mathbf{TSO}_\mathrm{B}$ relative to **INT-N** and **INT-O**, respectively) are the rate limiting step. The second barrier for the carbamate formation process is particularly high (54.4 kcal mol^{-1}), due to the energetic favorability of the **INT-N** species which seems to indicate that this multimolecular mechanism is likely not a catalyst for carbamate formation, as this barrier is higher than the uncatalyzed barrier through **TS1** in the H_2O + HNCO case. However, the barrier between $\mathbf{TSO}_\mathrm{B}$ and **INT-O** is only 39.5 kcal mol^{-1}, which might indicate that the imidic acid route is slightly favored with an additional isocyanate in the unsubstituted case. Additionally, the fact that the RNCO catalyzed reactions must proceed through two significant barriers would likely inhibit the catalytic efficiency of these pathways.

Substituents significantly influence the energetic landscape of the 2 RNCO + H_2O reaction. The barrier heights (0 K enthalpies and 298 K Gibbs energies) are presented in Figures 6.9 and 6.10 for each substituent case. Generally speaking, the barrier to form the $\mathbf{TSN}_\mathrm{A}$ and $\mathbf{TSO}_\mathrm{A}$ six-membered ring transition states are decreased by electron withdrawing groups and increased by electron donating groups, similar to the water catalyzed cases. In the carbamate formation mechanism, the barrier through $\mathbf{TSN}_\mathrm{A}$ is significantly less than the barrier through $\mathbf{TSN}_\mathrm{B}$ for all substituents. The lowest barrier through $\mathbf{TSN}_\mathrm{B}$ is the case of R=F (38.9 kcal mol^{-1}) which does not compete with the uncatalyzed FNCO + H_2O barrier. There is no clear single feature of either the **INT-N** or $\mathbf{TSN}_\mathrm{B}$ stationary points that definitively correlates with the $\mathbf{TSN}_\mathrm{B}$ barrier height. We do note many competing factors that might influence these barriers, such as the extent of delocalization of the carbonyl groups in **INT-N**, the strength of intramolecular hydrogen bonds in **INT-N**, steric repulsion of the intermediate's carbon atoms, the acidity of the proton transferred in the $\mathbf{TSN}_\mathrm{B}$ process, and the strength of the C−N bond breaking in $\mathbf{TSN}_\mathrm{B}$. The convoluting influence of all of these factors makes a clear explanation for the $\mathbf{TSN}_\mathrm{B}$ barrier height trends elusive.

The RNCO catalyzed imidic acid formation pathway exhibits many similarities to the carbamate formation pathway. The first barrier ($\mathbf{TSO}_\mathrm{A}$) follows a very similar pattern with respect to substituent choice, except that the barriers are generally higher than $\mathbf{TSN}_\mathrm{A}$, in each case by about 5 kcal mol^{-1}. The key difference in the imidic acid formation pathway is that the second barrier through $\mathbf{TSO}_\mathrm{B}$ is much smaller and is predicted to be between 34 and 42 kcal mol^{-1} across all the substituents considered. In every case, the energetic barriers from the respective intermediate structures through $\mathbf{TSO}_\mathrm{B}$ are much lower than the

barriers through $\mathbf{TSN}_\mathrm{B}$. This suggests that catalysis via an additional RNCO molecule could favor the imidic acid product over the carbamate product. This is certainly a consequence of the **INT-O** structure lying so much higher in energy than **INT-N**. Again, there are too many factors to provide a definite single explanation for how the substituents influence the $\mathbf{TSO}_\mathrm{B}$ barriers, but many of the same reasons mentioned in the previous paragraph are factors here as well.

Even though a mechanism exists for the $\mathrm{RNCO} + \mathrm{H_2O}$ reaction with an additional catalyst RNCO species, the energetic barriers likely remain too high to be significant without the involvement of other factors. The necessity of overcoming two substantial barriers, partially due to the favorable intermediate structures, implies that these pathways are not nearly as competitive as the water catalyzed pathways. Considering the 298 K Gibbs energy makes matters worse as the $\mathbf{TSN}_\mathrm{A}$ and $\mathbf{TSO}_\mathrm{A}$ barriers increase by about 20 kcal mol^{-1} due to the loss of translational entropy. With the Gibbs energy considered, the RNCO catalyzed pathway would hardly compete with the uncatalyzed $\mathrm{RNCO} + \mathrm{H_2O}$ reaction and would likely be far less efficient than catalysis with excess water based on our results. One should note however that Cheikh and coworkers studied the PhNCO + isopropyl alcohol reaction and determined that both excess alcohol and PhNCO catalyzed the formation of the carbamate product [256]. They note that solvent effects can significantly influence the energy of the hydrogen transfer process through $\mathbf{TSN}_\mathrm{B}$ and $\mathbf{TSO}_\mathrm{B}$ which confirms that other factors beyond gas phase predictions are necessary for a full description of these systems. Nevertheless, our substituent analysis of the $2\,\mathrm{RNCO} + \mathrm{H_2O}$ reactions should be insightful for further research on these systems in more complicated environments.

4.5 Conclusion

Our work presents the highest level *ab inito* study to date of the important $\mathrm{RNCO} + \mathrm{H_2O}$ reactions, with the goal of guiding experimental progress of novel isocyanate containing reactions. We characterize the fundamental stationary points of the reactions of water across the HNCO N=C bond to form carbamate and across the C=O bond to form imidic acid, with energies targeting the CCSDT(Q)/CBS level of theory. Composite method geometry optimizations consisting of MP2 and CCSD(T) refined with large basis CCSD(T) single points, describe the influence of 24 RNCO substituents on the barriers to carbamate and imidic acid formation. NBO analysis reveals that the occupation of the in-plane C–O* orbital of RNCO is strongly associated with higher **TS1** and **TS2** barriers, due to the the electron delocalization motifs present in the transition states. The most extreme electron withdrawing substituent (R=F) lowered both barriers ($\Delta \mathrm{H_{0K}}$) by at least 13 kcal mol^{-1} and most alkyl substituents lowered the barriers by around 4 kcal mol^{-1}.

The catalytic influence of one extra water, two extra waters, and one extra RNCO species are predicted

and discussed with respect to different substituted RNCO reactants. Electron withdrawing substituents significantly lower the barrier to carbamate and imidic acid formation in the water catalyzed cases. For the **TSN2** and **TSN3** barriers, the lower barriers are likely related to Bent's rule influencing the proclivity of the nitrogen to break the N=C bond. We also find that the **TSO2** and **TSO3** barriers are highly associated with the C−O* occupation in the RNCO reactant. The 298 K Gibbs energies predict very similar barriers in the one and two water catalyst cases. Thus, the ability of the second catalyst water molecule to lower the enthalpy of reaction is offset by the loss of translation entropy manifest at higher temperatures and confirms the findings of Wei and coworkers[258] that many additional waters produce only marginal catalytic efficiency. A catalytic RNCO molecule results in two-step mechanisms towards carbamate and imidic acid formation. Our results indicate that the highly favorable intermediate in the carbamate formation pathway (**INT-N**) makes the second barrier large and an unlikely path for carbamate formation regardless of substituent. However, the second barrier in the RNCO catalyzed imidic acid formation pathway (through $\mathbf{TSO}_{\mathrm{B}}$) is significantly lower and might indicate more efficient imidic acid production in the presence of excess RNCO species. The complex factors involved in these RNCO catalyzed pathways did not reveal any clear trends with respect to substituent implying needs for additional work on a case by case basis for more complicated systems in order to more fully understand these RNCO catalyzed pathways. Our work lays a firm theoretical foundation for the $\mathrm{RNCO} + \mathrm{H_2O}$ class of reactions and provides insights that might guide future experimental work on these industrially relevant species.

CHAPTER 5

METHANE COMBUSTION ENERGY: AN UNDERGRADUATE CHEMISTRY LAB EXPLORING COMPUTATIONAL COST AND ACCURACY

M. E. Wolf, W. Norris, H. Fynewever, J. M. Turney, H. F. Schaefer, To be submitted to *The Journal of Chemical Education*

5.1 Abstract

Over the past half century, computational chemistry has evolved from a niche field to a ubiquitous pillar of modern chemical research. Driven by the increased demand for computational chemistry in research settings, the undergraduate curriculum has evolved alongside, to ensure that students are well equipped for modern research. Towards this end, many excellent computational chemistry exercises have been developed that aim to teach students what kinds of questions computational chemistry can answer and how to properly interpret the results. However, there has been far less attention given to the complexities of determining how reliable computational results are and how constraining computational scaling can be. We present an undergraduate lab exercise that uses *ab initio* methods to predict the combustion energy of methane. The exercise walks students through the process of benchmarking error on a small system (methane), estimating the computational cost to perform the same analysis on a larger system (propane), and justifying an affordable yet accurate method for a hypothetical study of the larger system. Furthermore, students are introduced to other cost saving measures like basis set extrapolation and additive corrections. The entire exercise is intentionally designed to require little technical knowledge of computational chemistry and flexibly graft into a standard undergraduate curriculum. In order to ensure accessibility, the exercise utilizes the free open source software Psi4 (available on any OS) and provides a detailed installation and use guide for completing this lab. This lab will provide students the understanding for how to properly judge, select, and justify different computational models where cost and accuracy compete; a highly desirable set of skills that generalize to any computational science.

5.2 Introduction

Computational chemistry has become an increasingly important sub-disciple as theoretical results have consistently proven useful for a diverse array of applications such as drug design[299,300], natural product synthesis[301], material science[302,303], and high-accuracy spectroscopy.[304,305] Chemists have been so fruitful with these endeavors by taking full advantage of the rapid expansion of computational resources over the past few decades. As a consequence of this, undergraduate chemistry curricula now more frequently teach or at least expose students to computational chemistry and its applications. Some of the most natural places to graft computational chemistry training into an undergraduate curriculum have been general[306], organic[307–309], and physical chemistry courses[310–315], with students generally learning a handful of prototypical applications parallel to the main content of each respective course. However, most of these computational labs or exercises focus on teaching students very niche methods that they will likely not encounter outside of the classroom

unless they pursue a computational chemistry career. At the same time, many of these labs are designed so that students see what kinds of questions computational methods can answer rather than teaching them how to think like computational chemists. This pedagogy was historically appropriate when computational methods were merely auxiliary tools for chemical research. Now however, the ubiquity of computational methods for all sub-disciplines of chemistry (and most scientific fields in general) demands that chemistry students begin to develop the necessary skills to properly produce and interpret computational results.

In a seminal review concerning the introduction of computational chemistry into the classroom[316], DeKock and coworkers reference six steps of a computational chemistry study[317] (arguably generalizeable to computational studies in any field) proposed by Trinajstic: (1) Problem identification, (2) Theoretical framework development, (3) Numerical approach, (4) Relevant computation, (5) Interpretation of results, (6) Documentation. We suggest that much of the current material for computational chemistry education does an excellent job introducing students to items 1,5, and 6. Items 2 and 3 are understandably beyond the scope of most undergraduate audiences (primarily due to insufficient mathematics prerequisites.) and generally reserved for specialists in most cases. The fourth item, relevant computation, is somewhere in between. In one sense, significant focus is given to the kind of computation (e.g. geometry optimization or reaction energy) necessary to answer a question; so much so that Davidson warns against overemphasizing this step at the expense of the others and performing computations that lack necessary context.[318] On the other hand, little attention is usually given to the quality of selected methods and therefore their ability to properly address the research question, particularly in terms of comparison to experiment. Davidson goes on to warn that " "the user of the elaborate program packages and supercomputer resources must have a fairly deep understanding of the subject if sensible results are to be obtained. The user must specify the basis set, the initial geometry, the correlation method, etc. More importantly, the user must understand the relationship between the experiment and what can be reliably computed."[318] DeKock and coworkers concur and assert that failure to do so properly can lead students and even instructors to doubt the reliability of their own computational predictions[316], a detrimental takeaway for students learning computational chemistry. Some recent educational articles, for example, do address the complexities of determining a "relevant computation" by introducing students to the method dependence of results and even some advanced techniques like basis set extrapolation[313–315,319]. However, model selection (and the logic to do so properly) is often a tangential topic as students cannot be expected to understand the technical details of different models or have the time available to perform expensive benchmark analyses. This warrants creativity in developing lessons that guide students towards understanding the selection and justification for methods used in a computational study, which seems to persists as a weak spot in current undergraduate chemistry education. The exercise we propose will address this gap by focusing on how the goal of obtaining accurate results is always limited

by the constraint of computational cost using an *ab initio* study of a fundamental chemical process.

The constant tension between computational cost and the reliability of results is a reality present in every sort of computational science including *ab initio* quantum chemistry. Despite the fact that computational resources are consistently improving, the scaling of many quantum chemical models far outpaces these technological and algorithmic improvements. This practically results in a short-run scenario where computational resources are essentially fixed. Theoretical chemists are constantly improving the algorithms, paralellization, and development of new methods, but this takes time and rigorous benchmarking. At best, improvements usually reduce scaling factors slightly rather then change the nature of the exponential scaling exhibited by *ab initio* methods. Therefore, at any given time, chemists applying currently available quantum chemical models need to make difficult decisions about the complexity and paramaterization of their selected models and deal with the consequential error manifest in each. The important principles and decision making paradigms to do this well in an *ab initio* context can be generalized to other computational chemistry applications and many broader applications of computational science.

We propose an undergraduate physical chemistry laboratory exercise that utilizes high level *ab initio* techniques to study the combustion energy of methane ($CH_4 + 2\,O_2 \longrightarrow CO_2 + 2\,H_2O$) as a representation of the cost-benefit decision making process intrinsic to computational chemistry study. This exercise is appropriate for upper level undergraduate students who are already familiar with basic physical chemistry concepts like the Schrödinger Equation, reaction energetics, enthalpy, etc. Although this analysis relies on sophisticated methods such as coupled-cluster and technical concepts like basis set design and extrapolation, we intentionally avoid in-depth discussion on these topics and instead focus on the pragmatic interpretation of their results. The exercise focuses primarily on the balance between the error and computational cost of various combinations of method and basis set, aimed at predicting the reaction energy of methane combustion. These findings will be used as a benchmark for the hypothetical study of the combustion energy of propane. Students will hopefully accomplish the following by completing this lab exercise: 1) gain an appreciation for the necessity of testing different models and their relative errors rather than assuming computational chemistry is "black box" or "mindless"; 2) become familiar with the concept of computational scaling and how quickly an increase in parameters results in an infeasible computation time; 3) practice the skills needed to justify and communicate the quality of model selection based on a benchmark study. The flexibility of this lab allows instructors to discuss more advanced topics such as geometry optimization, basis set composition, specific *ab initio* methods, and various other technical topics. However, the lab can be completed independent or agnostic of many of these technical details depending on the context of the course and student background knowledge. The rest of this work will provide a simple summary of the theoretical background necessary for instructors to teach the content. This is followed by in-depth explanation of the

logic behind the design of each part of the exercise and the modifications possible to enhance student learning such that instructors are equipped to maximize the efficacy of the provided materials.

5.3 Brief Theoretical Background

The cornerstone of this exercise and all other *ab initio* studies relies upon ground state solutions of the time independent electronic Schrödinger equation, $\hat{H}\Psi = E\Psi$. Many-electron solutions of the Schröodinger equation cannot be solved analytically for all but the simplest of cases. To circumvent this, many approximate methods have been developed. These approximate methods rigorously describe the important physical features of the electronic wavefunction such as the required antisymmetry of electrons and their spin, as to not violate the Pauli principle. Numerical solutions are produced by describing the molecular system as a set of functions that model the atomic orbitals of each atom known as a basis set. Thus, every *ab initio* quantum chemical prediction is said to be obtained at some level of theory which is a combination of method and the basis set employed to produce the numerical solutions. The most fundamental *ab initio* method, Hartree–Fock, optimizes the "best" set of molecular orbitals, that is, the set of linear combinations of atomic orbitals that most minimize the electronic energy (i.e. in terms of the variational theorem) subject to the constraint of orbital orthonormality. The Hartree–Fock method results in a physical description of a molecular system that includes the nuclear-nuclear repulsion, electron-nuclear attraction, electron-electron repulsion, electron kinetic energy, and nuclear kinetic energy, based on the mean field approximation (each electron interacts with the average field of all others) and Born–Oppenheimer approximation (nuclei are fixed with respect to electron motion.). Because a basis set is used to numerically optimize the Hartree–Fock result, increasing the size of the basis set provides increased flexibility for the description of the molecular orbitals which results in a lower electronic energy. As a basis set size approaches the limit of infinity, the energy approaches what is known as the Hartree-Fock limit (or more generally, the Complete Basis Set (CBS) limit) and is negligibly lowered by additional basis functions. All computations use a finite basis which must necessarily introduce some error into the system, but this error can always be systematically removed simply by adding more basis functions. Doing so, however, will rapidly increase the memory requirements and number of mathematical operations required, which can result in unreasonable computation times and necessitate selection of some smaller basis set and its associated error.

The Hartree-Fock method, though a great start, fails to fully describe the simultaneous interactions of electrons in a system and instead approximates each electron as it interacts with the mean field of all other electrons. The result of this is an overestimation of the true electronic energy (i.e. the Full Configuration Interaction (FCI) energy). The difference between the Hartree-Fock energy and the FCI energy is known as

the correlation energy. A FCI computation can be performed, but is impossibly expensive for all but the smallest systems. Thus, many different methods (e.g. coupled cluster) have been developed that estimate the correlation energy of a molecular system and approach the FCI limit in a cost effective manner. A detailed explanation and derivation of these methods is far beyond the scope of an undergraduate context and we direct the reader to the work of Cramer for an excellent introduction to the theory.[2] For the purposes of this exercise, one must only know that certain *ab initio* methods are developed to systematically converge towards the exact FCI solution.

Though many approximations and sources of error exist in standard *ab initio* methods, the errors due to a finite basis set size and correlation energy are usually considered the most prominent. The exercises we propose will consider these two sources of error and predict the combustion energy of methane at various levels of theory. One of the benefits of computing relative energies is the fact that cancellation of error removes much of the absolute error that might be manifest in the products and reactants as the individual absolute energies approach the FCI/CBS limit. The methods considered will be Hartree-Fock, MP2, CCSD, and CCSD(T) in order of increasingly rigorous description of correlation energy. The basis sets utilized will be the Dunning family[11] of correlation consistent basis sets (cc-pVXZ X=D (double),T (triple),Q (quadruple),5 (quintuple)). The X refers to the cardinality, or size, of the basis sets. This small array of methods and basis sets will allow students to predict reaction energies at sixteen different levels of theory and observe the trends in the error and computational cost of each. All results can be obtained in the free open-source software package Psi4[196] with the aid of the accompanying tutorial. If the software or computing resources are unavailable, the supplementary information contains a spreadsheet of all of the data which can be given to students at the discussion of the instructor and used to complete the questions as if the students generated the data themselves.

5.4 Lab Design

5.4.1 Obtaining the data

The first part of the exercise asks students to compute the energy of each product and reactant at the aforementioned levels of theory using the provided Cartesian coordinates (optimized at the CCSD(T)/cc-pVTZ level of theory) and a provided step-by-step Psi4 tutorial. Though this section is not particularly enlightening, it allows students to interact with the input, output, and submission files of a computational chemistry software package and facilitates practice with logical organization of data in a file-system. The instructions encourage students to convert the results to units which they are likely more familiar with (kcal

mol^{-1}) and perform a quick sanity check to ensure that the results they obtained are free of any major mistakes. For example, if the O_2 MP2/cc-pVDZ energy was accidentally copied into the MP2/cc-pVTZ column as well, the predicted reaction energy would likely be at least an order of magnitude incorrect from the expected result. Mistakes made by students in this regard would provide excellent practical experience in troubleshooting computations results. Input files are provided for ease of use, and the experiment is designed to minimize the number of options that need to be changed, therefore minimizing mistakes. We also provide the cc-pV5Z row energies so that students have something to compare their data to and are spared from obtaining the most costly data.

It should be noted that the level of theory of the geometry optimizations could influence the energetic analysis performed in the following sections. It is common practice to obtain the best possible geometries and run energetic analyses "on top of" them. We provide geometries at a sufficiently high level of theory that this need not be a concern here but instructors should be aware of this fact and could use this as an opportunity to extend the study to geometry optimization as well. Further discussion of this extension is discussed in the "Suggested Modifications" section of this manuscript.

All of these computations should be able to be run on most standard laptops or workstations. The Psi4 guide provided in the supplementary information section provides detailed information on how to install and run Psi4 on Windows, Mac, and Linux systems. Psi4 is a free, open-source, program with excellent support and requires no special permissions to use. Furthermore, the guide should fully explain all of the information necessary for a beginner to complete this exercise in its entirety, including terminal navigation. Though we provide a firm foundation for instructors and students to use Psi4, we leave it to the instructor to choose what system the students will work on, that their machines have the necessary specifications to run properly, and that the job submission commands are well understood. In cases where the most expensive jobs might be too memory or time intensive for certain machines or when technical difficulties arise, we provide a spreadsheet containing all of the data which can be dispersed to the students at any time with the instructor's discretion.

5.4.2 Error Analysis

One of the hopeful assumptions in *ab initio* quantum chemistry is that the errors due to a finite basis set and correlation energy can be considered somewhat independent of one another, or at least treated independently. This section of the lab uses the previously obtained data to guide the students in observing the trends in methane combustion energies with respect to basis set and treatment of correlation energy, with the ultimate goal being to select levels of theory that produce minimal errors in their predictions. The

questions in this section are briefly prefaced to ensure the students realize that the methods and basis sets selected will converge to eliminate their respective errors if taken to infinite order, but do possess error given our truncation of each. We provide a highly converged estimate of the FCI/CBS energy obtained via focal point analysis[18–21] that is a reasonable proxy for the "exact theoretical answer" and can be used by the students to estimate the error in their combustion energies. Table 5.1 presents the focal point analysis for the methane combustion energy broken down by increments up to the valence CCSDTQ(P) level of theory and with methods up to CCSD(T) extrapolated to the CBS limit using a large cc-pV6Z basis set. Notice how higher order corrections decrease in magnitude (e.g δ(P) is 0.25 kcal mol^{-1} with a cc-pVDZ basis) and the jumps between the cc-pV6Z estimates and the extrapolated CBS limit are small (e.g HF/cc-pV6Z is only 0.03 kcal mol^{-1} from the HF/CBS estimate using a three-point extrapolation formula[24] and the CCSD(T) increment is only 0.01 kcal mol^{-1} from the CCSD(T)/CBS limit obtained with a two-point extrapolation formula.[23] resulting in a jump in the net column of only 0.53 kcal mol^{-1}) This justifies the tight convergence of our provided CCSDTQ(P)/CBS "exact answer" and its utilization to estimate error in lower levels of theory.

Table 5.1: Incremented focal point table for the methane combustion reaction. Values in brackets correspond to methane combustion energies that are extrapolated or are additive corrections. The bold value in the bottom right is the FCI/CBS estimate. Units are in kcal mol $^{-1}$.

$\mathbf{CH_4 + 2\,O_2 \longrightarrow CO_2 + 2\,H_2O}$									
	HF	$+\delta$ MP2	$+\delta$ CCSD	$+\delta$ (T)	$+\delta$ T	$+\delta$ (Q)	$+\delta$ Q	$+\delta$ (P)	NET
cc-pVDZ	–157.30	–21.46	8.93	–0.62	0.12	0.61	0.01	0.25	–169.46
cc-pVTZ	–161.68	–35.42	12.87	–1.19	0.23	0.74	[0.01]	[0.25]	[–184.19]
cc-pVQZ	–163.18	–41.57	14.41	–1.35	0.36	0.76	[0.01]	[0.25]	[–190.32]
cc-pV5Z	–163.58	–44.02	15.38	–1.42	[0.36]	[0.76]	[0.01]	[0.25]	[–192.26]
cc-pV6Z	–163.52	–44.77	15.72	–1.41	[0.36]	[0.76]	[0.01]	[0.25]	[–192.62]
CBS	[–163.49]	[–45.82]	[16.18]	[–1.40]	[0.36]	[0.76]	[0.01]	[0.25]	[**–193.15**]

Analysis begins with an examination of the trends in the predicted combustion energies. Students are first asked to consider the cc-pVDZ basis set only and comment on the overall trend in the reaction energies as the quality of method increases (See Table 5.2 for the results students will obtain). It is expected that a sufficient answer would address both the qualitative and quantitative trends so that students can further develop their intuition for the units generally used to describe reaction energies. Instructors should note that the trends observed are not necessarily monotonic and particular emphasis should be given to the magnitude of the differences between predicted values as convergence may not occur from one side to the true answer. For example, HF greatly overestimates the reaction energy while MP2 significantly underestimates it, but the CCSD and CCSD(T) estimates are usually within 2 kcal mol^{-1} of each other. Students are furthermore asked to consider if their observed trends for the cc-pVDZ row are the same for all basis sets. A key learning

goal from this section is that despite the systematic behavior of *ab initio* methods, the convergence patterns may exhibit more complicated motifs.

The second question asks for a description of the trends in combustion energy with respect to increasing basis set size. These trends are much clearer and are monotonic for each column. The final question of this section requires students to compare each reaction energy to the "exact answer" and rank the top five levels of theory, noting which are within chemical accuracy. This task is straightforward, but sets up the cost benefit analysis in part 4.

Table 5.2: Reaction energies for methane combustion computed at various levels of theory. The units are in kcal mol^{-1}.

$\mathbf{CH_4 + 2\,O_2 \longrightarrow CO_2 + 2\,H_2O}$ **Reaction Energies**				
Basis Set	Hartree-Fock	MP2	CCSD	CCSD(T)
cc-pVDZ	–157.30	–178.76	–169.83	–170.45
cc-pVTZ	–161.68	–197.09	–184.23	–185.41
cc-pVQZ	–163.18	–204.74	–190.33	–191.68
cc-pV5Z	–163.58	–207.59	–192.21	–193.63

5.4.3 Cost Analysis

The cost analysis section of the lab is the cornerstone of this exercise and is designed to go beyond most standard computational labs and investigate how computational scaling (and the associated cost) complicates the selection of level of theory for any study. Most computational chemistry labs generally focus on quickly obtaining some computationally inexpensive data and the rest of the time is spent analyzing and answering chemical questions with this data. Though the pedagogical value of these kinds of exercises cannot be overstated, the problem with stopping here is that students are left agnostic to the practical difficulties associated with ensuring a computational study can be completed with available resources. The consequence of this approach at best fails to train students how to think like a computational scientists and at worst provides the false impression that all computational studies are quick, easy, and always produce clean and meaningful results regardless of level of theory. The following few questions are designed to allow students to interact conceptually with how quickly computational cost becomes a limiting factor for theoretical studies of even small molecules, without wasting their time to experience it first hand. Memory can also be severely limiting for many of these computations, but we ignore its influence for the sake of simplicity. It is very likely that students could interact with algorithms in other computational science fields that scale poorly but have low memory requirements, thus a cost analysis agnostic of memory is of great pedagogical value.

Computational scaling is conceptually introduced in this section without requiring the students to have

any background knowledge of computer science. Our focus is not on the complexities of algorithm design, but rather on how the consequences of computational scaling manifest and are dealt with in practice. The N^x notation (N is the number of basis functions and x is the scaling factor) is introduced and assumes an understanding of exponential functions. Without requiring participants to actually perform unreasonably long computations on larger systems, we simulate this by requiring students to estimate the time it would take to compute the energy of propane with all sixteen previously considered levels of theory. Hypothetical timings are provided for the energy computations of each method with the cc-pVDZ basis set, assuming constant resources. The number of basis functions necessary to describe the propane molecule is provided for each basis set cardinality along with the scaling factors for each method: HF (x=4), MP2 (x=5), CCSD (x=6), and CCSD(T) (x=7). Using Equation 5.1 (where T denotes the time, N denotes the number of basis functions, x specifies the scaling factor, and the subscripts denote the basis set) timing predictions can be made for all for all levels of theory using Equation 5.1 and should reproduce Table 5.3.

$$T_{\mathrm{cc-pVDZ}}(\frac{N_{\mathrm{cc-pVDZ}}}{N_{\mathrm{cc-pVXZ}}})^x = T_{\mathrm{cc-pVXZ}} \tag{5.1}$$

Instructors should make it clear that all timing predictions are only estimates and the true computation times may depend on many factors such as the specific machine used or memory allocation. It should also be noted that each method is not independent. For example, all perform an initial Hartree–Fock and the CCSD(T) procedure requires CCSD to be completed first. Regardless, the most sophisticated method still determines how each computation scales and warrants this sort of approximation, at least for a rough estimation of timings. For computational procedures that are not built atop one another, this kind of analysis would be even more reliable and is therefore worth completing for pedagogical purposes despite the oversimplified assumptions. Question 4 consists of four parts that are intended to simulate various considerations one might make when conducting a theoretical study. Parts a, b and d are very practical while part c considers a unreasonable year-long CCSDT computation and tests one's intuition as to how quickly computations can become infeasible. Part d of this question is the most complex and asks students to predict what size alkane could be computed with the CCSD/cc-pVTZ level of theory if the time depended only on the number of basis functions and the baseline CCSD/cc-pVDZ, assuming exact scaling. In reality, due to the mathematics of these *ab initio* methods, the additional electrons added to the system would also increase the computation time significantly, but this exercise is nevertheless of educational benefit for judging the viability of a research project and its time constraints.

Table 5.3: Estimated timings for the computation of the electronic energy of propane at based on hypothetical cc-pVDZ timings, the number of basis functions, and the scaling factors. The units are in seconds.

Propane Single Point Energy Estimated Timings					
Basis Set	$N_{\mathrm{cc-pVXZ}}$	HF (x=4)	MP2 (x=5)	CCSD (x=6)	CCSD(T) (x=7)
cc-pVDZ	82	0.1	0.5	2	2.3
cc-pVTZ	202	4	45	447	1266
cc-pVQZ	405	60	1470	29032	164898
cc-pV5Z	712	572	24851	864339	8642862

5.4.4 Balancing Cost and Accuracy

This section of the exercise considers the hypothetical situation where students are given a deadline of one week and asked to obtain the most reliable combustion energy of propane that they can using the previous levels of theory employed for methane. These goals and constraints mimic what computational chemists do on a daily basis. That is, make decisions in order to produce the most accurate and meaningful results that fully answer scientific questions in a reasonable amount of time. In practice these kinds of decisions are made based on reliable error benchmarks of small related systems combined with cost estimates or test timings of the larger systems (i.e. the findings of the previous two parts of this exercises.). The proposed scenario makes students follow the thought process computational chemists must undergo in the real-world research settings. It is likely that students will select the obvious CCSD(T)/cc-pVQZ level of theory which exhibits an error of 1.47 kcal mol^{-1} for methane and is predicted to finish in less than two days. The part of the question that requires much more thought begins with a reminder that the validity of this analysis rests on the assumption that trends in methane combustion energies behave similar to the trends in propane combustion energies with respect to level of theory. Students are asked to suggest how they might provide evidence that tests this assumption before coming to lab next week. The goal of this question is for students to realize that most of the levels of theory are extremely cheap to obtain even for propane and they could use them to predict propane combustion energies for must combinations of basis set and method. These results, and their manifest trends, could be compared where possible to the methane case (similar to part 1). Their agreement or lack thereof would be an easy way to strengthen or weaken the assumption, respectively, that methane combustion is an appropriate benchmark for the propane system.

This section also asks a question concerning what might be an anomaly at first glace. The MP2/cc-pVTZ methane combustion energy is predicted to have an error of –4 kcal mol^{-1} relative to the "exact answer" while the adjacent CCSD/cc-pVTZ prediction has and error of 9 kcal mol^{-1} and the adjacent MP2/cc-pVQZ prediction has an error of –12 kcal mol^{-1}. This seems to contradict the assertion that increasing the basis set size and treatment of correlation energy will remove error from the prediction. Students are asked to

provide a possible explanation for this. The systematic converge of basis set cardinality and coupled cluster series towards the FCI/CBS limit is not falsified by this (see table 5.1), but it just so happens that the MP2/cc-pVTZ case benefits from a fortuitous cancellation of error. One way to understand this (given as a hint in the problem) is to consider what the values are approaching in each respective column and row. The values in the cc-pVTZ row approach the FCI/cc-pVTZ limit which would eventually treat all correlation energy but has significant basis set error relative to the FCI/CBS limit. Likewise, the values in the MP2 column approach the MP2/CBS limit which has no error due to basis set size, but significant error due to treatment of correlation energy. Essentially, MP2/cc-pVTZ is far from being converged with either source of yet they cancel out nicely. Though all of our predictions rely on cancellation of error to some extent, doing so with such a low-fidelity method and small basis set could have disastrous consequence for the accuracy of predictions made for a larger system like propane. Small changes in both sources of error are expected when moving to a larger system and could significantly increase the total error by removing the fortuitous cancellation or error. Predictions that are tightly converged along both axes are expected to be far more robust to minor differences and ensure the integrity of the final prediction with much more certainty. A numerical example is worked out in the accompanying answer key to demonstrate this fact and can be used by the instructor to ensure students understand the concept. Thus, it is far more reliable to benchmark against predictions that demonstrate convergence towards the FCI/CBS limit than "accidental" cases of error cancellation.

5.4.5 Cost Saving Measures

The lab ends with a final section that introduces students to some commonly applied techniques that either cut cost with minimal loss in accuracy or improve accuracy with minimal increase in cost. The presented techniques are contextualized as representative examples of the kinds of developments, approximations, and tools that computational scientists are consistently utilizing. The first example is the inclusion of additive corrections which could treat a variety of physical effects that we have not yet considered such as: core correlation, relativistic effects, or spin-orbit coupling. We focus on the inclusion of a 298 K enthalpy correction to our predicted methane combustion energies that allows comparison to an experimental result. Students are asked to lookup the experimental ΔH_{298K} combustion energy of methane from the active thermochemical tables of Ruscic and coworkers.[320] A link is provided and the database can be easily navigated. The experimental value considers water in the liquid phase and will need to be adjusted using the 298 K enthalpies of liquid and gas phase water and Hess's Law. The first part of the question asks for the difference between our "best" CCSD(T)/cc-pV5Z energy and the experimental value, which disagree by approximately

1.8 kcal mol^{-1}. This is an unsurprisingly large deviation due to our exclusion of any treatment for the rotational, vibrational, and translational energy present in the species at a temperature above absolute zero. Next, students are instructed to compute the harmonic vibrational frequencies for each species in the reaction at the modest MP2/cc-pVDZ level of theory (guidance for this is given in the Psi4 tutorial) and use the results to produce a 298 K enthalpy additive correction which is to be amended to the CCSD(T)/cc-pV5Z reaction energies and produce a theoretical combustion enthalpy. The predicted enthalpy of combustion the students will obtain is in remarkable agreement with experiment exhibiting an error of less than 0.5 kcal mol^{-1}. This simple example highlights how the inclusion of additional physical features of the system, even at a lower level of theory, can improve the accuracy of our predictions an allow meaningful comparison to experiment. Certainly, many other improvements could be made, but most are beyond the scope of this work. The instructor should ensure students appreciate how effective *ab initio* methods can be at predicting experimental thermodynamics and how a proper benchmark can make the difference between trustworthy predictions of experimental values (i.e. within chemical accuracy) and significantly wrong predictions that could even lead to qualitatively wrong conclusions.

The second technique introduced is basis set extrapolation. This tool can be utilized because the Dunning basis sets are built in such a way that energies obtained from a given method with basis sets of consecutive cardinality can be fit to a curve and extrapolated to predict the CBS limit. The details of this are technical and beyond the scope of this manuscript or the lab exercise, but previous work defends their reliability and applicability in cases like our study of methane combustion.[22] In order to understand this approach in a way that clearly communicates the point, students are provided with a value that is given as the CCSD(T)/CBS limit. In fact, this value (–194.52 kcal mol^{-1}) is obtained via the highly converged two-point extrapolation of the correlation energy (X = 5, 6) and three-point extrapolation of the HF energy (X = Q, 5, 6). Obviously, this value could be improved with even larger basis sets, but is sufficient for the students to compare to as the "true CCSD(T)/CBS" value. An astute observer might notice that the CCSD(T)/CBS value is actually farther from the provided FCI/CBS "exact answer" . This is because the CCSD(T)/CBS value lacks the higher order corrections contained in the FCI/CBS estimate, warranting comparison to a separate CCSD(T)/CBS value. Students are prompted to produce a table of error due to basis set incompleteness for all of their methane CCSD(T) combustion energies relative to the provided CCSD(T)/CBS value and record their observations. Next, students are instructed to use the line provided in the Psi4 guide to compute a CCSD(T)/cc-pV{T,Q}Z estimate of the CCSD(T)/CBS energy for each species, calculate a reaction energy from these extrapolated energies, and compare to the provided CCSD(T)/CBS energy. The CCSD(T)/cc-pV{T,Q}Z combustion energy ends up being closer to the CCSD(T)/CBS value than the CCSD(T)/cc-pV5Z prediction, demonstrating how useful basis set extrapolation can be. To drive home the point on the cost

savings of this method, students are asked to use their timing estimates for propane and determine how much time is saved between consecutive CCSD(T)/cc-pVTZ and CCSD(T)/cc-pVQZ jobs compared to a single CCSD(T)/cc-pV5Z job. The section ends with a cautionary statement on basis set extrapolation and notes that even this procedure is not "black box" and often requires comparisons and rigorous benchmarking between different extrapolation schemes.

5.5 Suggested Modifications

The design of this lab is intentional to maximize its flexibility within and undergraduate chemistry curriculum. The content as presented is primarily aimed at a standard physical chemistry student. However, we recognize that certain institutions have varying degrees of rigor in their physical chemistry programs and might need to modify the presented content to meet the academic needs of students. All materials, including a comprehensive example answer key, are provided as Microsoft Word files and we encourage modification of any and all parts of the exercise as necessary to maximize learning potential. The following discussion focuses on suggested extensions for advanced students and an alternate simplified version that could be presented to general chemistry students to "get their feet wet" with computational chemistry.

One of the most obvious and important extensions of this lab is to consider how the level of theory influences the optimized equilibrium structures of each species. One possible route of study could involve optimizing the geometries of CH_4, O_2, CO_2, and H_2O at various levels of theory. All of the bond distances (only four distinct: C-H, O-O, C-O, and H-O) could be recorded for each level of theory and a root-mean squared error could be obtained relative to the provided CCSD(T)/cc-pVTZ geometries. This would provide students the opportunity to see what kind of error is manifest in the geometry at different levels of theory and how long each optimization takes. Instructions for preforming these kinds of computations are provided in an appendix of the Psi4 guide. A similar analysis could be performed with the harmonic vibrational frequencies (CCSD(T)/cc-pVTZ frequencies are provided in the SI for reference as well). The instructor should use great care in selecting which levels of theory should be considered in these analyses as geometry optimization and frequency computations are far more costly than single-point energies, limiting the levels of theory accessible in a classroom setting. These kinds of extensions could be completed by dividing each optimization or frequency computation amongst students and compiling the data to analyze collectively. Also, because this laboratory touches on may quantum chemical concepts, instructors might use this lab to introduce a more rigorous study of concepts such as basis sets, thermochemistry, or Hartree–Fock theory quite naturally.

Alternatively, this lab can be simplified for potential use in a general chemistry lab. In this case, students

could complete the same lab exercises but, instead of learning Psi4, simply be given the data and asked to draw the same conclusions. Instructors could also provide the timings in the cost analysis section and not require the students to compute them. Likewise, the basis set extrapolation section could be a small standalone exercises that provides a straightforward example of basis set effects in quantum chemistry. Learning how to use Psi4 for the first time and obtaining all the required data could be overwhelming to a general chemistry student. Nevertheless, as the ubiquity of computational science increases, it would be worthwhile for the students to be exposed to running any sort of computation as early in their academic career as possible.

5.6 Conclusion

We present a novel physical chemistry laboratory exercise that focuses on the tension and balance between accurate and computationally affordable results using an *ab initio* study of methane combustion. This lab fills in a weak point in most undergraduate chemistry curricula and provides an invaluable opportunity to teach students to think like computational scientists rather than mindlessly obtain data. A major benefit of this exercise is that its primary learning goals are extremely generalizeable to most computational sciences rather than constrained to a niche computational chemistry application. The lab is designed with optimal flexibility such that it can easily merge into existing physical chemistry courses. The open-source nature of Psi4 allows for this entire lab exercise to be completed on a personal laptop, ensuring no accessibility barriers are present. A Comprehensive guide and template input files are provided to ensure that Psi4 can be easily installed and employed by first time users. An example answer key is also provided to aid instructors in grading and providing hints for students. We also highlight numerous modifications that can extend the scope of the lab for advanced students or possibly simplify the content for a general chemistry audience. It is likely that the content of this lab could significantly bolster the preparedness of chemistry students entering any of the rapidly expanding computational science fields.

CHAPTER 6

CONCLUSION

This work encompasses high-level *ab initio* investigations of important chemical systems with a particular emphasis on electronic structure trends. We present the highest level *ab initio* determination of the hydrogen and halogen bonded complexes of hypohalous acid and water, with results nearly converged to the FCI/CBS limit. The hydrogen bonds are stronger in all cases and as halogen size increases, the halogen bonded complex becomes increasingly competitive (for the iodine case, the two interaction types differ by less than 1 kcal mol^{-1} at the CCSD(T)/aug-cc-pVTZ-X2C level of theory). Moreover, SAPT analysis reveals that the dispersion and induction percent contributions are together more important in the halogen bond, despite both types of interactions being primarily of electrostatic origin. The second project concerns the first theoretical study to date of the full $Pn(CH)_3$ (Pn = N, P, As, Sb, Bi) tetrahedrane series. All were determined to be stationary points at the CCSD(T)/aug-cc-pwCVTZ(-PP) level of theory, with the strain energy of each species decreasing as pnictogen size increases. The trends in geometry, dipole moment, NBO composition, and harmonic vibrational frequencies are characterized and discussed for the complete series. Electron delocalization from the Pn-C bonds into the C-H* orbitals is investigated using NBO methods and is determined to be an important feature of these systems. Replacing the hydrogens with a small set of substituents reveals a clear relationship between the Pn-C bond lengths and the extent of electron delocalization into substituent antibonding orbitals; this has consequence for selecting substituents for possible $Pn(CR)_3$ synthetic targets. The third project characterizes important stationary points on the $HNCO + H_2O$ potential energy surface with energetics approaching the CCSDT(Q)/CBS limit that can serve as benchmark data for future study of related systems. Composite methods are utilized to study the influence of substituents on the barriers towards carbamate and imidic acid formation. NBO predictions reveal strong associations between these barrier heights and the occupation of the C−O* isocyanate orbital. Similar trends are characterized with the inclusion of one and two catalyst water molecules, contrasted with the inclusion of a catalyst isocyanate molecule that manifests two-step pathways towards carbamate and imidic acid formation. Finally, we present a flexible and accessible undergraduate lab exercise that utilizes *ab initio* methods to predict the

methane combustion energy. The primary focus of the exercise is the balance between computational cost and accuracy, with learning goals that are transferable to any computational science field. This dissertation is composed of a diverse set of *ab initio* quantum chemistry applications, with many of the key insights related to chemical trends. The findings posses a high possibility of assisting and informing novel chemical research across various sub-disciplines.

Bibliography

[1] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications, Mineola, N.Y, Reprint Edition edn, 1996.

[2] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 2nd edn, 2004.

[3] L. Pauling and E. B. W. Jr, *Introduction to Quantum Mechanics with Applications to Chemistry*, Dover Publications, New York, N.Y, 3134th edn, 1985.

[4] I. Levine, *Quantum Chemistry*, Pearson, Boston, 7th edn, 2013.

[5] E. Schrödinger, *Phys. Rev.*, 1926, **28**, 1049–1070.

[6] J. C. Slater, *Phys. Rev.*, 1929, **34**, 1293–1322.

[7] J. Čížek, *J. Chem. Phys.*, 1966, **45**, 4256–4266.

[8] J. Čížek, *Adv. Chem. Phys.*, John Wiley & Sons, Ltd, 1969, pp. 35–89.

[9] J. Čižek and J. Paldus, *Int. J. Quantum Chem.*, 1971, **5**, 359–379.

[10] J. F. Stanton, *Chem. Phys. Lett.*, 1997, **281**, 130–134.

[11] T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

[12] D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1995, **103**, 4572–4585.

[13] T. H. Dunning, K. A. Peterson and A. K. Wilson, *J. Chem. Phys.*, 2001, **114**, 9244–9253.

[14] R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.

[15] E. Papajak and D. G. Truhlar, *J. Chem. Theory Comput.*, 2011, **7**, 10–18.

[16] K. A. Peterson and T. H. Dunning, *J. Chem. Phys.*, 2002, **117**, 10548–10560.

[17] K. A. Peterson, D. Figgen, E. Goll, H. Stoll and M. Dolg, *J. Chem. Phys.*, 2003, **119**, 11113–11123.

[18] M. S. Schuurman, S. R. Muir, W. D. Allen and H. F. Schaefer, *J. Chem. Phys.*, 2004, **120**, 11586–11599.

[19] J. M. Gonzales, C. Pak, R. S. Cox, W. D. Allen, H. F. Schaefer, A. G. Császár and G. Tarczay, *Chem.-Eur. J.*, 2003, **9**, 2173–2192.

[20] A. G. Császár, W. D. Allen and H. F. Schaefer, *J. Chem. Phys.*, 1998, **108**, 9751–9764.

[21] A. L. L. East and W. D. Allen, *J. Chem. Phys.*, 1993, **99**, 4638–4650.

[22] D. Feller, K. A. Peterson and T. D. Crawford, *J. Chem. Phys.*, 2006, **124**, 054107.

[23] T. Helgaker, W. Klopper, H. Koch and J. Noga, *J. Chem. Phys.*, 1997, **106**, 9639–9646.

[24] D. Feller, K. A. Peterson and T. D. Crawford, *J. Chem. Phys*, 2006, **124**, 054107.

[25] L. W. Bertels, J. Lee and M. Head-Gordon, *J. Chem. Theory Comput.*, 2021, **17**, 742–755.

[26] A. P. Rahalkar, B. K. Mishra, V. Ramanathan and S. R. Gadre, *Theor. Chem. Acc.*, 2011, **130**, 491–500.

[27] C. E. Warden, D. G. A. Smith, L. A. Burns, U. Bozkaya and C. D. Sherrill, *J. Chem. Phys.*, 2020, **152**, 124109.

[28] A. Sengupta, R. O. Ramabhadran and K. Raghavachari, *J. Comput. Chem.*, 2016, **37**, 286–295.

[29] M. A. Bartlett, A. H. Kazez, H. F. Schaefer and W. D. Allen, *J. Chem. Phys.*, 2019, **151**, 094304.

[30] B. Z. Abbott, P. R. Hoobler and H. F. Schaefer, *Phys. Chem. Chem. Phys.*, 2019, **21**, 26438–26452.

[31] F. Weinhold and C. R. Landis, *Chem. Edu. Res. Pract.*, 2001, **2**, 91–104.

[32] F. Weinhold, *J. Comput. Chem.*, 2012, **33**, 2363–2379.

[33] F. Weinhold and C. R. Landis, *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*, Cambridge University Press, 2005.

[34] E. D. Glendening, C. R. Landis and F. Weinhold, *Wires. Comput. Mol. Sci.*, 2012, **2**, 1–42.

[35] E. D. Glendening, C. R. Landis and F. Weinhold, *J. Comput. Chem.*, 2013, **34**, 1429–1437.

[36] P.-O. Löwdin, *Phys. Rev.*, 1955, **97**, 1474–1489.

[37] A. E. Reed, R. B. Weinstock and F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735–746.

[38] E. D. Glendening, J. K. Badenhoop and F. Weinhold, *J. Comput. Chem.*, 1998, **19**, 628–646.

[39] E. D. Glendening and F. Weinhold, *J. Comput. Chem.*, 1998, **19**, 610–627.

[40] E. D. Glendening and F. Weinhold, *J. Comput. Chem.*, 1998, **19**, 593–609.

[41] M. Wolf, B. Zhang, J. M. Turney and H. F. Schaefer, *Phys. Chem. Chem. Phys.*, 2019, **21**, 6160–6170.

[42] J. P. D. Abbatt, *Geophys. Res. Lett.*, 1994, **21**, 665–668.

[43] J. W. Adams, N. S. Holmes and J. N. Crowley, *Atmos. Chem. Phys.*, 2002, **2**, 79–91.

[44] S. Solomon, R. R. Garcia, F. Sherwood Rowland and D. Wuebbles, *Nature*, 1986, **321**, 755–758.

[45] Elliot, Scott, Cicerone, Ralph, Turco, Richard, Drdla, Katja and Tabazadeh, Azaddeh, *J. Geophys. Res.*, 1994.

[46] R. Hossaini, M. P. Chipperfield, S. A. Montzka, A. Rap, S. Dhomse and W. Feng, *Nat. Geosci.*, 2015, **8**, 186–190.

[47] K. A. Read, A. S. Mahajan, L. J. Carpenter, M. J. Evans, B. V. E. Faria, D. E. Heard, J. R. Hopkins, J. D. Lee, S. J. Moller, A. C. Lewis, L. Mendes, J. B. McQuaid, H. Oetjen, A. Saiz-Lopez, M. J. Pilling and J. M. C. Plane, *Nature*, 2008, **453**, 1232–1235.

[48] J. A. Schmidt, D. J. Jacob, H. M. Horowitz, L. Hu, T. Sherwen, M. J. Evans, Q. Liang, R. M. Suleiman, D. E. Oram, M. L. Breton, C. J. Percival, S. Wang, B. Dix and R. Volkamer, *J. Geophys. Res.-Atmos.*, 2016, **121**, 11,819–11,835.

[49] T. Sherwen, J. A. Schmidt, M. J. Evans, L. J. Carpenter, K. Grobmann, S. D. Eastham, D. J. Jacob, B. Dix, T. K. Koenig, R. Sinreich, I. Ortega, R. Volkamer, A. Saiz-Lopez, C. Prados-Roman, A. S. Mahajan and C. Ordonez, *Atmos. Chem. Phys.*, 2016, **16**, 12239–12271.

[50] T. K. Koenig, R. Volkamer, S. Baidar, B. Dix, S. Wang, D. C. Anderson, R. J. Salawitch, P. A. Wales, C. A. Cuevas, R. P. Fernandez, A. Saiz-Lopez, M. J. Evans, T. Sherwen, D. J. Jacob, J. Schmidt, D. Kinnison, J.-F. Lamarque, E. C. Apel, J. C. Bresch, T. Campos, F. M. Flocke, S. R. Hall, S. B. Honomichl, R. Hornbrook, J. B. Jensen, R. Lueb, D. D. Montzka, L. L. Pan, J. M. Reeves, S. M. Schauffler, K. Ullmann, A. J. Weinheimer, E. L. Atlas, V. Donets, M. A. Navarro, D. Riemer, N. J. Blake, D. Chen, L. G. Huey, D. J. Tanner, T. F. Hanisco and G. M. Wolfe, *Atmos. Chem. Phys.*, 2017, 1–46.

[51] W. R. Simpson, U. Frieß, M. E. Goodsite, D. Heard, M. Hutterli, A. Richter, H. Roscoe, R. Sander, P. Shepson, J. Sodeau, A. Steffen, T. Wagner and E. Wolff, *Atmos. Chem. Phys.*, 2007, 44.

[52] P. S. Monks, F. L. Nesbitt, M. Scanlon and L. J. Stief, *J. Phys. Chem.*, 1993, **97**, 11699–11705.

[53] R. Müller, J.-U. Grooß, A. M. Zafar, S. Robrecht and R. Lehmann, *Atmos. Chem. Phys.*, 2018, **18**, 2985–2997.

[54] W. A. Prütz, *Arch. Biochem. Biophys.*, 1999, **371**, 107–114.

[55] M. Whiteman, A. Jenner and B. Halliwell, *Chem.Res.Toxicol.*, 1997, **10**, 1240–1246.

[56] J. M. Albrich, C. A. McCarthy and J. K. Hurst, *Proc Natl Acad Sci U S A*, 1981, **78**, 210–214.

[57] W. A. Prütz, R. Kissner, T. Nauser and W. H. Koppenol, *Arch. Biochem. Biophys.*, 2001, **389**, 110–122.

[58] S. Sakarya, N. Gunay, B. Ozturk and B. Ertugrul, *Wounds*, 2014, **26**, 10.

[59] G. Bhave, C. F. Cummings, R. M. Vanacore, C. Kumagai-Cresse, I. A. Ero-Tolliver, M. Rafi, J.-S. Kang, V. Pedchenko, L. I. Fessler, J. H. Fessler and B. G. Hudson, *Nat. Chem. Biol.*, 2012, **8**, 784–790.

[60] E. L. Thomas, *Infect Immun*, 1979, **23**, 522–531.

[61] C. L. Perrin and J. B. Nielson, *Annu. Rev. Phys. Chem.*, 1997, **48**, 511–544.

[62] N. V. Belkova, E. S. Shubina and L. M. Epstein, *Acc. Chem. Res.*, 2005, **38**, 624–631.

[63] G. R. Desiraju, *Acc. Chem. Res.*, 2002, **35**, 565–573.

[64] K. Müller-Dethlefs and P. Hobza, *Chem. Rev.*, 2000, **100**, 143–168.

[65] T. Kato and J. M. J. Frechet, *J. Am. Chem. Soc.*, 1989, **111**, 8533–8534.

[66] A. M. S. Riel, D. A. Decato, J. Sun, C. J. Massena, M. J. Jessop and O. B. Berryman, *Chem. Sci.*, 2018, **9**, 5828–5836.

[67] Y. Kim, K. H. Kim and C.-Y. Park, *B. Kor. Chem. Soc.*, 2005, **26**, 1953–1961.

[68] T. S. Dibble and J. S. Francisco, *J. Phys. Chem.-US.*, 1995, **99**, 1919–1922.

[69] M. Ortiz-Repiso, R. Escribano and P. C. Gómez, *J. Phys. Chem. A*, 2000, **104**, 600–609.

[70] C. M. P. Santos, R. B. Faria, W. B. De Almeida, J. O. Machuca-Herrera and S. P. Machado, *Can. J. Chem.*, 2003, **81**, 961–970.

[71] G. A. Jeffrey, *An introduction to hydrogen bonding*, Oxford University Press, New York, 1997.

[72] S. Scheiner, *Hydrogen bonding: a theoretical perspective*, Oxford University Press, 1997.

[73] A. D. Buckingham, J. E. Del Bene and S. A. C. McDowell, *Chem. Phys. Lett.*, 2008, **463**, 1–10.

[74] K. B. Moore, K. Sadeghian, C. D. Sherrill, C. Ochsenfeld and H. F. Schaefer, *J. Chem. Theory Comput.*, 2017, **13**, 5379–5395.

[75] A. Priimagi, G. Cavallo, P. Metrangolo and G. Resnati, *Acc. Chem. Res.*, 2013, **46**, 2686–2695.

[76] A. C. Legon, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7736–7747.

[77] L. N. Anderson, F. W. Aquino, A. E. Raeber, X. Chen and B. M. Wong, *J. Chem. Theory Comput.*, 2018, **14**, 180–190.

[78] G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati and G. Terraneo, *Chem. Rev.*, 2016, **116**, 2478–2601.

[79] M. G. Chudzinski and M. S. Taylor, *J. Org. Chem.*, 2012, **77**, 3483–3491.

[80] M. H. Kolar and P. Hobza, *Chem. Rev.*, 2016, **116**, 5155–5187.

[81] S. Kozuch and J. M. L. Martin, *J. Chem. Theory Comput.*, 2013, **9**, 1918–1931.

[82] V. Oliveira, E. Kraka and D. Cremer, *Phys. Chem. Chem. Phys.*, 2016, **18**, 33031–33046.

[83] V. Oliveira and E. Kraka, *J. Phys. Chem. A*, 2017, **121**, 9544–9556.

[84] V. Oliveira, E. Kraka and D. Cremer, *Inorg. Chem.*, 2017, **56**, 488–502.

[85] V. Oliveira and D. Cremer, *Chem. Phys. Let.*, 2017, **681**, 56–63.

[86] J. Thirman, E. Engelage, S. M. Huber and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2018, **20**, 905–915.

[87] L. P. Wolters, P. Schyman, M. J. Pavan, W. L. Jorgensen, F. M. Bickelhaupt and S. Kozuch, *Wires. Comput. Mol. Sci.*, 2014, **4**, 523–540.

[88] L. P. Wolters and F. M. Bickelhaupt, *ChemistryOpen*, 2012, **1**, 96–105.

[89] A. Lange, J. Heidrich, M. O. Zimmermann, T. E. Exner and F. M. Boeckler, *J. Chem. Inf. Model.*, 2019.

[90] V. Angarov and S. Kozuch, *New J. Chem.*, 2018, **42**, 1413–1422.

[91] N. Galland, G. Montavon, J. L. Questel and J. Graton, *New J.Chem.*, 2018, **42**, 10510–10517.

[92] Z. P. Shields, J. S. Murray and P. Politzer, *Int. J. Quantum Chem.*, 2010, **110**, 2823–2832.

[93] Q.-Z. Li, B. Jing, R. Li, Z.-B. Liu, W.-Z. Li, F. Luan, J.-B. Cheng, B.-A. Gong and J.-Z. Sun, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2266–2271.

[94] Y. Geboes, F. De Proft and W. A. Herrebout, *Acta. Cryst. B*, 2017, **73**, 168–178.

[95] S. W. L. Hogan and T. v. Mourik, *J. Comput. Chem.*, 2016, **37**, 763–770.

[96] C. C. Robertson, J. S. Wright, E. J. Carrington, R. N. Perutz, C. A. Hunter and L. Brammer, *Chem. Sci.*, 2017, **8**, 5392–5398.

[97] M. Domagala, A. Lutynska and M. Palusiak, *J. Phys. Chem. A*, 2018, **122**, 5484–5492.

[98] K. V. Chance, D. G. Johnson and W. A. Traub, *J. Geophys. Res.*, 1989, **94**, 11059.

[99] G. C. Toon and C. B. Farmer, *Geophys. Res. Lett.*, 1989, **16**, 1375–1377.

[100] M. R. Le Breton, M. W. Gallagher, D. E. Shallcross, M. J. Evans, L. Carpenter, S. Andrews, R. T. Lidster, N. R. P. Harris and C. Percival, *AGU Fall Meeting Abstracts*, 2014, **23**, A23L–3434.

[101] M. Dorf, A. Butz, C. Camy-Peyret, M. P. Chipperfield, L. Kritten and K. Pfeilsticker, *Atmos. Chem. Phys.*, 2008, 7.

[102] E. H. Appelman, *Acc. Chem. Res.*, 1973, **6**, 113–117.

[103] Q. Li, X. Xu, T. Liu, B. Jing, W. Li, J. Cheng, B. Gong and J. Sun, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6837–6843.

[104] X. An, H. Zhuo, Y. Wang and Q. Li, *J Mol Model*, 2013, **19**, 4529–4535.

[105] A. Zabardasti, A. Sharifi-Rad and A. Kakanejadifard, *Spectrochim. Acta. A*, 2015, **151**, 746–759.

[106] A. Zabardasti, Y. A. Tyula and H. Goudarziafshar, *J. Sulfur Chem.*, 2017, **38**, 119–133.

[107] M. Solimannejad, I. Alkorta and J. Elguero, *Chem. Phys. Lett.*, 2007, **449**, 23–27.

[108] A. Kakanejadifard, S. Japelaghi, M. Ghasemian and A. Zabardasti, *Struct. Chem.*, 2015, **26**, 23–33.

[109] Q. Li, Q. Lin, W. Li, J. Cheng, B. Gong and J. Sun, *Chem. Phys. Chem.*, 2008, **9**, 2265–2269.

[110] Q.-Z. Li, R. Li, P. Guo, H. Li, W.-Z. Li and J.-B. Cheng, *Comput. Theor. Chem.*, 2012, **980**, 56–61.

[111] Q.-Z. Li, J.-L. Zhao, B. Jing, R. Li, W.-Z. Li and J.-B. Cheng, *J. Comput. Chem.*, **32**, 2432–2440.

[112] A. Zabardasti, Y. A. Tyula and H. Goudarziafshar, *B. Chem. Soc. Ethiopia*, 2017, **31**, 241–252.

[113] A. Zabaradsti, A. Kakanejadifard and M. Ghasemian, *Comput. Theor. Chem.*, 2012, **989**, 1–6.

[114] J. J. Panek and S. Berski, *Chem. Phys. Lett.*, 2008, **467**, 41–45.

[115] Z. Zhang, J. Shen, N. Jin, L. Chen and Z. Yang, *Comput. Theor. Chem.*, 2012, **999**, 48–54.

[116] L. Ying and X. Zhao, *J. Phys. Chem. A*, 1997, **101**, 3569–3573.

[117] C. M. P. Santos, R. Faria, S. P. Machado and W. B. De Almeida, *J. Chem. Phys.*, 2004, **121**, 141.

[118] S. Qiao, L. Zhen, Z. Xiao Qing, G. Mao Fa and W. Dian Xu, *Chinese J. Chem.*, 2005, **23**, 483–490.

[119] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, Cambridge University Press, 2009.

[120] R. J. Bartlett, J. D. Watts, S. A. Kucharski and J. Noga, *Chem. Phys. Lett.*, 1990, **165**, 513–522.

[121] K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479–483.

[122] M. E. Harding, T. Metzroth, J. Gauss and A. A. Auer, *J. Chem. Theory Comput.*, 2008, **4**, 64–74.

[123] J. F. Stanton, J. Gauss, L. Cheng, M. E. Harding, D. A. Matthews and P. G. Szalay, *CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package*, With contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch, K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de.

[124] Q. Sun, W. Liu, Y. Xiao and L. Cheng, *J. Chem. Phys.*, 2009, **131**, 081101.

[125] L. Cheng and J. Gauss, *J. Chem. Phys.*, 2011, **135**, 084114.

[126] M. Iliaš and T. Saue, *J. Chem. Phys.*, 2007, **126**, 064102.

[127] `http://slater.chemie.uni-mainz.de/cfour/index.php?n=Main.RecontractedCorrelation-consistentBasisFu`

[128] K. A. Peterson, C. Krause, H. Stoll, J. G. Hill and H.-J. Werner, *Molecular Physics*, 2011, **109**, 2607–2623.

[129] M. K. Kesharwani, D. Manna, N. Sylvetsky and J. M. L. Martin, *J. Phys. Chem. A*, 2018, **122**, 2184–2197.

[130] N. J. DeYonker, K. A. Peterson and A. K. Wilson, *J. Phys. Chem. A*, 2007, **111**, 11383–11393.

[131] N. C. Handy, Y. Yamaguchi and H. F. Schaefer, *J. Chem. Phys.*, 1986, **84**, 4481.

[132] H. Sellers and P. Pulay, *Chem. Phys. Lett.*, 1984, **103**, 463–465.

[133] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kús, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio Jr., H. Dop, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharmaa, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhou, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xua, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Mol. Phys.*, 2015, **113**, 184–215.

[134] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

[135] E. Willighagen, E. Willighagen and M. Howard, *Nature Precedings*, 2007.

[136] E. G. Hohenstein and C. D. Sherrill, *J Chem Phys*, 2010, **133**, 014101.

[137] B. Jeziorski, R. Moszynski and K. Szalewicz, *Chem. Rev.*, 1994, **94**, 1887–1930.

[138] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, *J. Chem. Theory Comput.*, 2017, **13**, 3185–3197.

[139] P. Politzer, J. S. Murray and M. C. Concha, *J. Mol. Model.*, 2007, **13**, 643–650.

[140] E. G. Hohenstein and C. D. Sherrill, *Wires. Comput. Mol. Sci.*, 2012, **2**, 304–326.

[141] K. E.Riley and P. Hobza, *Phys. Chem. Chem. Phys.*, 2013, **15**, 17742–17751.

[142] K. E. Riley, J. S. Murray, J. Fanfrlik, J. Rezac, R. J. Sola, M. C. Concha, F. M. Ramos and P. Politzer, *J. Mol. Model*, 2013, **19**, 4651–4659.

[143] P. Politzer, K. E. Riley, F. A. Bulat and J. S. Murray, *Comput.Theor. Chem.*, 2012, **998**, 2–8.

[144] A. J. Stone, *Journal of the American Chemical Society*, 2013, **135**, 7005–7009.

[145] M. E. Wolf, E. A. Doty, J. M. Turney and H. F. Schaefer III, *J. Phys. Chem. A.*, 2021, **125**, 2612–2621.

[146] M. D. Newton, *Applications of Electronic Structure Theory*, Springer US, Boston, MA, 1977, pp. 223–275.

[147] V. Balaji and J. Michl, *Pure. Appl. Chem.*, 1988, **60**, 189–194.

[148] C.-Y. Zhao, Y. Zhang and X.-Z. You, *J. Phys. Chem. A*, 1997, **101**, 5174–5182.

[149] W. Adcock, M. J. Brunger, M. T. Michalewicz and D. A. Winkler, *Aust. J. Phys.*, 1998, **51**, 707–728.

[150] O. Brummel, D. Besold, T. Döpper, Y. Wu, S. Bochmann, F. Lazzari, F. Waidhas, U. Bauer, P. Bachmann, C. Papp and et al., *Chem. Sus. Chem.*, 2016, **9**, 1424–1432.

[151] K. Chen, X. Huang, S. B. J. Kan, R. K. Zhang and F. H. Arnold, *Science*, 2018, **360**, 71–75.

[152] *IUPAC Compendium of Chemical Terminology: Gold Book*, ed. M. Nič, J. Jirát, B. Košata, A. Jenkins and A. McNaught, IUPAC, Research Triagle Park, NC, 2nd edn, 2009.

[153] M.-L. Y. Riu, R. L. Jones, W. J. Transue, P. Müller and C. C. Cummins, *Sci. Adv.*, 2020, **6**, eaaz3168.

[154] A. Greenberg and J. F. Liebman, *Strained Organic Molecules: Organic Chemistry: A Series of Monographs, Vol. 38*, Academic Press, 2013.

[155] H. Hopf and 1940, *Classics in hydrocarbon chemistry*, Wiley-VCH, 2000.

[156] A. Nemirowski, H. P. Reisenauer and P. R. Schreiner, *Chem-Eur. J.*, 2006, **12**, 7411–7420.

[157] S. Nagase, M. Nakano and T. Kudo, *J. Chem. Soc., Chem. Commun.*, 1987, 60–62.

[158] R. Haunschild and G. Frenking, *Mol. Phys.*, 2009, **107**, 911–922.

[159] S. Rayne and K. Forest, *Nat. Preced.*, 2011, 1–1.

[160] I. Alkorta, J. Elguero, I. Rozas and A. T. Balaban, *J. Mol. Struct.*, 1990, **208**, 63–77.

[161] N. K. V. Monteiro, J. F. d. Oliveira and C. L. Firme, *New J. Chem.*, 2014, **38**, 5892–5904.

[162] S. Rayne and K. Forest, *Propell. Explos. Pyrot.*, 2011, **36**, 410–415.

[163] G. N. Srinivas and E. D. Jemmis, *J. Am. Chem. Soc.*, 1997, **119**, 12968–12973.

[164] B. M. Cossairt, M.-C. Diawara and C. C. Cummins, *Science*, 2009, **323**, 602–602.

[165] G. Hierlmeier, P. Coburger, M. Bodensteiner and R. Wolf, *Agnew. Chem. Int. Edit.*, 2019, **58**, 16918–16922.

[166] G. Maier, S. Pfriem, U. Schäfer and R. Matusch, *Agnew. Chem. Int. Edit.*, 1978, **17**, 520–521.

[167] G. Maier, J. Neudert, O. Wolf, D. Pappusch, A. Sekiguchi, M. Tanaka and T. Matsuo, *J. Am. Chem. Soc.*, 2002, **124**, 13819–13826.

[168] T. Ochiai, M. Nakamoto, Y. Inagaki and A. Sekiguchi, *J. Am. Chem. Soc.*, 2011, **133**, 11504–11507.

[169] M. Nakamoto, Y. Inagaki, M. Nishina and A. Sekiguchi, *J. Am. Chem. Soc.*, 2009, **131**, 3172–3173.

[170] J. L. Menke, E. V. Patterson and R. J. McMahon, *J. Phys. Chem. A*, 2010, **114**, 6431–6437.

[171] M. Balci, M. L. McKee and P. v. R. Schleyer, *J. Phys. Chem. A*, 2000, **104**, 1246–1255.

[172] G. Gayatri, Y. Soujanya, I. Fernández, G. Frenking and G. N. Sastry, *J. Phys. Chem. A*, 2008, **112**, 12919–12924.

[173] I. Novak, *Chem. Phys. Lett.*, 2003, **380**, 258–262.

[174] P. R. Schreiner, L. V. Chernish, P. A. Gunchenko, E. Y. Tikhonchuk, H. Hausmann, M. Serafin, S. Schlecht, J. E. P. Dahl, R. M. K. Carlson and A. A. Fokin, *Nature*, 2011, **477**, 308–311.

[175] A. S. Ivanov, K. V. Bozhenko and A. I. Boldyrev, *J. Chem. Theory Comput.*, 2012, **8**, 135–140.

[176] M. E. Lahm, P. R. Hoobler, J. M. Turney, K. A. Peterson and H. F. Schaefer, *Phys. Chem. Chem. Phys.*, 2018, **20**, 21881–21889.

[177] D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1993, **98**, 1358–1371.

[178] T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

[179] D. Feller, *J. Comput. Chem.*, 1996, **17**, 1571–1586.

[180] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li and T. L. Windus, *J. Chem. Inf. Model.*, 2007, **47**, 1045–1052.

[181] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibsom and T. L. Windus, *J. Chem. Inf. Model.*, 2019, **59**, 4814–4820.

[182] K. A. Peterson and K. E. Yousaf, *J. Chem. Phys.*, 2010, **133**, 174116.

[183] B. Metz, H. Stoll and M. Dolg, *J. Chem. Phys.*, 2000, **113**, 2563–2569.

[184] S. Rayne and K. Forest, *Theor. Chem. Acc.*, 2010, **127**, 697–709.

[185] S. E. Wheeler, K. N. Houk, P. v. R. Schleyer and W. D. Allen, *J. Am. Chem. Soc.*, 2009, **131**, 2547–2560.

[186] J. Howell, J. D. Goddard and W. Tam, *Tetrahedron*, 2009, **65**, 4562–4568.

[187] R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.

[188] A. K. Wilson, D. E. Woon, K. A. Peterson and T. H. Dunning, *J. Chem. Phys.*, 1999, **110**, 7667–7676.

[189] A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

[190] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.

[191] Y. Zhao, N. E. Schultz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2006, **2**, 364–382.

[192] A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

[193] H. S. Yu, X. He, S. L. Li and D. G. Truhlar, *Chem Sci*, 2016, **7**, 5032–5051.

[194] J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.

[195] C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.

[196] D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. Schaefer, A. Y. Sokolov, K. Patkowski, A. E. DePrince, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, *J. Chem. Phys.*, 2020, **152**, 184108.

[197] G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, T. Harville, S. Irle, J. Ivanic, K. Kowalski, S. S. Leang, H. Li, W. Li, J. J. Lutz, I. Magoulas, J. Mato, V. Mironov, H. Nakata, B. Q. Pham, P. Piecuch, D. Poole, S. R. Pruitt, A. P. Rendell, L. B. Roskop, K. Ruedenberg, T. Sattasathuchana, M. W. Schmidt, J. Shen, L. Slipchenko, M. Sosonkina, V. Sundriyal, A. Tiwari, J. L. Galvez Vallejo, B. Westheimer, M. Wloch, P. Xu, F. Zahariev and M. S. Gordon, *J. Chem. Phys.*, 2020, **152**, 154102.

[198] B. Metz, H. Stoll and M. Dolg, *J. Chem. Phys.*, 2000, **113**, year.

[199] F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, year.

[200] J. E. Del Bene, I. Alkorta and J. Elguero, *Molecules*, 2020, **25**, 2846.

[201] B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán and S. Alvarez, *Dalton T.*, 2008, 2832–2838.

[202] E. D. Jemmis and P. v. R. Schleyer, *J. Am. Chem. Soc.*, 1982, **104**, 4781–4788.

[203] M. Elian and R. Hoffmann, *Inorg. Chem.*, 1975, **14**, 1058–1076.

[204] K. B. Wiberg, R. F. W. Bader and C. D. H. Lau, *J. Am. Chem. Soc.*, 1987, **109**, 1001–1012.

[205] M. Nakamoto, Y. Inagaki, T. Ochiai, M. Tanaka and A. Sekiguchi, *Heteroatom Chem.*, 2011, **22**, 412–416.

[206] G. Zhou, J.-L. Zhang, N.-B. Wong and A. Tian, *J. Mol. Struct*, 2004, **668**, 189–195.

[207] M. Wolf, J. E. Vandezande and H. F. Schaefer, *Phys. Chem. Chem. Phys.*, 2021, **23**, 18535–18546.

[208] C. Ludwig Schmidt and M. Jansen, *J. Mater. Chem.*, 2010, **20**, 110–116.

[209] E. Demir, O. Sari, Y. Çetinkaya, U. Atmaca, S. S. Erdem and M. Çelik, *Beilstein J. Org. Chem.*, 2020, **16**, 1805–1819.

[210] D. Quan, E. Herbst, Y. Osamura and E. Roueff, *Astrophys. J*, 2010, **725**, 2101.

[211] Y. Qin, B. Lu, G. Rauhut, M. Hagedorn, K. Banert, C. Song, X. Chu, L. Wang and X. Zeng, *Angew. Chem. Int. Edit.*, 2019, **58**, 17277–17281.

[212] L. Kolesniková, Z. Kisiel, E. R. Alonso, J. C. Guillemin, J. L. Alonso, I. R. Medvedev and M. Winnewisser, *Astrophys. J. Suppl. S.*, 2019, **245**, 31.

[213] R. Martín-Doménech, V. M. Rivilla, I. Jiménez-Serra, D. Quénard, L. Testi and J. Martín-Pintado, *Mon. Not. R. Astron. Soc.*, 2017, **469**, 2230–2234.

[214] W.-H. Fang and G.-Q. Shao, *J. Mol. Struct-Theochem.*, 1996, **369**, 183–188.

[215] W. Sun, W. G. D. P. Silva and J. van Wijngaarden, *J. Phys. Chem. A*, 2019, **123**, 2351–2360.

[216] G. A. Martínez Córdoba, L. A. Ramos, S. E. Ulic, J. L. Jios, C. O. DellaVédova, J. Pepino, M. A. Burgos Paci, G. A. Argüello, M. Ge, H. Beckers and H. Willner, *J. Phys. Chem. A*, 2011, **115**, 8608–8615.

[217] H.-J. Knölker, T. Braxmeier and H. Oberhammer, *J. Mol. Struct.*, 1997, **413-414**, 211–216.

[218] M. E. Bailey, V. Kirss and R. G. Spaunburgh, *Ind. Eng. Chem.*, 1956, **48**, 794–797.

[219] D.-C. Fang and X.-Y. Fu, *J. Mol. Struct-Theochem.*, 1999, **459**, 15–21.

[220] S. G. Entelis and O. V. Nesterov, *Russ. Chem. Rev.*, 1966, **35**, 917–930.

[221] T. M. Klapötke and A. Schulz, *Inorg. Chem.*, 1996, **35**, 7897–7904.

[222] K. Yokoyama, S.-y. Takane and T. Fueno, *B. Chem. Soc. Jpn.*, 1991, **64**, 2230–2235.

[223] J. H. Teles, G. Maier, B. A. Hess, L. J. Schaad, M. Winnewisser and B. P. Winnewisser, *Chemische Berichte*, 1989, **122**, 753–766.

[224] M. D. Leslie, M. Ridoli, J. G. Murphy and N. Borduas-Dedekind, *Environ. Sci-Prod. Imp.*, 2019, **21**, 793–808.

[225] J. M. Roberts, P. R. Veres, A. K. Cochran, C. Warneke, I. R. Burling, R. J. Yokelson, B. Lerner, J. B. Gilman, W. C. Kuster, R. Fall and J. d. Gouw, *P. Natl. Acad. Sci. USA*, 2011, **108**, 8966–8971.

[226] R. F. Hems, C. Wang, D. B. Collins, S. Zhou, N. Borduas-Dedekind, J. A. Siegel and J. P. D. Abbatt, *Environ. Sci-Prod. Imp.*, 2019, **21**, 1334–1341.

[227] S. Rosanka, G. H. T. Vu, H. M. T. Nguyen, T. V. Pham, U. Javed, D. Taraborrelli and L. Vereecken, *Atmos. Chem. Phys.*, 2020, **20**, 6671–6686.

[228] Z. Wang, B. Yuan, C. Ye, J. Roberts, A. Wisthaler, Y. Lin, T. Li, C. Wu, Y. Peng, C. Wang, S. Wang, S. Yang, B. Wang, J. Qi, C. Wang, W. Song, W. Hu, X. Wang, W. Xu, N. Ma, Y. Kuang, J. Tao, Z. Zhang, H. Su, Y. Cheng, X. Wang and M. Shao, *Environ. Sci. Technol.*, 2020, **54**, 11818–11826.

[229] P. Glarborg, P. G. Kristensen, S. H. Jensen and K. Dam-Johansen, *Combust. Flame*, 1994, **98**, 241–258.

[230] R. A. Perry and D. L. Siebers, *Nature*, 1986, **324**, 657–658.

[231] L. Majumdar, J.-C. Loison, M. Ruaud, P. Gratier, V. Wakelam and A. Coutens, *Mon. Not. R. Astron. Soc. Let.*, 2018, **473**, L59–L63.

[232] M. Ferus, V. Laitl, A. Knizek, P. Kubelík, J. Sponer, J. Kára, J. E. Sponer, B. Lefloch, G. Cassone and S. Civiš, *Astron. Astrophys.*, 2018, **616**, A150.

[233] K. A. Haupa, G. Tarczay and Y.-P. Lee, *J. Am. Chem. Soc.*, 2019, **141**, 11614–11620.

[234] A. Coutens, J. K. Jørgensen, M. H. D. v. d. Wiel, H. S. P. Müller, J. M. Lykke, P. Bjerkeli, T. L. Bourke, H. Calcutt, M. N. Drozdovskaya, C. Favre, E. C. Fayolle, R. T. Garrod, S. K. Jacobsen, N. F. W. Ligterink, K. I. Öberg, M. V. Persson, E. F. v. Dishoeck and S. F. Wampfler, *Astron. Astrophys.*, 2016, **590**, L6.

[235] T. Wang and J. H. Bowie, *Org. Biomol. Chem.*, 2012, **10**, 652–662.

[236] F. S. Brigiano, Y. Jeanvoine, A. Largo and R. Spezia, *Astron. Astrophys.*, 2018, **610**, A26.

[237] F. W. Hoover and H. S. Rothrock, *J. Org. Chem.*, 1964, **29**, 143–145.

[238] J. E. Rode and J. C. Dobrowolski, *J. Phys. Chem. A*, 2006, **110**, 3723–3737.

[239] A. S. Narula and K. Ramachandran, *Encyclopedia of Reagents for Organic Synthesis*, American Cancer Society, 2001.

[240] T. Uchimaru, S. Yamane, J. Mizukado and S. Tsuzuki, *RSC Advances*, 2020, **10**, 15955–15965.

[241] P. I. Kordomenos and J. E. Kresta, *Macromolecules*, 1981, **14**, 1434–1437.

[242] P. I. Kordomenos, J. E. Kresta and K. C. Frisch, *Macromolecules*, 1987, **20**, 2077–2083.

[243] H. Ni, A. D. Skaja, R. A. Sailer and M. D. Soucek, *Macromol. Chem. Phys.*, 2000, **201**, 722–732.

[244] M. A. Bahili, E. C. Stokes, R. C. Amesbury, D. M. C. Ould, B. Christo, R. J. Horne, B. M. Kariuki, J. A. Stewart, R. L. Taylor, P. Andrew Williams, M. D. Jones, K. D. M. Harris and B. D. Ward, *Chem. Commun.*, 2019, **55**, 7679–7682.

[245] X. Xu, X. Ni and Z. Shen, *Polym. Bull.*, 2005, **53**, 81–88.

[246] F. Sanda, T. Takata and T. Endo, *J. Polym. Sci. Pol. Chem.*, 1995, **33**, 2353–2358.

[247] A. A. Nabulsi, D. Cozzula, T. Hagen, W. Leitner and T. E. Müller, *Polym. Chem-UK*, 2018, **9**, 4891–4899.

[248] F. E. Golling, R. Pires, A. Hecking, J. Weikard, F. Richter, K. Danielmeier and D. Dijkstra, *Polym. Int.*, 2019, **68**, 848–855.

[249] M. Çoban and F. A. S. Konuklar, *Comput. Theor. Chem.*, 2011, **963**, 168–175.

[250] J. Nagy, E. Pusztai and O. Wagner, *Eur. Chem. Bull.*, 2013, **2**, 985–992.

[251] A. Y. Samuilov, L. A. Zenitova, Y. D. Samuilov and A. I. Konovalov, *Russ. J. Org. Chem.*, 2008, **44**, 1316.

[252] G. Raspoet, M. T. Nguyen, M. McGarraghy and A. F. Hegarty, *J. Org. Chem.*, 1998, **63**, 6878–6885.

[253] C. A. Tsipis and P. A. Karipidis, *J. Am. Chem. Soc.*, 2003, **125**, 2307–2318.

[254] A. Nicolle, S. Cagnina and T. de Bruin, *Chem. Phys. Lett.*, 2016, **664**, 149–153.

[255] G. Raspoet, M. T. Nguyen, M. McGarraghy and A. F. Hegarty, *J. Org. Chem.*, 1998, **63**, 6867–6877.

[256] W. Cheikh, Z. B. Rózsa, C. O. Camacho López, P. Mizsey, B. Viskolcz, M. Szőri and Z. Fejes, *Polymers*, 2019, **11**, 1543.

[257] C. Gertig, E. Erdkamp, A. Ernst, C. Hemprich, L. C. Kröger, J. Langanke, A. Bardow and K. Leonhard, *ChemistryOpen*, 2021, open.202000150.

[258] X.-G. Wei, X.-M. Sun, X.-P. Wu, S. Geng, Y. Ren, N.-B. Wong and W.-K. Li, *J. Mol. Model.*, 2011, **17**, 2069–2082.

[259] H. Sardon, A. C. Engler, J. M. W. Chan, J. M. García, D. J. Coady, A. Pascual, D. Mecerreyes, G. O. Jones, J. E. Rice, H. W. Horn and J. L. Hedrick, *J. Am. Chem. Soc.*, 2013, **135**, 16235–16241.

[260] H. Sardon, A. Pascual, D. Mecerreyes, D. Taton, H. Cramail and J. L. Hedrick, *Macromolecules*, 2015, **48**, 3153–3165.

[261] E. Levent, O. Sala, L. F. B. Wilm, P. Löwe and F. Dielmann, *Green Chem.*, 2021, **23**, 2747–2755.

[262] R. K. Sahoo, N. Sarkar and S. Nembenna, *Angew. Chem.*, 2021, **133**, 12098–12107.

[263] R. Devendra, N. R. Edmonds and T. Söhnel, *J. Mol. Catal. A-Chem.*, 2013, **366**, 126–139.

[264] R. Devendra, N. R. Edmonds and T. Söhnel, *RSC Advances*, 2015, **5**, 48935–48945.

[265] M. A. McAllister and T. T. Tidwell, *J. CHEM. SOC. PERKIN TRANS.*, 1994, 10.

[266] A. Y. Samuilov, L. A. Zenitova, Y. D. Samuilov and A. I. Konovalov, *Russ. J. Org. Chem.*, 2009, **45**, 68–73.

[267] A. Y. Samuilov, F. B. Balabanova, T. A. Kamalov, Y. D. Samuilov and A. I. Konovalov, *Russ. J. Org. Chem.*, 2010, **46**, 1452–1460.

[268] A. Y. Samuilov, T. A. Kamalov, F. B. Balabanova, Y. D. Samuilov and A. I. Konovalov, *Russ. J. Org. Chem.*, 2012, **48**, 158–163.

[269] A. Y. Samuilov, F. B. Balabanova, Y. D. Samuilov and A. I. Konovalov, *Russ. J. Org. Chem.*, 2012, **48**, 164–174.

[270] A. Y. Samuilov, F. B. Balabanova, Y. D. Samuilov and A. I. Konovalov, *Russ. J. Org. Chem.*, 2012, **48**, 1512–1517.

[271] A. Y. Samuilov, S. V. Nesterov, F. B. Balabanova, Y. D. Samuilov and A. I. Konovalov, *Russ. J. Org. Chem.*, 2013, **49**, 968–973.

[272] A. Y. Samuilov, F. B. Balabanova, Y. D. Samuilov and A. I. Konovalov, *Russ. J. Org. Chem.*, 2013, **49**, 22–27.

[273] A. Y. Samuilov, S. V. Nesterov, F. B. Balabanova, Y. D. Samuilov and A. I. Konovalov, *Russ. J. Org. Chem.*, 2014, **50**, 155–159.

[274] Y. Zhao and G. J. Suppes, *Polym. Eng. Sci.*, 2015, **55**, 1420–1428.

[275] F. Kössl, M. Lisaj, V. Kozich, K. Heyne and O. Kühn, *Chem. Phys. Lett.*, 2015, **621**, 41–45.

[276] P. Cysewski, P. Król and A. Shyichuk, *Macromol. Theor. Simul.*, 2007, **16**, 541–547.

[277] T. H. Dunning, K. A. Peterson and A. K. Wilson, *J. Chem. Phys.*, 2001, **114**, 9244–9253.

[278] D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1993, **98**, 1358–1371.

[279] T. Helgaker, W. Klopper, H. Koch and J. Noga, *J. Chem. Phys.*, 1997, **106**, 9639–9646.

[280] J. Noga and R. J. Bartlett, *J. Chem . Phys*, 1987, **86**, 7041–7050.

[281] G. E. Scuseria and H. F. Schaefer, *Chem. Phys. Lett.*, 1988, **152**, 382–386.

[282] Y. J. Bomble, J. F. Stanton, M. Kállay and J. Gauss, *J. Chem . Phys*, 2005, **123**, 054101.

[283] M. Kállay and J. Gauss, *J. Chem . Phys*, 2008, **129**, 144101.

[284] K. G. Dyall, *J. Chem . Phys*, 1997, **106**, 9618–9626.

[285] W. Kutzelnigg and W. Liu, *J. Chem . Phys*, 2005, **123**, 241102.

[286] W. Liu and D. Peng, *J. Chem . Phys*, 2006, **125**, 044102.

[287] W. Liu and D. Peng, *J. Chem . Phys*, 2009, **131**, 031104.

[288] J. Sikkema, L. Visscher, T. Saue and M. Iliaš, *J. Chem . Phys*, 2009, **131**, 124116.

[289] H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. Heßelmann, D. Kats, A. Köhn, T. Korona, D. A. Kreplin, Q. Ma, T. F. Miller, A. Mitrushchenkov, K. A. Peterson, I. Polyak, G. Rauhut and M. Sibaev, *J. Chem . Phys*, 2020, **152**, 144107.

[290] G. W. Spitznagel, T. Clark, P. v. R. Schleyer and W. J. Hehre, *J. Comput. Chem.*, 1987, **8**, 1109–1116.

[291] W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem . Phys*, 1972, **56**, 2257–2261.

[292] P. C. Hariharan and J. A. Pople, *Theor. chim. acta*, 1973, **28**, 213–222.

[293] M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro and W. J. Hehre, *J. Am. Chem. Soc.*, 1982, **104**, 2797–2803.

[294] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *J. Chem . Phys*, 1982, **77**, 3654–3665.

[295] T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. R. Schleyer, *J. Comput. Chem.*, 1983, **4**, 294–301.

[296] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. D. Jr, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. V. Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Molecular Physics*, 2015, **113**, 184–215.

[297] D. Rappoport and F. Furche, *J. Chem . Phys*, 2010, **133**, 134105.

[298] P. V. Bharatam, R. Moudgil and D. Kaur, *J. Phys. Chem. A*, 2003, **107**, 1627–1634.

[299] X. Lin, X. Li and X. Lin, *Molecules*, 2020, **25**, 1375.

[300] S. J. Y. Macalino, V. Gosu, S. Hong and S. Choi, *Arch. Pharm. Res.*, 2015, **38**, 1686–1701.

[301] M. Elkin and T. R. Newhouse, *Chem. Soc. Rev.*, 2018, **47**, 7830–7844.

[302] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa and A. Fazzio, *J. Phys. Mat.*, 2019, **2**, 032001.

[303] M. E. Wolf, J. E. Vandezande and H. F. Schaefer, *Phys. Chem. Chem. Phys.*, 2021, **23**, 18535–18546.

[304] P. R. Franke, J. T. Brice, C. P. Moradi, H. F. Schaefer and G. E. Douberly, *J. Phys. Chem. A.*, 2019, **123**, 3558–3568.

[305] P. R. Franke and G. E. Douberly, *J. Phys. Chem. A.*, 2018, **122**, 148–158.

[306] S. E. Feller, R. F. Dallinger and P. C. McKinney, *J. Chem. Ed.*, 2004, **81**, 283.

[307] B. J. Esselman and N. J. Hill, *J. Chem. Ed.*, 2016, **93**, 932–936.

[308] B. Albrecht, *J. Chem. Ed.*, 2014, **91**, 2182–2185.

[309] C. D. Montgomery, *J. Chem. Ed.*, 2013, **90**, 661–664.

[310] M. E. Karpen, J. Henderleiter and S. A. Schaertel, *J. Chem. Ed.*, 2004, **81**, 475.

[311] L. S. Lever, J. J. Howe and D. M. Whisnant, *J. Chem. Ed.*, 2000, **77**, 1648.

[312] J. J. Howe, L. S. Lever and D. M. Whisnant, *J. Chem. Ed.*, 2000, **77**, 199.

[313] R. P. Orenha and S. E. Galembeck, *J. Chem. Ed.*, 2014, **91**, 1064–1069.

[314] B. J. Esselman and N. J. Hill, *J. Chem. Ed.*, 2015, **92**, 660–663.

[315] A. M. Halpern, *J. Chem. Ed.*, 2010, **87**, 174–179.

[316] R. L. Dekock, J. D. Madura, F. Rioux and J. Casanova, *Reviews in Computational Chemistry*, John Wiley & Sons, Ltd, 1993, pp. 149–228.

[317] N. Trinajstić, *Journal of Molecular Structure: THEOCHEM*, 1989, **200**, 219–224.

[318] E. R. Davidson, *Reviews in Computational Chemistry*, John Wiley & Sons, Ltd, 1990, pp. 373–382.

[319] J. K. Pearson, *J. Chem. Ed.*, 2007, **84**, 1323.

[320] P. B. Changala, T. L. Nguyen, J. H. Baraban, G. B. Ellison, J. F. Stanton, D. H. Bross and B. Ruscic, *J. Phys. Chem. A*, 2017, **2017**, year.

APPENDIX

6.1 Chapter 1

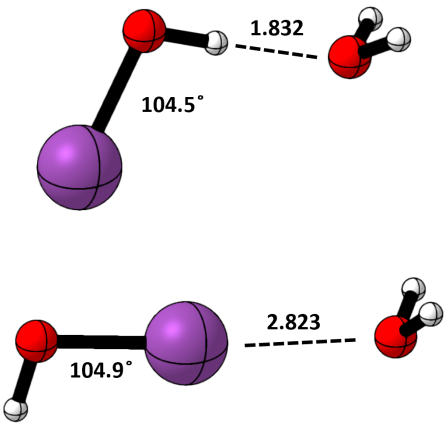

Figure 6.1: The Iodine binary complexes optimized at the CCSD(T)/aug-cc-pVTZ-X2C level of theory. Bond distances are in Angstroms.

Table 6.1: Harmonic vibrational frequencies in cm $^{-1}$ for all binary complex structures. Each structure is C_S symmetry with each vibrational modes corresponding to either the a′ or a″ irreducible representation.

Symmetry	Number	FOH (HB)	ClOH (HB)	BrOH (HB)	IOH (HB)	HOCl(XB)	HOBr (XB)	HOI (XB)
				syn-				
a′	1	3809	3805				3903	
	2	3570	3582				3796	
	3	1646	1646				1644	
	4	1478	1382				1173	
	5	913	734				622	
	6	233	246				220	
	7	164	193				126	
	8	85	73				88	
a″	9	3916	3911				3903	
	10	701	710				272	
	11	212	212				85	
	12	40	24				27	
				anti-				
a′	1	3799	3799	3802	3800	3800	3794	3797
	2	3523	3552	3592	3607	3783	3786	3791
	3	1650	1649	1647	1647	1646	1645	1645
	4	1499	1396	1315	1239	1253	1174	1082
	5	908	731	637	594	721	622	577
	6	310	295	242	243	205	244	262
	7	237	228	190	188	112	129	134
	8	83	74	63	59	86	91	91
a″	9	3903	3902	3906	3904	3908	3900	3897
	10	746	735	698	689	225	279	314
	11	227	227	206	199	77	86	87
	12	42	14	28	31	71	72	73

Table 6.2: Estimated equilibrium constants for *anti* HOX and water complexes using 200 K and 298 K temperatures as bounds for stratosphere and troposphere temperatures. Electronic energies are at the CCSDT(Q)/CBS level of theory for X = F, Cl, Br and CCSD(T)/aug-cc-pVTZ-X2C level of theory for X = I. All thermodynamic corrections are at the CCSD(T)/aug-cc-pVTZ-X2C level of theory.

Complex	200 K K_{eq}	298 K K_{eq}
$FOH\cdots H_2O$	2.9×10^{1}	2.5×10^{-1}
$ClOH\cdots H_2O$	2.6×10^{1}	2.5×10^{-1}
$BrOH\cdots H_2O$	1.1×10^{1}	1.3×10^{-1}
$IOH\cdots H_2O$	1.2×10^{1}	1.4×10^{-1}
$HOCl\cdots H_2O$	1.5×10^{-3}	3.8×10^{-4}
$HOBr\cdots H_2O$	1.8×10^{-2}	1.4×10^{-3}
$HOI\cdots H_2O$	4.5×10^{-1}	1.0×10^{-2}

6.2 Chapter 2

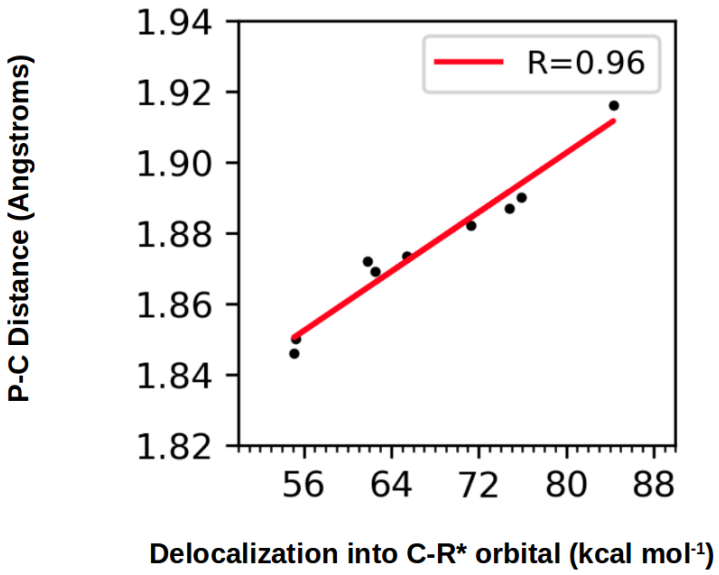

Figure 6.2: Scatterplot of the NBO deletion energy corresponding to the sum of the deletion for all C-C → C-H* and P-C → C-H* delocalizations vs. the P-C bond distance for $P(CR)_3$ (R = H, *t*-Bu, F, Cl, CH_3, CCH, NH_2, CF_3, and CCF) structures optimized at the MP2/aug-cc-pV(T+d)Z level of theory.

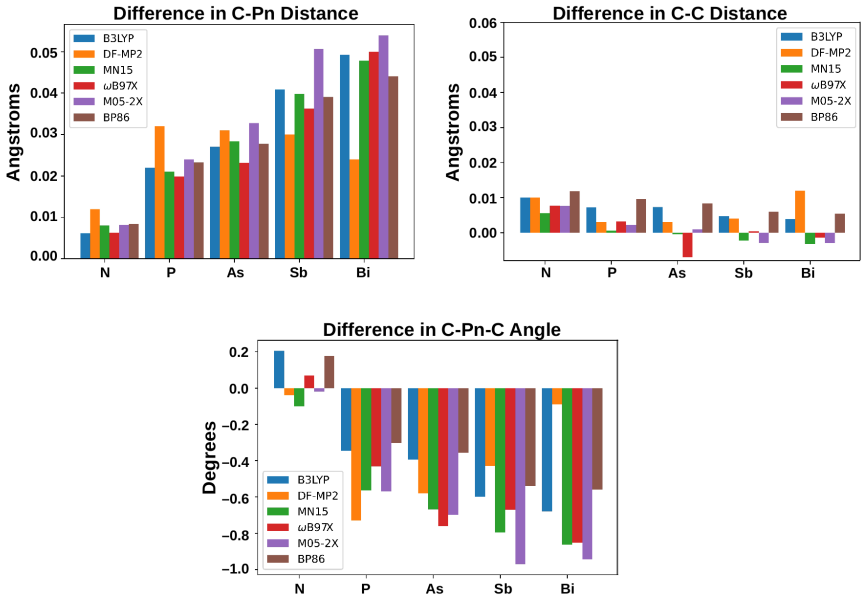

Figure 6.3: Changes in key geometric parameters upon *t*-Bu substitution relative to the parent tetrahedrane with various methods. An aug-cc-pV(T+d)Z basis set describes the central tetrhedrane moiety and the adjacent *t*-Bu carbons, cc-pVTZ on the other carbons, and cc-pVDZ on the *t*-Bu hydrogens.

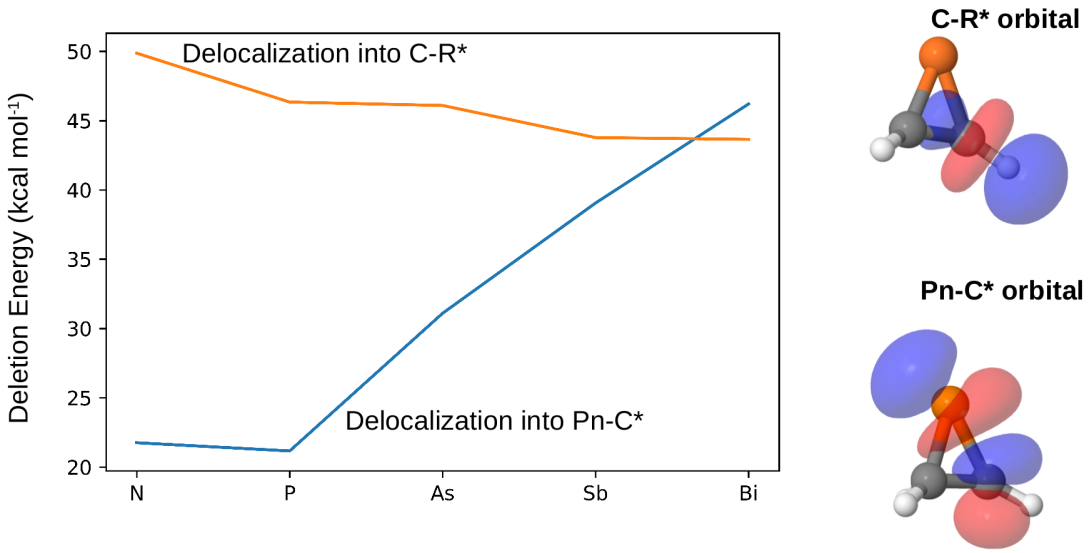

Figure 6.4: The NBO deletion energies into the C-R* and Pn-C* antibonding orbitals for the entire $Pn(CH)_3$ series. Values are in kcal mol^{-1}. The corresponding orbitals are depicted on the right side of the figure.

6.3 Chapter 3

6.3.1 Extended Methods

A benchmark is performed to justify the reliability of the results obtained with the composite methods as the size and lack of symmetry of these structures preclude a pure coupled cluster treatment with a large basis set. Using the reactants (H_2O + HNCO), hydrolysis transition states (**TS1**, **TS2**) and respective products (**M1**, **M2**), as well as the transition state connecting the two hydrolysis products (**TS3**), we benchmark this composite method against the CCSD(T)/cc-pVQZ geometries obtained from CFOUR 2.0. The MP2[TZ,QZ] + ΔCCSD(T)/DZ method exhibited a RMSE of 0.0020 Å for bond lengths and a RMSE of 0.172° for bond angles, slightly outperforming conventional CCSD(T)/cc-pVTZ (RMSE = 0.0022 Å , 0.528°) and significantly outperforming CCSD(T)/cc-pVDZ (RMSE = 0.0112 Å , 1.746°), therefore justifying the reliability of geometries obtained via composite methods for this system. CCSD(T)/CBS single points are obtained on top of these composite geometries and likely sufficiently capture the correlation energy of each stationary point given the small CCSDT and CCSDT(Q) corrections obtained in the focal point tables for the parent case. In the catalyzed sections, CCSD(T)/aug-cc-pV(T+d)Z single points are run on the optimized geometries and not extrapolated due to the cost of the larger systems and the additional augmented functions.

The Cartesian coordinates for all optimized structures can be found in the accompanying files labeled: "Coupled Cluster Geometries" for the uncalatlyzed high level geometries, "Uncatalyzed Substituent Geometries" for the substituted RNCO + H_2O geometries, and "Catalyzed Geometries" for the water and RNCO catalyzed structures.

6.3.2 Additional Figures and Tables

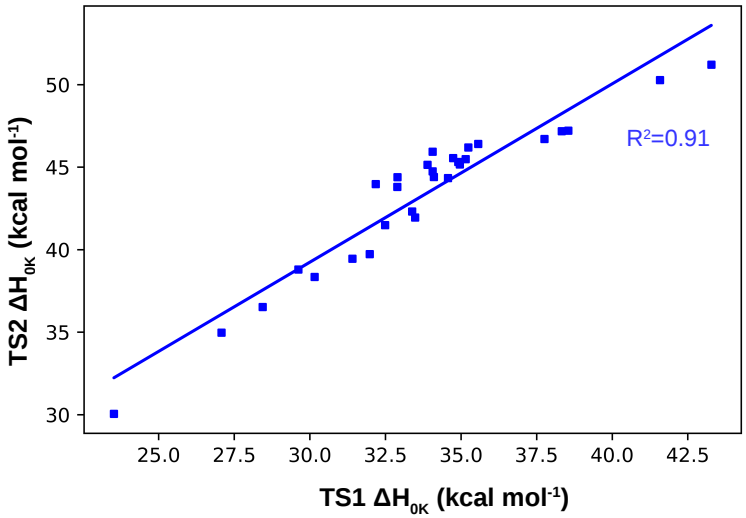

Figure 6.5: Scatter plot of the **TS1** vs **TS2** ΔH_{0K} values at the CCSD(T)/CBS//MP2[TZ,QZ] + ΔCCSD(T)/cc-pV(D+d)Z level of theory.

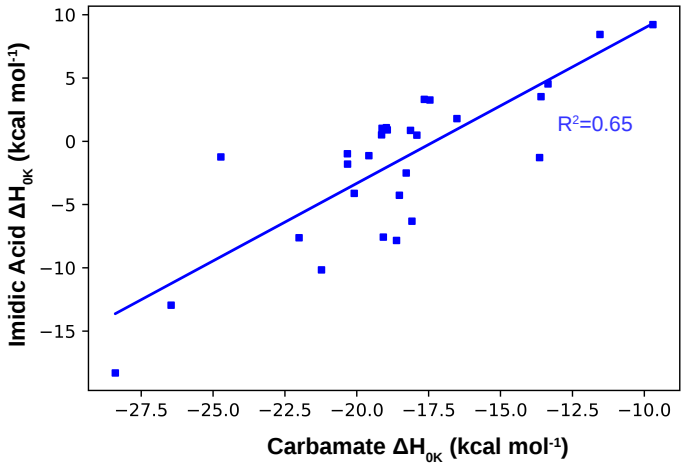

Figure 6.6: Scatter plot of the **M1** vs **M2** ΔH_{0K} values at the CCSD(T)/CBS//MP2[TZ,QZ] + ΔCCSD(T)/cc-pV(D+d)Z level of theory.

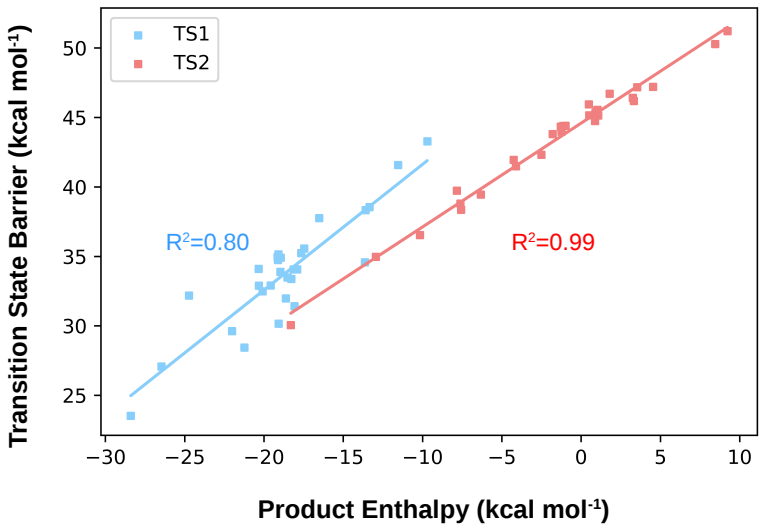

Figure 6.7: Bell-Evans-Polanyi plot of the product relative enthalpies vs the transition state barrier heights for **TS1** (Blue) and **TS2** (Red). ΔH_{0K} values are obtained at the CCSD(T)/CBS//MP2[TZ,QZ] + ΔCCSD(T)/cc-pV(D+d)Z level of theory.

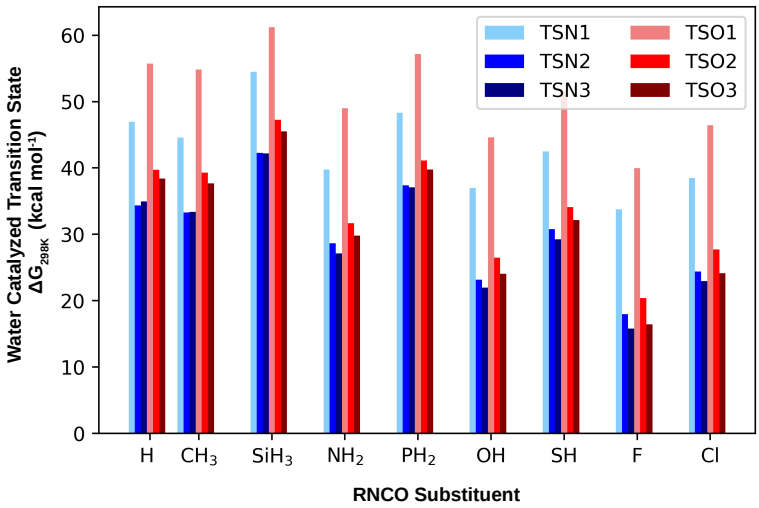

Figure 6.8: ΔG_{298K} barrier heights relative to separated products for the uncatalyzed and water catalyzed cases at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + ΔCCSD(T)/6-31+G** level of theory. The number following the "TSN" or "TSO" labels corresponds to the number of water molecules involved in the reaction.

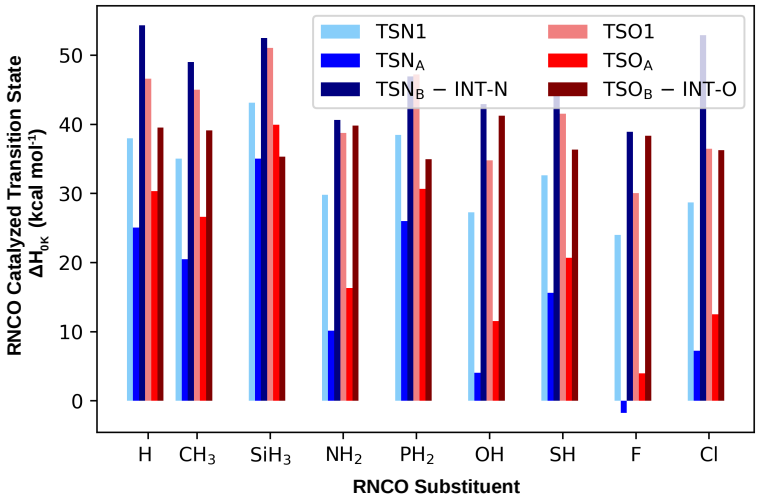

Figure 6.9: ΔH_{0K} barrier heights relative to separated products for the uncatalyzed (**TSN1** and **TSO1**) and RNCO catalysed cases at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + ΔCCSD(T)/6-31+G** level of theory. The barrier with the A subscript corresponds to the initial transition state forming the intermediate species and the second barrier corresponds to the hydrogen transfer process from the intermediate to the final produce. See SI Figure 8.

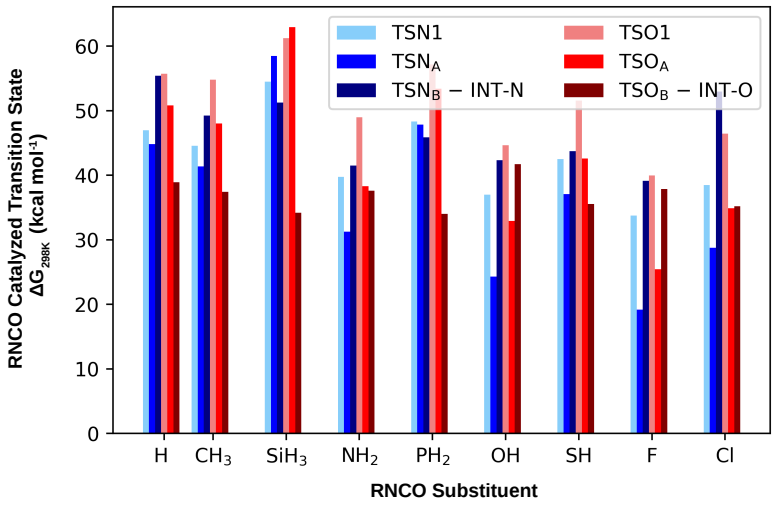

Figure 6.10: ΔG_{298K} barrier heights relative to separated products for the uncatalyzed (**TSN1** and **TSO1**) and RNCO catalysed cases at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + ΔCCSD(T)/6-31+G** level of theory. The barrier with the A subscript corresponds to the initial transition state forming the intermediate species and the second barrier corresponds to the hydrogen transfer process from the intermediate to the final produce. See SI Figure 8.

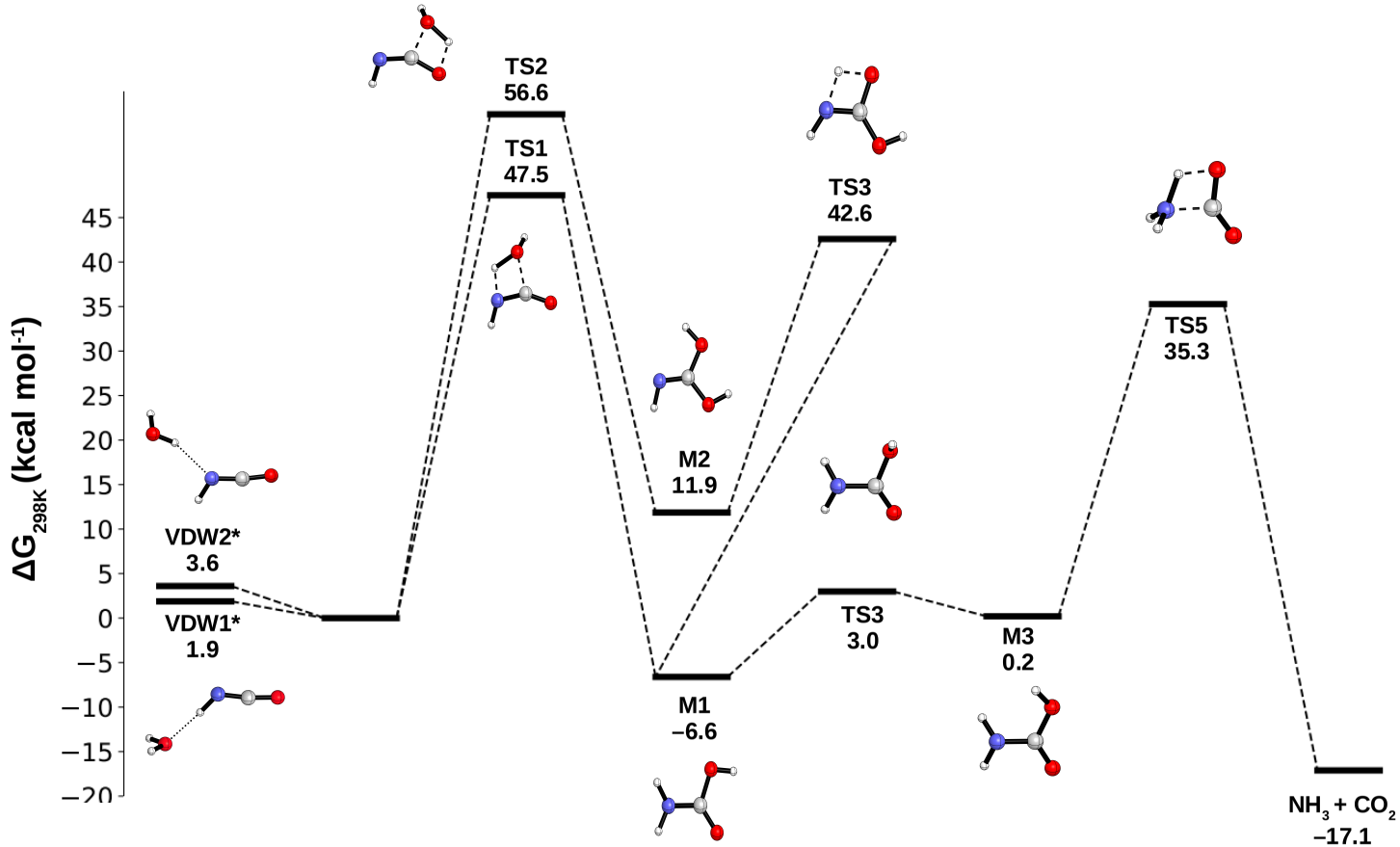

Figure 6.11: Predicted 298 K Gibbs energies for stationary points on the HNCO + H_2O potential energy surface. All values are in kcal mol^{-1} and target the CCSDT(Q)/CBS limit. Structures denoted with an asterisk are computed using augmented basis functions.

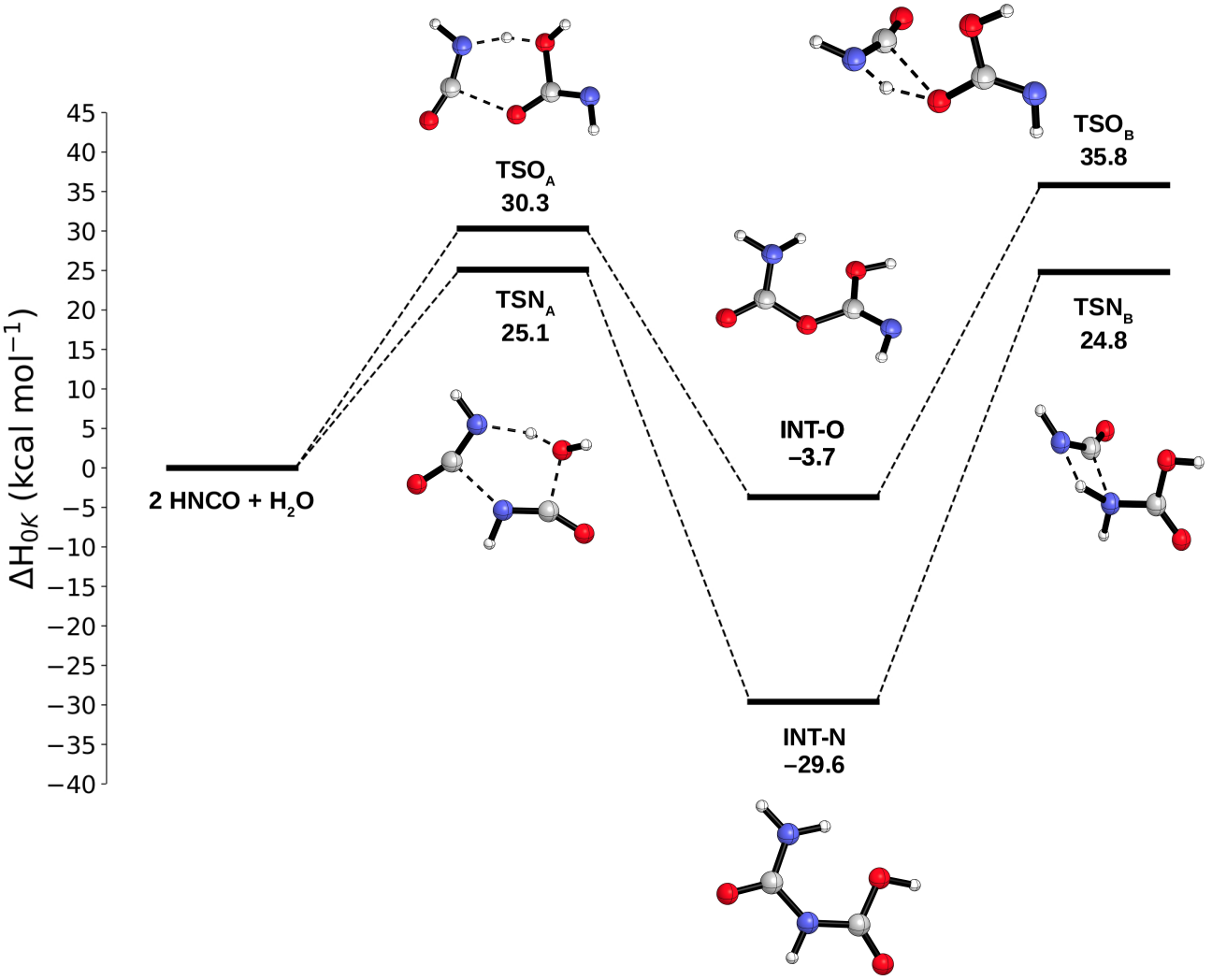

Figure 6.12: Predicted ΔH_{0K} energies for stationary points on the 2HNCO + H_2O potential energy surface. All values are in kcal mol^{-1} and are obtained at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + ΔCCSD(T)/6-31+G** level of theory.

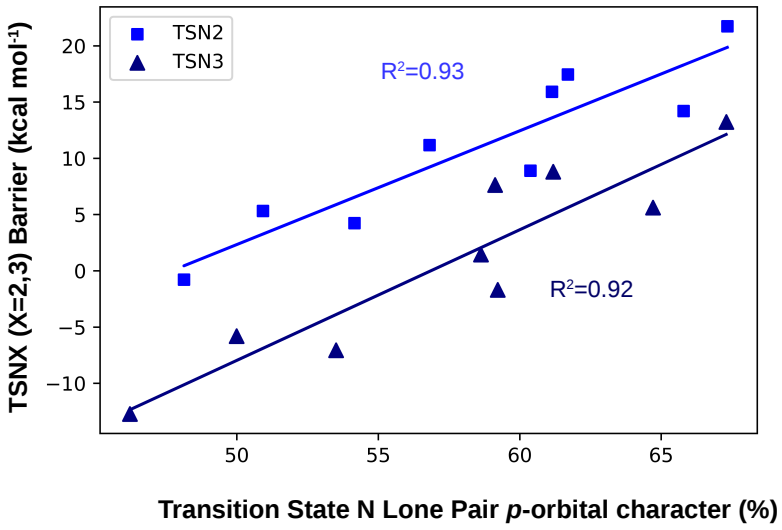

Figure 6.13: Scatter plot of ΔH_{0K} barrier heights at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + ΔCCSD(T)/6-31+G** level of theory versus the *p*-orbital character of the nitrogen lone pair in the **TSN2** and **TSN3** transition states.

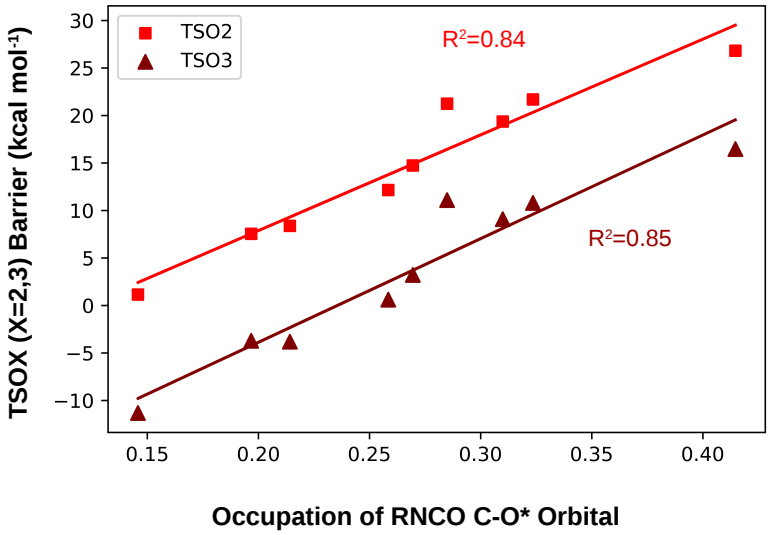

Figure 6.14: Scatter plot of ΔH_{0K} barrier heights at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + ΔCCSD(T)/6-31+G** level of theory of **TSO2** and **TSO3** versus the C-O* occupancy in the RNCO species.

Table 6.3: Reactant data for RNCO structures at the CCSD(T)/CBS//MP2[TZ,QZ] + Δ CCSD(T)/cc-pV(D+d)Z level of theory. The CCSD(T)/CBS energies are in hartree and all other values are in kcal mol^{-1}. The final two columns are the ZPVE and Gibbs energy corrections.

Substituent	CCSD(T)/CBS Energy (hartree)	ZPVE (kcal/mol)	Gibbs Correction (kcal/mol)
Water	-76.37451504	13.45	1.97
H	-168.5125432	13.31	-1.09
F	-267.5823254	8.28	-7.97
Cl	-627.6254199	8.04	-8.84
Br	-2740.697359	7.84	-9.81
OH	-243.5943375	16.23	-0.18
SH	-566.2526638	13.72	-3.61
SeH	-2568.726513	12.86	-5.45
NH_2	-223.7575312	24.52	7.35
PH_2	-510.0436352	19.50	1.74
AsH_2	-2403.679436	18.29	-0.32
CH_3	-207.7588253	31.84	14.67
SiH_3	-458.8220168	24.30	5.59
GeH_3	-2245.437783	23.51	3.81
CH_2CH_3	-247.0202977	49.83	31.06
$CH(CH_3)_2$	-286.2831192	67.42	47.58
$CH_2CH_3CH_{3A}$	-286.2789689	67.81	48.03
$CH_2CH_3CH_{3B}$	-286.2786727	67.70	47.40
$C(CH_3)_3$	-325.5461258	84.64	63.88
C_6H_5	-399.2229168	64.55	44.10
$CHCH_{2A}$	-245.7869241	34.72	16.89
$CHCH_{2B}$	-245.7880954	34.60	16.85
COH_A	-281.7191908	20.26	2.76
COH_B	-281.717051	20.03	2.55
CF_3	-505.329789	17.65	-2.17
$COOH_A$	-356.9146381	23.94	5.47
$COOH_B$	-356.9133746	23.82	5.31
$COOH_C$	-356.9113867	23.82	5.32
SO_2Cl	-1175.764373	15.68	-4.91
CN	-260.6169319	13.23	-3.73

Table 6.4: Transition state data for **TS1** structures at the CCSD(T)/CBS//MP2[TZ,QZ] + Δ CCSD(T)/cc-pV(D+d)Z level of theory. The CCSD(T)/CBS energies are in hartree and all other values are in kcal mol^{-1}. The final two columns are the 0 K enthalpies and 298 K Gibbs energies relative to water plus RNCO.

Substituent	CCSD(T)/CBS	ZPVE	Gibbs Correction	ΔH_{0K}	ΔG_{298K}
H	-244.8287516	27.92	11.43	37.76	47.13
F	-343.9210018	22.76	5.18	23.52	33.67
Cl	-703.9555678	22.10	3.89	28.44	38.60
Br	-2817.024493	21.72	2.76	30.16	40.32
OH	-319.9270006	30.50	12.76	27.08	37.23
SH	-642.5758392	27.45	9.16	32.49	43.01
SeH	-2645.047793	26.28	7.16	33.38	44.04
NH_2	-300.086129	38.78	20.89	29.62	40.38
PH_2	-586.3571232	32.98	14.57	38.33	49.15
AsH_2	-2479.992573	31.77	11.97	38.55	48.83
CH_3	-284.0781844	45.64	27.19	34.96	45.16
SiH_3	-535.127442	37.68	17.59	43.28	53.38
GeH_3	-2321.745813	36.83	16.04	41.58	51.97
CH_2CH_3	-323.3391955	63.53	43.28	35.15	45.15
$CH(CH_3)_2$	-362.6012946	81.09	60.45	35.57	46.25
$CH_2CH_3CH_A3$	-362.5981755	81.46	60.49	34.91	45.20
$CH_2CH_3CH_{3B}$	-362.5976352	81.03	60.97	34.74	46.45
$C(CH_3)_3$	-401.8648262	98.30	76.85	35.24	46.02
C_6H_5	-475.5437995	78.40	56.89	34.06	44.47
$CHCH_{2A}$	-322.1082346	48.67	29.79	33.89	44.31
$CHCH_{2B}$	-322.1091209	48.59	29.65	34.10	44.40
COH_A	-358.0399829	34.06	15.24	34.06	44.22
COH_B	-358.0412747	34.11	15.33	32.18	42.38
CF_3	-581.6515997	31.51	10.61	33.48	43.88
$COOH_A$	-433.2370571	37.61	18.33	32.90	43.58
$COOH_B$	-433.2359372	37.55	18.28	32.89	43.60
$COOH_C$	-433.231119	37.46	18.15	34.57	45.24
SO_2Cl	-1252.088853	29.15	8.15	31.41	42.49
CN	-336.9415025	27.32	8.89	31.98	41.99

Table 6.5: Transition state data for **TS2** structures at the CCSD(T)/CBS//MP2[TZ,QZ] + Δ CCSD(T)/cc-pV(D+d)Z level of theory. The CCSD(T)/CBS energies are in hartree and all other values are in kcal mol^{-1}. The final two columns are the 0 K enthalpies and 298 K Gibbs energies relative to water plus RNCO.

Substituent	CCSD(T)/CBS	ZPVE	Gibbs Correction	ΔH_{0K}	ΔG_{298K}
H	-244.8150062	28.24	11.90	46.71	56.23
F	-343.9109923	23.01	5.63	30.05	40.39
Cl	-703.943128	22.37	4.38	36.53	46.90
Br	-2817.011946	22.03	3.31	38.35	48.75
OH	-319.9149665	30.83	13.26	34.97	45.27
SH	-642.5618673	27.68	9.49	41.49	52.11
SeH	-2645.034135	26.64	7.66	42.31	53.11
NH_2	-300.0717896	38.96	21.48	38.80	49.96
PH_2	-586.3433179	33.17	14.81	47.18	58.06
AsH_2	-2479.978563	31.65	12.51	47.21	58.17
CH_3	-284.0621398	45.77	28.23	45.16	56.27
SiH_3	-535.1150897	37.85	17.98	51.21	61.52
GeH_3	-2321.732256	37.01	16.39	50.28	60.83
CH_2CH_3	-323.3230821	63.76	43.89	45.49	55.87
$CH(CH_3)_2$	-362.5844575	81.36	61.04	46.41	57.41
$CH_2CH_3CH_{3A}$	-362.5819633	81.70	60.89	45.32	55.77
$CH_2CH_3CH_{3B}$	-362.5813915	81.65	60.42	45.55	56.10
$C(CH_3)_3$	-401.8477692	98.55	77.40	46.19	57.27
C_6H_5	-475.5271148	78.63	57.12	44.75	55.17
$CHCH_{2A}$	-322.0906486	48.89	29.97	45.14	55.53
$CHCH_{2B}$	-322.0930109	48.78	29.94	44.40	54.80
COH_A	-358.0212365	34.17	14.77	45.94	55.51
COH_B	-358.0224206	34.06	15.86	43.97	54.73
CF_3	-581.638544	31.79	11.30	41.95	52.76
$COOH_A$	-433.2186297	37.53	17.87	44.39	54.69
$COOH_B$	-433.2185062	37.53	18.28	43.80	54.61
$COOH_C$	-433.2162209	37.88	18.15	44.34	54.26
SO_2Cl	-1252.076479	29.42	8.57	39.45	50.67
CN	-336.9295628	27.58	9.30	39.73	49.89

Table 6.6: Data for **M1** structures at the CCSD(T)/CBS//MP2[TZ,QZ] + Δ CCSD(T)/cc-pV(D+d)Z level of theory. The CCSD(T)/CBS energies are in hartree and all other values are in kcal mol^{-1}. The final two columns are the 0 K enthalpies and 298 K Gibbs energies relative to water plus RNCO.

Substituent	CCSD(T)/CBS	ZPVE	Gibbs Correction	ΔH_{0K}	ΔG_{298K}
H	-244.9223369	32.37	15.94	-16.52	-7.08
F	-344.0104823	27.00	9.51	-28.40	-18.15
Cl	-704.0413428	26.24	8.15	-21.23	-10.97
Br	-2817.109632	25.90	7.09	-19.08	-8.77
OH	-320.0194913	35.00	17.38	-26.46	-16.19
SH	-642.6667584	31.91	13.41	-20.09	-9.80
SeH	-2645.137434	30.87	11.41	-18.28	-7.96
NH_2	-300.175309	43.11	25.18	-22.01	-11.30
PH_2	-586.4470794	37.52	18.61	-13.59	-3.26
AsH_2	-2480.082129	36.07	15.67	-13.35	-3.66
CH_3	-284.1714021	50.03	31.46	-19.14	-9.07
SiH_3	-535.219224	42.29	23.55	-9.70	1.74
GeH_3	-2321.837858	41.47	21.94	-11.54	0.12
CH_2CH_3	-323.4327719	68.01	48.56	-19.09	-8.29
$CH(CH_3)_2$	-362.6928146	85.49	64.93	-17.46	-6.70
$CH_2CH_3CH_{3A}$	-362.6910781	85.91	65.11	-18.94	-8.48
$CH_2CH_3CH_{3B}$	-362.6911861	85.88	65.00	-19.11	-8.22
$C(CH_3)_3$	-401.9561566	102.73	82.40	-17.65	-5.75
C_6H_5	-475.6339218	82.77	61.21	-18.13	-7.77
$CHCH_{2A}$	-322.1999121	53.34	34.52	-18.97	-8.49
$CHCH_{2B}$	-322.2028075	52.95	34.13	-20.33	–9.91
COH_A	-358.1297526	38.42	19.68	-17.91	-7.67
COH_B	-358.1391052	38.58	20.03	-24.73	-14.32
CF_3	-581.7408786	35.53	14.93	-18.52	-7.82
$COOH_A$	-433.3276372	41.96	22.27	-19.58	-9.32
$COOH_B$	-433.3280236	42.13	22.48	-20.32	-9.98
$COOH_C$	-433.3144565	41.55	21.68	-13.64	-3.53
SO_2Cl	-1252.174745	33.55	12.30	-18.08	-7.26
CN	-337.0281872	31.11	12.58	-18.62	-8.72

Table 6.7: Data for **M2** structures at the CCSD(T)/CBS//MP2[TZ,QZ] + Δ CCSD(T)/cc-pV(D+d)Z level of theory. The CCSD(T)/CBS energies are in hartree and all other values are in kcal mol^{-1}. The final two columns are the 0 K enthalpies and 298 K Gibbs energies relative to water plus RNCO.

Substituent	CCSD(T)/CBS	ZPVE	Gibbs Correction	ΔH_{0K}	ΔG_{298K}
H	-244.8928581	32.19	15.83	1.80	11.30
F	-343.9936087	26.49	9.06	-18.31	-8.02
Cl	-704.0233283	26.01	8.04	-10.16	0.23
Br	-2817.090977	25.71	7.03	-7.57	2.88
OH	-319.996893	34.33	16.73	-12.95	-2.66
SH	-642.6407336	31.56	13.28	-4.11	6.41
SeH	-2645.111934	30.64	11.56	-2.51	8.19
NH_2	-300.1515626	42.60	24.97	-7.62	3.40
PH_2	-586.4193259	37.22	18.63	3.53	14.18
AsH_2	-2480.053357	35.90	16.52	4.54	15.24
CH_3	-284.1393282	49.55	30.58	0.51	10.18
SiH_3	-535.1885146	41.95	22.75	9.23	20.21
GeH_3	-2321.805411	41.09	21.39	8.45	19.93
CH_2CH_3	-323.4001512	67.62	48.21	0.98	11.83
$CH(CH_3)_2$	-362.6593198	85.19	65.09	3.26	14.48
$CH_2CH_3CH_{3A}$	-362.6588153	85.51	64.82	0.90	11.48
$CH_2CH_3CH_{3B}$	-362.6584223	85.46	64.66	1.03	12.00
$C(CH_3)_3$	-401.922214	102.40	81.48	3.32	14.64
C_6H_5	-475.6029005	82.30	60.77	0.87	11.26
$CHCH_{2A}$	-322.1666475	52.52	33.51	1.08	11.38
$CHCH_{2B}$	-322.1712841	52.51	33.70	-0.98	9.44
COH_A	-358.0998243	38.04	19.23	0.49	10.66
COH_B	-358.1005879	37.90	19.82	-1.24	9.65
CF_3	-581.7181383	35.52	15.34	-4.26	6.86
$COOH_A$	-433.2978517	41.71	21.78	-1.14	8.88
$COOH_B$	-433.2978517	41.71	21.78	-1.80	8.25
$COOH_C$	-433.2950816	41.75	21.94	-1.28	8.89
SO_2Cl	-1252.156086	33.60	12.54	-6.32	4.69
CN	-337.0112838	31.28	13.05	-7.84	2.36

Table 6.8: Electronic Energies of water catalyzed transition states at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + Δ CCSD(T)/6-31+G* level of theory in units of hartree.

Substituent	TSN1	TSN2	TSN3	TSO1	TSO2	TSO3
H	-244.7299225	-321.11195	-397.4720773	-244.7166555	-321.1042313	-397.4673102
F	-343.7834132	-420.1699453	-496.5353876	-343.7741942	-420.1673929	-496.533823
Cl	-703.8261242	-780.2103513	-856.5746171	-703.8141246	-780.2060733	-856.5721726
OH	-319.7986074	-396.1827096	-472.5474073	-319.7871743	-396.1773712	-472.5421348
SH	-642.4542413	-718.8347154	-795.197708	-642.4406524	-718.8303508	-795.1952002
NH_2	-299.9658781	-376.3456907	-452.7092074	-299.9521234	-376.3413361	-452.7062571
PH_2	-586.2412929	-662.6211356	-738.9820406	-586.2277139	-662.6153609	-738.9793609
CH_3	-283.964123	-360.3438706	-436.703493	-283.9485841	-360.3359537	-436.6989573
SiH_3	-535.014325	-611.3953131	-687.7556607	-535.0021096	-611.3879234	-687.7515707

Table 6.9: ZPVE corrections of the water catalyzed transition states at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + Δ CCSD(T)/6-31+G* level of theory in units of kcal mol^{-1}.

Substituent	TSN1	TSN2	TSN3	TSO1	TSO2	TSO3
H	28.01	44.28	60.58	28.31	44.75	61.06
F	22.83	39.19	55.17	23.07	39.53	55.60
Cl	22.16	38.49	54.56	22.41	38.88	55.03
OH	30.59	47.20	63.35	30.92	47.16	63.38
SH	27.60	43.50	60.17	27.97	44.32	60.37
NH_2	38.86	54.92	71.05	39.21	55.44	71.51
PH_2	33.16	49.13	65.54	33.38	49.74	65.86
CH_3	45.70	61.77	77.46	45.92	61.96	78.09
SiH_3	37.65	53.95	70.16	37.91	54.40	70.83

Table 6.10: 298 K Gibbs energy corrections of the water catalyzed transition states at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + Δ CCSD(T)/6-31+G* level of theory in units of kcal mol^{-1}.

Substituent	TSN1	TSN2	TSN3	TSO1	TSO2	TSO3
H	11.54	26.19	40.31	11.98	26.73	40.78
F	5.26	19.58	34.26	5.69	20.39	33.92
Cl	3.98	18.55	33.21	4.42	19.20	32.87
OH	12.89	27.62	42.83	13.35	27.61	41.59
SH	9.08	23.64	37.43	9.63	24.22	38.76
NH_2	20.99	35.78	49.91	21.60	36.05	50.71
PH_2	14.27	29.20	42.94	14.60	29.33	43.93
CH_3	27.28	41.84	55.12	27.78	42.90	56.60
SiH_3	19.06	33.45	47.03	18.11	33.79	47.78

Table 6.11: Electronic energies of the RNCO catalyzed transition states at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + Δ CCSD(T)/6-31+G* level of theory in units of hartree.

Substituent	TSN_A	INT-N	TSN_B	TSO_A	INT-O	TSO_B
H	-413.202777	-413.2966795	-413.204246	-413.1926975	-413.2549686	-413.1848742
F	-611.3072504	-611.38562	-611.3178039	-611.2981323	-611.3637812	-611.2956689
Cl	-1331.39357	-1331.467287	-1331.376764	-1331.382973	-1331.44117	-1331.376764
OH	-563.3381657	-563.4283122	-563.353249	-563.3266384	-563.3980269	-563.3249864
SH	-1208.64858	-1208.731445	-1208.653818	-1208.639777	-1208.700815	-1208.636545
NH_2	-523.6714178	-523.7567986	-523.686666	-523.660973	-523.7276022	-523.6571952
PH_2	-1096.225175	-1096.305785	-1096.224111	-1096.215638	-1096.271357	-1096.209557
CH_3	-491.6688729	-491.7641649	-491.6796428	-491.6572195	-491.7223581	-491.6536195
SiH_3	-993.773049	-993.8587526	-993.767101	-993.7630792	-993.8176374	-993.7555198

Table 6.12: ZPVE corrections of the RNCO catalyzed transition states at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + Δ CCSD(T)/6-31+G* level of theory in units of kcal mol^{-1}.

Substituent	TSN_A	INT-N	TSN_B	TSO_A	INT-O	TSO_B
H	45.20	49.47	45.81	44.11	49.2	44.73
F	34.37	38.11	34.46	34.37	37.9	33.48
Cl	33.24	36.75	32.82	31.88	36.5	32.34
OH	49.52	54.19	50.03	49.77	53.9	49.28
SH	43.86	48.23	43.72	43.40	47.8	43.86
NH_2	66.31	70.33	66.96	65.95	70.2	65.88
PH_2	55.24	59.23	54.92	53.94	59.0	55.16
CH_3	80.42	84.93	80.89	79.24	84.2	80.15
SiH_3	65.09	69.22	64.20	63.74	68.4	64.78

Table 6.13: 298 K Gibbs corrections of RNCO catalyzed transition states at the CCSD(T)/aug-cc-pV(T+d)Z//MP2/jul-cc-pV([TZ,QZ]+d)Z + Δ CCSD(T)/6-31+G* level of theory in units of kcal mol^{-1}.

Substituent	TSN_A	INT-N	TSN_B	TSO_A	INT-O	TSO_B
H	25.12	29.16	26.58	24.76	29.4	24.30
F	11.75	16.38	12.97	12.29	16.0	11.14
Cl	9.90	13.93	10.11	9.40	13.4	8.12
OH	25.97	32.55	27.75	27.37	31.7	27.54
SH	19.61	25.17	20.20	19.59	24.2	19.36
NH_2	42.84	47.47	44.94	43.35	49.1	42.54
PH_2	30.63	36.04	30.67	30.22	34.5	29.76
CH_3	56.49	61.96	58.15	55.84	61.9	56.15
SiH_3	39.66	45.18	38.93	37.87	43.3	38.54

Table 6.14: Focal point analyses for the CCSD(T)/cc-pVQZ transition states. Below each table the CCSDT(Q)/CBS energy is printed followed by the ZPVE correction, frozen core correction, scalar relativistic correction, and the DBOC. Units are kcal mol $^{-1}$.

TS1

	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
cc-pVDZ	49.263	-16.692	3.394	-2.052	0.037	-0.160	33.790
cc-pVTZ	52.518	-18.058	4.017	-2.757	[0.037]	[-0.160]	[35.596]
cc-pVQZ	53.668	-18.072	3.978	-2.877	[0.037]	[-0.160]	[36.575]
cc-pV5Z	54.078	-17.988	3.879	-2.889	[0.037]	[-0.160]	[36.958]
cc-pV6Z	54.176	[-17.950]	[3.835]	[-2.894]	[0.037]	[-0.160]	[37.044]
CBS	[54.197]	[-17.899]	[3.775]	[-2.901]	[0.037]	[-0.160]	[37.049]

$E_{CCSDT(Q)/CBS+\Delta}$ =37.05+1.26+0.24+0.00$_4$–0.02 = **38.54**

TS2

	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
cc-pVDZ	57.665	-13.752	1.418	-1.414	0.086	-0.084	43.919
cc-pVTZ	59.574	-15.228	1.970	-2.068	[0.086]	[-0.084]	[44.251]
cc-pVQZ	60.690	-15.179	1.919	-2.156	[0.086]	[-0.084]	[45.276]
cc-pV5Z	61.021	-15.087	1.809	-2.156	[0.086]	[-0.084]	[45.589]
cc-pV6Z	61.102	[-15.046]	[1.760]	[-2.156]	[0.086]	[-0.084]	[45.662]
CBS	[61.120]	[-14.990]	[1.693]	[-2.155]	[0.086]	[-0.084]	[45.669]

$E_{CCSDT(Q)/CBS+\Delta}$ =45.67+1.55+0.32+0.00$_1$–0.03 = **47.51**

TS3

	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
cc-pVDZ	37.920	-8.822	0.412	-0.494	0.091	-0.018	29.089
cc-pVTZ	41.231	-10.701	0.840	-1.092	[0.091]	[-0.018]	[30.350]
cc-pVQZ	42.343	-10.911	0.723	-1.179	[0.091]	[-0.018]	[31.049]
cc-pV5Z	42.6717	-10.967	0.605	-1.192	[0.091]	[-0.018]	[31.190]
cc-pV6Z	42.753	[-10.992]	[0.552]	[-1.197]	[0.091]	[-0.018]	[31.189]
CBS	[42.771]	[-11.026]	[0.480]	[-1.205]	[0.091]	[-0.018]	[31.095]

$E_{CCSDT(Q)/CBS+\Delta}$ = 31.10+2.21+0.11+0.08–0.03= **33.47**

TS4

	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
cc-pVDZ	-15.566	4.551	-2.787	1.472	0.058	0.286	-11.986
cc-pVTZ	-13.274	3.219	-2.924	1.221	[0.058]	[0.286]	[-11.414]
cc-pVQZ	-12.449	3.028	-3.195	1.171	[0.058]	[0.286]	[-11.102]
cc-pV5Z	-12.151	3.042	-3.3897	1.171	[0.058]	[0.286]	[-10.983]
cc-pV6Z	-12.070	[3.048]	[-3.474]	[1.171]	[0.058]	[0.286]	[-10.981]
CBS	[-12.051]	[3.056]	[-3.592]	[1.171]	[0.058]	[0.286]	[-11.071]

$E_{CCSDT(Q)/CBS+\Delta}$ = –11.07+4.79+0.04+0.11–0.01= **–6.13**

TS5

	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
cc-pVDZ	33.063	-13.569	1.741	-1.389	0.059	-0.055	19.850
cc-pVTZ	35.835	-14.068	2.059	-1.864	[0.059]	[-0.055]	[21.966]
cc-pVQZ	36.960	-13.892	1.923	-1.921	[0.059]	[-0.055]	[23.073]
cc-pV5Z	37.354	-13.790	1.781	-1.915	[0.059]	[-0.055]	[23.435]
cc-pV6Z	37.440	[-13.744]	[1.719]	[-1.912]	[0.059]	[-0.055]	[23.506]
CBS	[37.459]	[-13.682]	[1.633]	[-1.909]	[0.059]	[-0.055]	[23.505]

$E_{CCSDT(Q)/CBS+\Delta}$ =23.51+2.43+0.25+0.01–0.01 = **26.19**

Table 6.15: Focal point analyses for the CCSD(T)/cc-pVQZ minima. Below each table the CCSDT(Q)/CBS energy is printed followed by the ZPVE correction, frozen core correction, scalar relativistic correction, and the DBOC. Units are kcal mol $^{-1}$.

M1

	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
cc-pVDZ	-26.966	4.817	-2.451	1.545	0.088	0.289	-22.678
cc-pVTZ	-24.267	3.542	-2.665	1.325	[0.088]	[0.289]	[-21.688]
cc-pVQZ	-23.352	3.322	-2.974	1.282	[0.088]	[0.289]	[-21.345]
cc-pV5Z	-23.035	3.301	-3.170	1.280	[0.088]	[0.289]	[-21.247]
cc-pV6Z	-22.950	[3.292]	[-3.256]	[1.279]	[0.088]	[0.289]	[-21.258]
CBS	[-22.929]	[3.279]	[-3.375]	[1.277]	[0.088]	[0.289]	[-21.371]

$E_{CCSDT(Q)/CBS+\Delta}$ =–21.37+5.62+0.01+0.12–0.02 = **−15.64**

M2

	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
cc-pVDZ	-6.097	3.783	-3.473	1.408	0.098	0.257	-4.026
cc-pVTZ	-3.575	2.025	-3.438	1.073	[0.098]	[0.257]	[-3.561]
cc-pVQZ	-2.721	1.947	-3.676	1.055	[0.098]	[0.257]	[-3.040]
cc-pV5Z	-2.449	2.002	-3.858	1.070	[0.098]	[0.257]	[-2.881]
cc-pV6Z	-2.377	[2.026]	[-3.939]	[1.077]	[0.098]	[0.257]	[-2.859]
CBS	[-2.360]	[2.059]	[-4.050]	[1.086]	[0.098]	[0.257]	[-2.911]

$E_{CCSDT(Q)/CBS+\Delta}$ = –2.91+5.47+0.13+0.10–0.02= **2.77**

M3

	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
cc-pVDZ	-18.908	4.087	-2.700	1.404	0.060	0.286	-15.771
cc-pVTZ	-16.498	2.796	-2.825	1.154	[0.060]	[0.286]	[-15.027]
cc-pVQZ	-15.696	2.693	-3.101	1.125	[0.060]	[0.286]	[-14.633]
cc-pV5Z	-15.386	2.738	-3.298	1.133	[0.060]	[0.286]	[-14.467]
cc-pV6Z	-15.309	[2.758]	[-3.384]	[1.136]	[0.060]	[0.286]	[-14.453]
CBS	[-15.291]	[2.786]	[-3.504]	[1.141]	[0.060]	[0.286]	[-14.522]

$E_{CCSDT(Q)/CBS+\Delta}$ = –14.52+5.53+0.08+0.10–0.00$_7$= **−8.83**

CO_2 + NH_3

	HF	+δ MP2	+δ CCSD	+δ (T)	+δ T	+δ (Q)	NET
cc-pVDZ	-25.854	-0.261	0.277	-0.004	0.037	0.022	-25.784
cc-pVTZ	-24.249	1.891	-0.072	0.214	[0.037]	[0.022]	[-22.156]
cc-pVQZ	-23.713	2.533	-0.281	0.306	[0.037]	[0.022]	[-21.097]
cc-pV5Z	-23.495	2.813	-0.416	0.352	[0.037]	[0.022]	[-20.688]
cc-pV6Z	-23.435	[2.937]	[-0.476]	[0.372]	[0.037]	[0.022]	[-20.543]
CBS	[-23.420]	[3.107]	[-0.558]	[0.400]	[0.037]	[0.022]	[-20.412]

$E_{CCSDT(Q)/CBS+\Delta}$ =–20.41+2.07+0.11–0.07+0.05 = **−18.27**

Table 6.16: Focal point analyses for the CCSD(T)/aug-cc-pVTZ van der Waals complexes. Below each table the CCSDT(Q)/CBS energy is printed followed by the ZPVE correction, frozen core correction (Obtained at the CCSD(T)/aug-cc-pCVTZ level of theory), scalar relativistic correction (obtained using the decontracted aug-cc-pCVTZ basis set), and the DBOC (at the SCF/aug-cc-pVTZ level of theory). Units are kcal mol $^{-1}$.

VDW1

	HF	$+\delta$ MP2	$+\delta$ CCSD	$+\delta$ (T)	$+\delta$ T	$+\delta$ (Q)	NET
aug-cc-pVDZ	-4.596	-1.965	0.518	-0.354	0.037	-0.008	-6.367
aug-cc-pVTZ	-4.380	-2.088	0.485	-0.334	[0.037]	[-0.008]	[-6.289]
aug-cc-pVQZ	-4.335	-2.005	0.504	-0.330	[0.037]	[-0.008]	[-6.137]
aug-cc-pV5Z	-4.306	[-1.976]	[0.510]	[-0.328]	[0.037]	[-0.008]	[-6.070]
aug-cc-pV6Z	[-4.295]	[-1.963]	[0.513]	[-0.327]	[0.037]	[-0.008]	[-6.042]
CBS	[-4.290]	[-1.945]	[0.517]	[-0.326]	[0.037]	[-0.008]	[-6.015]

$E_{CCSDT(Q)/CBS+\Delta}$ =–6.02+1.72–0.07++0.02–0.02 = **−4.36**

VDW2

	HF	$+\delta$ MP2	$+\delta$ CCSD	$+\delta$ (T)	$+\delta$ T	$+\delta$ (Q)	NET
aug-cc-pVDZ	-2.014	-1.683	0.282	-0.256	0.010	-0.001	-3.662
aug-cc-pVTZ	-1.844	-1.673	0.281	-0.244	[0.010]	[-0.001]	[-3.471]
aug-cc-pVQZ	-1.802	-1.552	0.310	-0.240	[0.010]	[-0.001]	[-3.276]
aug-cc-pV5Z	-1.779	[-1.509]	[0.320]	[-0.239]	[0.010]	[-0.001]	[-3.198]
aug-cc-pV6Z	[-1.771]	[-1.490]	[0.325]	[-0.238]	[0.010]	[-0.001]	[-3.165]
CBS	[-1.767]	[-1.464]	[0.331]	[-0.238]	[0.010]	[-0.001]	[-3.128]

$E_{CCSDT(Q)/CBS+\Delta}$ =–3.17+1.43–0.03+0.00_1–0.01 = **−1.73**

6.4 Chapter 4

6.4.1 Lab Exercise Handout

Introduction

Every single application of scientific computing, whether in fields such as biology, chemistry, physics, or meteorology walks a delicate balance between the computing time (known as cost) and accuracy of model results. In an ideal world, one would simply increase the computational resources used until a desired accuracy can be obtained in an acceptable time frame. Unfortunately, limitations due to economic cost, scaling of algorithms, and the physical design of computational technology often preclude this approach. Instead, some compromise between cost and accuracy must be chosen. Understanding the logic and justification behind these kinds of difficult decisions is an invaluable skill for any scientist interacting with computational methods. This lab utilizes *ab initio* quantum chemistry to study the combustion energy of the methane as a representative example of how computational quantum chemists approach this dilemma. The primary learning goals are less about the technicalities of computational quantum chemistry and more about developing the necessary skills to be an effective computational scientist in any field.

Instructions

This lab exercise will require the accompanying Excel sheet. Complete all of the tasks below and answer the Questions in a separate lab write-up that will be turned in with your completed Excel notebook.

Part 1: Obtaining the Data

Computational chemists make chemical predictions based on solutions of the Schrödinger equation, which correspond to the electronic energy of some spatial arrangement of atoms (usually forming a molecule). The Schrödinger equation cannot be solved analytically for all but the simplest cases, thus approximate methods are utilized. An approximate solution is said to be obtained at a particular level of theory, i.e a particular method and basis set combination. The method is the algorithm describing the physics of the system and can have significantly varying complexity. The basis set refers to the mathematical functions used to describe the orbitals and provides better predictions as its size increases. The first part of the lab involves computing the methane combustion reaction energies at various levels of theory. The methods (Hartree-Fock (HF), MP2, CCSD, and CCSD(T)) and basis sets (cc-pVDZ, cc-pVTZ, cc-pVQZ, and cc-pV5Z) for this lab are listed in order of increasing sophistication and size, respectively.

Below are the Cartesian coordinates for the equilibrium geometries of each reactant (CH_4 and O_2) and product (CO_2 and H_2O). Using the empty tables in the Excel sheet and the Psi4 tutorial, compute and record the energy of each geometry at every level of theory (combination of basis set and method). Consider the following tips as you perform the computations.

1. Hartree-Fock, MP2, and CCSD values are computed as a byproduct of each CCSD(T) job and are printed at the bottom of the output file. You should only have to run one computation for each basis set and molecule pair.

2. The ground state of O_2 is a triplet two electrons are unpaired), so the spin multiplicity needs to be set to 3. See the Psi4 tutorial for guidance on where this change is made.

3. The cc-pV5Z row is already filled out for you in the Excel sheet and can be used to check that your results are in the ballpark. The cc-pV5Z values should not differ with any other energies in the same column by more than 1 hartree.

Cartesian Coordinates:

CH_4

C	0.0000000000	–0.0000000000	–0.0000000029
H	0.0000000000	–0.8891689480	0.6287338751
H	–0.0000000000	0.8891689480	0.6287338751
H	0.8891689588	0.0000000000	–0.6287338578
H	–0.8891689588	–0.0000000000	–0.6287338578

O_2 (Triplet)

O	0.0000000000	0.0000000000	-0.6061050518
O	0.0000000000	0.0000000000	0.6061050518

CO_2

O	0.0000000000	0.0000000000	-1.1662885018
C	0.0000000000	0.0000000000	0.0000000000
O	0.0000000000	0.0000000000	1.1662885018

H_2O

H	0.0000000000	0.7538827233	-0.5270184796
O	0.0000000000	0.0000000000	0.0664139108
H	0.0000000000	-0.7538827233	-0.5270184796

Once you have all of the data, compute the reaction energy at each level of theory (Products – Reactants, assuming one unit of CH_4 is reacting). Note that the energy is in hartree units which is likely disconnected from your chemical intuition. Convert the table of reaction energies to the more practical kcal mol^{-1} units commonly used in thermodyanmics. Quickly "Google" the energy of this reaction to make sure you are qualitatively in the ballpark. We will closely compare to a reliable experimental value later.

Part 2: Analyzing the Error

There are many sources of error and approximations made in our computations, but most of these are beyond the scope of this lab. The two primary and largest sources of error that we will consider are basis set size and sophistication of method (i.e. correlation energy). In our example, the methods and basis sets are built such that they systematically converge to an answer that completely eliminates both of these sources of error in the absolute energies if we were to continue with higher order methods or larger basis sets in the series. As the predicted absolute energies for each species approach the error free limit, we can observe what happens to the energy of reaction. In practice, computing a result that is converged to the "exact answer" is incredibly expensive and usually impossible to obtain. We provide a reference "exact answer" for comparison in the Excel sheet which has been converged as tight as possible with respect to both sources of error and can be used to benchmark the total error in your predictions. Generally, most high level theoretical studies seek error that is within chemical accuracy, or less than 1 kcal mol^{-1}.

Question 1: Consider the first row (i.e. the cc-pVDZ basis set) of your energy of methane combustion table. Describe the effect of increasing the quality of the method both qualitatively and quantitatively on the predicted reaction energies. Be sure to focus on the jumps between consecutive predictions. Does the same overall trend hold for all basis sets?

Question 2: Now consider the Hartree–Fock column of the energy of combustion table. Describe the trends in combustion energy due to basis set size, again commenting on the qualitative and quantitative features. Do the same trends hold for all methods?

Question 3: Rank the top five level of theory that are in most agreement with the "exact answer". What is the error of the best case result? Is it within chemical accuracy?

Part 3: Analyzing the Cost

One of the reasons we do an error analysis like this, also called a benchmark, is so that we are able to predict the error that might exist in larger systems by studying smaller similar systems on which high level computations are affordable. What if we wanted to study the combustion energy of larger hydrocarbons like propane or butane instead of methane? Computational cost becomes a major roadblock for studying these larger molecules. Below we will estimate how expensive it would be to obtain energy computations for propane and then combine these estimates with our error analysis in Part 4 to determine a level of theory that might be a reasonable choice to study propane combustion with.

You might have noticed that some of the computations you ran in Part 1 took different lengths of time. The computational cost to produce a single energy value is dependent on the number of electrons,

the symmetry of the molecules (i.e. the point group), the method employed, the size of the basis set, the memory allotment, and the speed of the computer. In our case, holding all else equal, we are going to focus on basis set size and level of theory, which also happen to be the primary cost prohibiting factors for obtaining higher quality results. The reason for this is that improving the method requires solving of increasingly complicated mathematical operations and larger basis sets increase the number of times these operations must be performed. Therefore, the propane molecule (eleven atoms and 26 electrons) would take significantly longer to compute than methane (five atoms and 10 electrons).

Fortunately, there are ways to estimate how long certain computations will take using exponential functions and the fact that each method scales approximately according to a function of the form: $time=(number\ of\ basis\ functions)^x$. The exponent ($x$) is known as the scaling factor and is an integer determined from the complexity of the algorithm underlying each theoretical method. (x= 4, 5, 6, and 7 for HF, MP2, CCSD, and CCSD(T), respectively). Thus, for any given method, If we have a reference computation time with a small number of basis functions, we can determine a rough estimate for the time it will take to perform the same computation with a larger basis set, allowing us to make an educated guess about what levels of theory can be afforded. Though we only consider time here, be aware that computer memory would be severely limiting as well, but it is not considered for the sake of simplicity.

Question 4: The accompanying spreadsheet has possible HF, MP2, CCSD, and CCSD(T) computation times (in seconds) of propane with a cc-pVDZ basis set on a hypothetical computer. Using this information, the scaling factors, and the number of basis functions necessary to describe propane for each cardinality (also in the spreadsheet), calculate estimates for the time it would take to compute propane at each level of theory.

a) Which propane jobs would not finish in a day?

b) If you started it today, what calendar day would the propane CCSD(T)/cc-pV5Z job finish?

c) The next method we could have included in this study is CCSDT which has a scaling factor of 8. Make an approximate guess how many basis functions it would take for a propane CCSDT computation to finish in one year, assuming that a CCSDT/cc-pVDZ job takes 5 seconds to compute? Now approximate a value using the scaling formula. Does the result surprise you?

d) Propane is described by 202 basis functions and a CH_2 group is described by 58 basis functions using the cc-pVTZ basis set. Assume you have a month (30 days) to determine the CCSD energy of various alkanes and that the computation time only depends on the number of basis functions. What is the largest single branched alkane you could compute in the given time frame based on your estimate?

Part 4: Balancing Cost and Accuracy

You have thus far performed a small benchmark analysis for the combustion of propane using error estimates of methane combustion and estimated computation times for propane. This is the point where the analyses are combined and a level of theory must be selected given the constraints and the research goal. Though the final decision might be obvious, the fact that benchmark data reinforces a selection of level of theory is vital to a computational study. The following questions allow you to make that sort of decision for the study of propane and also ask you to consider some more interesting concepts related to this benchmark that you might not have considered.

Question 5: Pretend that your task between now and next week's lab (7 days) will be to compute the combustion energy of propane as accurately as possible. What level of theory would you select to produce the best possible results and ensure you finish the assignment on time based on your previous error and cost analysis? Provide justification. For this analysis to be valid, we assume that the patterns in the combustion energy of methane are similar to that of propane, with respect to level of theory. How could you test the validity of this assumption between now and next week's lab? (Hint: you are not restricted to one computation.)

Question 6 : Consider the predicted MP2/cc-pVTZ methane combustion energy, which exhibits a remarkably low error relative to the "exact answer". We have emphasized that increasing the basis set and the method will eventually converge to an answer free of error due to finite basis size or correlation energy. Knowing this, suggest an explanation for why the error increases going from MP2/cc-pVTZ to CCSD/cc-pVTZ and MP2/cc-pVQZ and why MP2/cc-pVTZ might not be a wise selection for larger systems. (Hint: consider that the MP2 column and cc-pVTZ row are approaching.)

Part 5: Some Cost Saving Measures:

By now you should have an appreciation for the tension between the cost and accuracy trade-off as it applies to computational chemistry. In the long term, the hope is that better methods or faster computers become widely available. In the meantime, there are always clever tools being developed that can be used to improve accuracy without significantly increasing the cost, or lower the cost without significantly lowering the accuracy. Below details two small exercises that introduce two such tools: additive corrections and basis set extrapolation.

Additive corrections: Our energy computations contain many approximations that exclude the treatment of certain physical effects. We can sometimes compute the effect of treating one or more of these features at a lower level of theory, add its influence to the reaction energy (Product-Reactants), and improve

our predicted energy with minimal cost.

Question 7: So far we have only considered predictions of electronic energies, which cannot be directly compared to experiment (experimental results include rotational and vibrational energy). Use the following NIST **website** to find the experimental 298 K enthalpy (ΔH) of methane combustion. (Note, the experimental combustion of methane value involves the liquid phase of water. You will need to use the same site to find the 298 K enthalpy values for water in the gas phase and liquid phase and adjust the experimental value to correspond with your theoretical predictions. Remember your balanced reaction has two units of water.) How far is your CCSD(T)/cc-pV5Z methane combustion energy away from the experimental value? Though the electronic energy we have obtained might be excellent, exclusion of an enthalpy correction results in significant disagreement from experiment. Using guidance from the PSI4 tutorial, compute the harmonic vibrational frequencies at the MP2/cc-pVDZ level of theory. Then calculate the additive 298 K enthalpy correction using the results from the harmonic frequency computation. How does the 298 K enthalpy of combustion compare to the experimental value? Is it within chemical accuracy? You could obtain this correction at a higher level of theory for even better results, but we can often assume that the error due to level of theory in a correction is far less severe than excluding a correction altogether.

Basis set extrapolation: Certain basis sets, such as the ones used thus far, are built to systematically converge to the complete basis set (CBS) limit. The CBS limit is the energy result at a given method obtained with an infinite basis set size, if such a computation could be performed. The CBS limit can be estimated by increasing the size of the basis set until the energy does not change anymore (this is very expensive to do) or by the use of specially designed basis set extrapolation functions. We will consider the use of basis set extrapolation techniques to estimate a CCSD(T) methane combustion energy free from error due to basis set size and evaluate the cost savings of such an approach.

Question 8:

Take as a given that the CCSD(T)/CBS methane combustion energy is –194.52 kcal mol^{-1} (obtained with extremely large basis sets). The astute observer might notice that this value, though considered free from basis set error, is significantly different from the "exact answer" given previously. This should be noted, but is of no concern for our consideration of basis set extrapolation because the "exact answer" consists of higher order methods beyond CCSD(T). Therefore, we will only focus on the CCSD(T) column and the CCSD(T)/CBS value given at the start of this paragraph.

Make a small table and determine the error between the CCSD(T)/cc-pVXZ (X= D, T, Q, 5) methane combustion energies and the given CCSD(T)/CBS prediction. Now, consult the PSI4 tutorial. Correctly modify the input files (instructions given in the tutorial) to compute an extrapolated estimate of the CCSD(T)/CBS energy for each species using a two-point extrapolation of CCSD(T)/cc-pVTZ and

CCSD(T)/cc-pVQZ energies. We will refer to this value as CCSD(T)/cc-pV{T,Q}Z. Use these predictions to compute a combustion energy (products – reactants). Ammend the table you just made to compare the CCSD(T)/cc-pV{T,Q}Z prediction to the given CCSD(T)/CBS energy. Comment on the errors in your table you just made, giving specific attention to the CCSD(T)/cc-pV{T,Q}Z and CCSD(T)/cc-pV5Z cases.

Consider your estimated timings for propane. How long would it take to compute the CCSD(T)/cc-pVTZ and CCSD(T)/cc-pVQZ values one after another? How much time would be saved by performing a T,Q extrapolation rather than a CCSD(T)/cc-pV5Z computation?

Though this cost savings is impressive, there is not an absolute guarantee that the larger propane molecule will converge to the CBS at the same rate as methane. Nevertheless, extrapolation is an excellent tool to estimate values that are far too expensive to compute explicitly. Even though basis set extrapolation can be a wonderful tool, it also must be benchmarked and utilized appropriately to produce meaningful results. For example, a cc-pV{D,T}Z extrapolation is known to produce poor results even though it would be extremely fast. In practice one might try a variety of extrapolation methods and average the results.

6.4.2 Psi4 Tutorial

The following will be a guide for utilizing the Psi4 quantum chemistry package for the purposes of completing the "Computational Methane Combustion" laboratory exercise.

Structure of the input file

A simple example of an input file is shown in Figure 6.15. There are three principle parts of the input file that we will focus on: the molecule, the computation parameters, and the extra options. The molecule section receives a list of atoms followed by their respective x, y, and z coordinates (An example for N_2 is presented). Directly preceding the coordinates are a pair of numbers that specify the charge and spin multiplicity of the molecule, respectively. These need not be changed from the default charge (0) and multiplicity (1) except for the the O_2 molecule which has a triple ground state (multiplicity =3).

The computation parameters specify the kind of quantum chemistry computation that you would like to perform along with the method and basis set it will be obtained with. Most settings in Psi4 are changed by typing the word "set" on a line followed by the option you are specifying and the setting you would like. The "basis" option is no different, and you simply enter the name of the basis set you desire to use. To call a particular type of computation requires a slightly different syntax; a line that denotes what you would like to compute, followed by parentheses that contain the desired method in quotes. The example shown in Figure 6.16 would specify a single energy computation of N_2 at the CCSD(T)/cc-pVDZ level of theory.

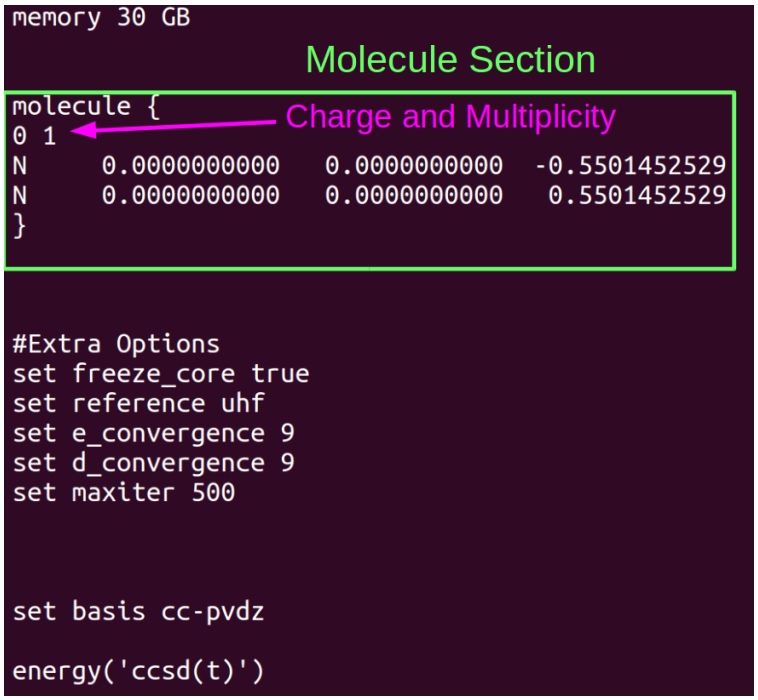

Figure 6.15: The input file structure with the molecule block highlighted in green and the charge and multiplicity options denoted by the pink text.

The extra options section in the middle of the file contains more advanced details of the computation. The extra options currently specified in Figure 6.15 should be used as default for the entire exercises. Curious students may want to explore these and other options after completing the assignment, but should not do so when gathering the lab data to avoid unnecessary errors.

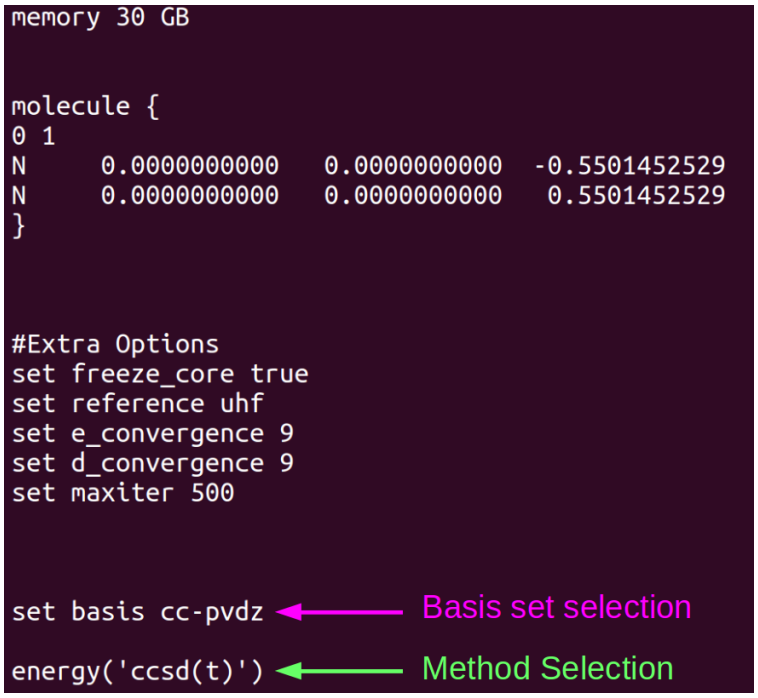

Figure 6.16: The input file structure with the basis set and method specifications denoted with pink and green arrows, respectively.

Submitting jobs

This exercise can be completed with any computer that has the Psi4 program installed properly. Installation instructions can be found on the Psi4 website here: (https://psicode.org/) with versions are available for all standard OS options (Windows, Apple, and Linux). The end of this guide contains detailed information on how to install, navigate, and run jobs using each operating system's terminal.

Users should create a new folder for each job that they will run and leave nothing inside but an input file named "input.dat". If the user is in an empty directory with an input file, Psi4 is installed, and the proper environment is active, users simply enter"psi4 input.dat" to begin the computation. If working correctly, the program should start and an "output.dat" file should be generated in the job folder. This file can be periodically checked to ensure the job is running correctly and analyzed upon completion for the desired data (example below). One should wait until the first job is finished running before submitting a second job, but most should be quite fast.

N.B. It is good practice to not overwrite completed successful jobs. The best protect from this is to put each job in its own directory and make sure the directories are well organized. Figure 6.17 presents a suggested directory structure for this experiment.

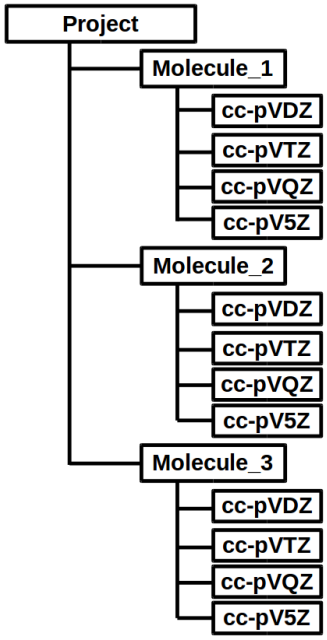

Figure 6.17: Example directory structure.

Collecting the data

The output file can be intimidating and cryptic to the untrained eye. This section will only focus on finding the results that are needed to complete this lab and should not be considered an exhaustive guide to searching the output file. Most of the lab entails generating HF, MP2, CCSD, and CCSD(T) energies at different geometries and basis sets. Note that a single CCSD(T) computation will provide the HF, MP2, and CCSD results for "free". To find these values, open the output file of a CCSD(T) job and go to the bottom. You should see something similar to Figure reftutorial4. The colored boxes highlight where each of the desired values are for this study. Note that all values here are in hartree units. Be sure to grab at least eight decimal places of precision.

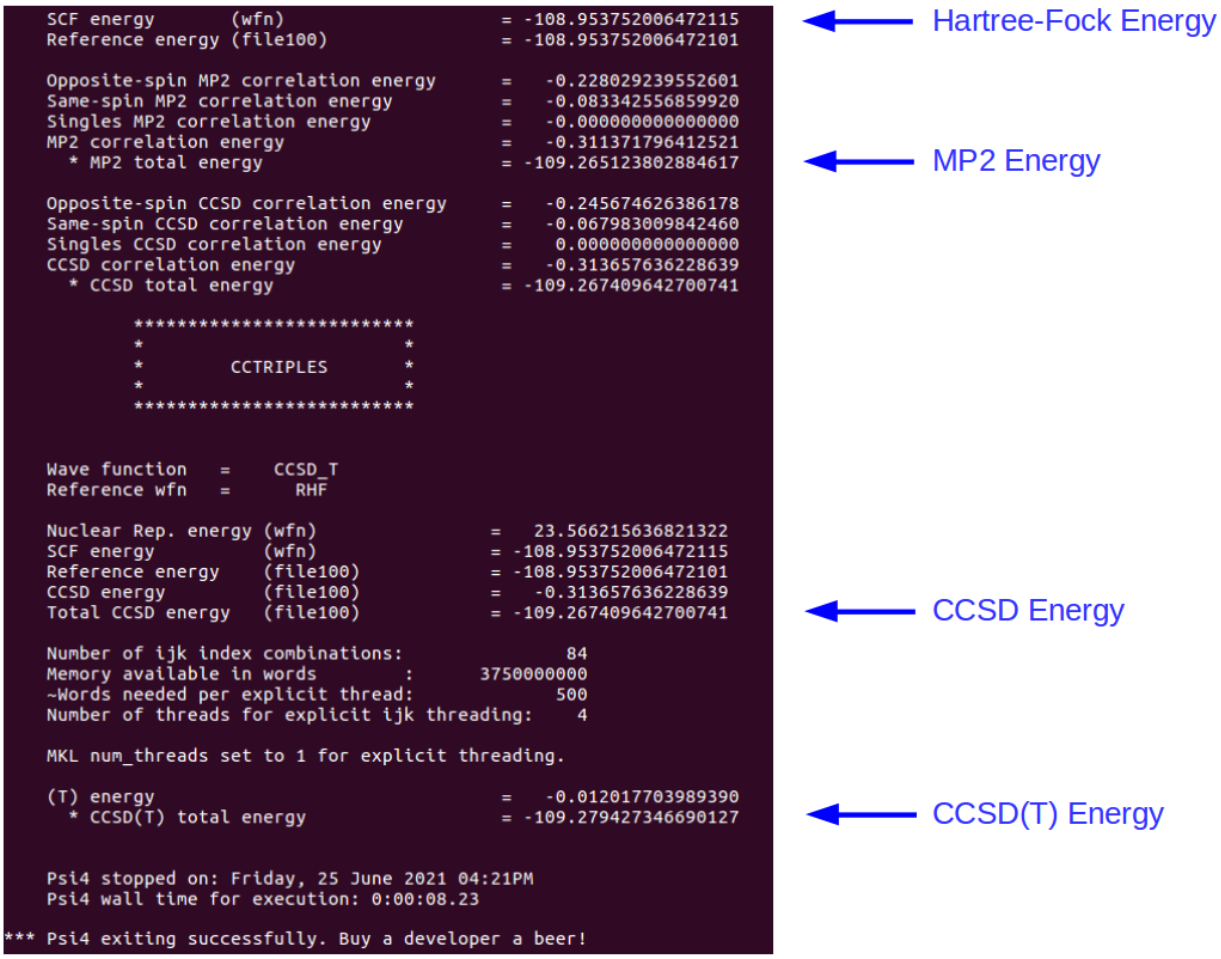

Figure 6.18: Example output file with the HF, MP2, CCSD, and CCSD(T) energy locations denoted.

Advanced computations

The last part of the lab requires a harmonic vibrational frequency computation to obtain the 298 K enthalpy correction for each species, which allows for comparison to experimental results. To do so, simply replace the "energy('method')" line with a line that says "opt('method')" followed by another line that says "freq('method')" in the input file (the geometry needs to be optimized at the desired level of theory first). Make sure you enter the correct method and basis set as well. The output files for these jobs are quite different and and example is shown in Figure. Be sure to grab the value on the leftmost column otherwise the units will be incorrect.

In order to compute the CCSD(T) energy extrapolated to the CBS limit, the line specifying the basis set should be deleted and the line calling an energy computation should be replaced with the following: "energy(cbs, corl_wfn='ccsd(t)', corl_basis='cc-pv([t,q]+d)z', corl_scheme=corl_xtpl_helgaker_2)" The final result can be obtained from the location denoted in Figure 6.20.

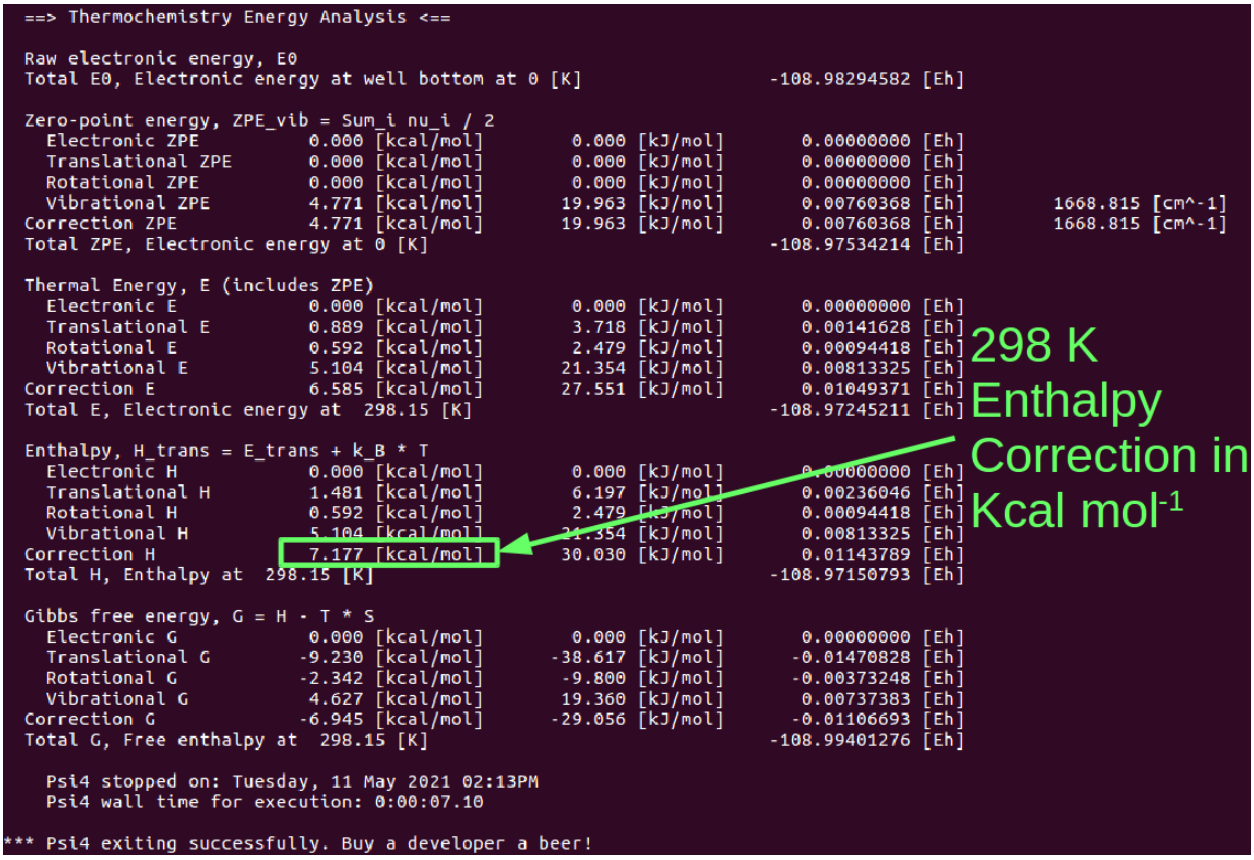

Figure 6.19: Example output file for a vibrational frequency computation. The green box denotes the 298 K enthalpy correction.

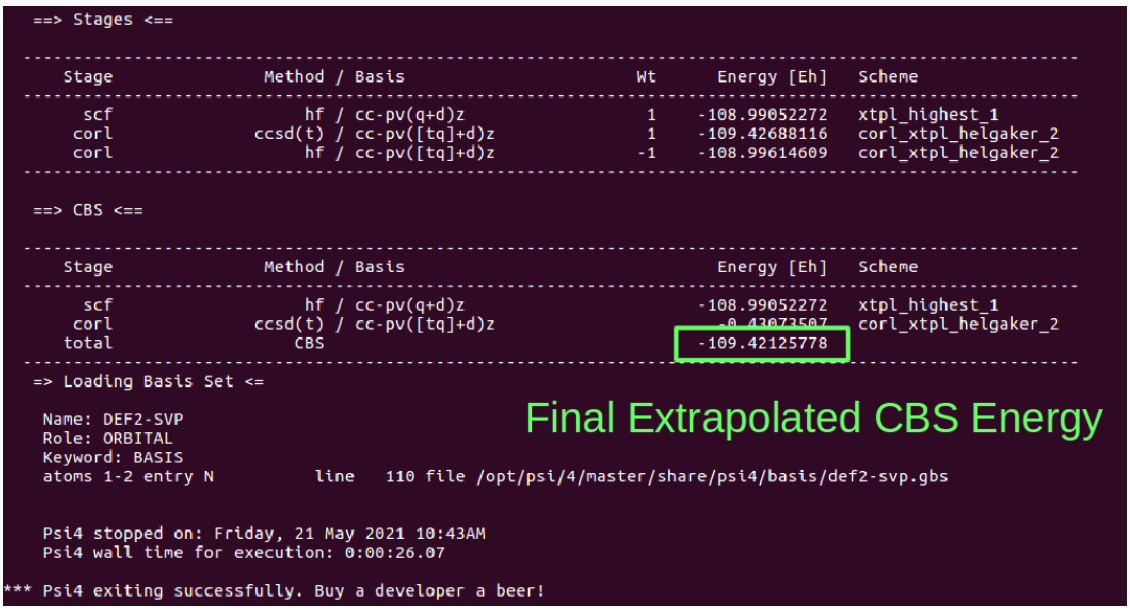

Figure 6.20: Example output file for a calculation of the CCSD(T)/CBS(3,4) energy.

Troubleshooting

This lab exercise is designed so that all of the computations should be well behaved if completed correctly. However, there is always room for human error. If an output file does not complete and raises and error, it may be difficult to discern the issue with an untrained eye. Carefully look over your input file and make sure that your syntax matches the example exactly. Use any hints from the error message to guide your inspection. If you cannot find the issue, copy an input file from a peer who successfully completed a job and modify their input.

Detailed installation and terminal instructions (all system prerequisites)

For each operating system, a similar installation and work-flow is suggested. However, the details will be slightly different for each case. Below is a general startup guide for each OS which should allow students to get Psi4 working on their personal laptops. We encourage instructors to have a basic working knowledge of these procedures ahead of time and try to minimize the number of systems spanned by the students (The number of issues would drastically decrease if students are all using the same version of an OS in a computer lab, for example). If Psi4 is already installed, students can skip to the sections that focus on navigating files and submitting jobs.

The following srteps should be performed: 1) go to psicode.org and select the download page. Select your OS, the conda tab, version 3.8, and the nightly build tab, as shown in Figure 6.21. 2) We will utilize the miniconda program to install psi4. Select the gray miniconda link which will redirect to the website where miniconda can be downloaded. Select the "Python 3.8" Miniconda3 option from your desired operating system. This will download a file that can be used to install Miniconda like you would any other program. Once that is finished, continue with the following steps pertaining to your operating system.

Detailed installation and terminal instructions (Windows)

Search for "Anaconda Prompt" in the windows program search box, which should find a link to open a miniconda terminal. In this terminal, enter the full conda line provided in the gray box on the Psi4 website. Please ensure that you are still selecting the correct OS, Psi4 version, and Python version. The program will begin to install, and eventually reach a point where the terminal asks if it should proceed. Enter "y" and continue. You can test Psi4 when it is finished installing by entering "psi4 –test" into the terminal and ensuring all of the required tests pass successfully.

Now you are ready to submit jobs of your own. We suggest leaving the conda terminal open to actually initiate Psi4, and a standard file browser open to edit and read the input and output files. In the file browser,

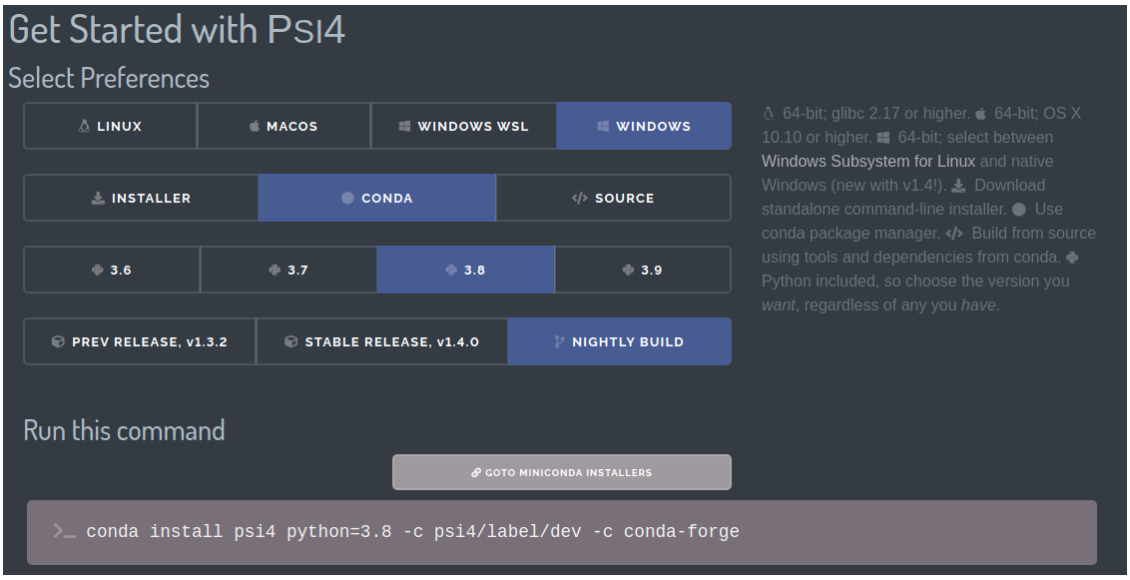

Figure 6.21: PSI4 installation webpage

find an empty directory and copy in or make the input.dat file from a plain text editor file. (be sure that the file is not saved with the .txt ending)

The job is now ready to submit, and we just need to navigate the terminal into the directory with the prepared input file and call PSI4. The terminal should start in your user home directory. You can see where you are at by entering "cd". To see what other folders are in your current directory, type the command "dir". To change directory, enter "cd target directory". To go back one layer from your current location, enter "cd ../". For example, considering the sample file structure shown above, if you are in the Project directory, you could enter "cd Molecule_2/cc-pV5Z" to enter a specific job folder and then "cd ../../" to go back to the project directory. Using these tools, navigate to the folder with the input file and enter "psi4 input.dat". The jobs should proceed and free up the terminal when it is finished. Use the open file browser to inspect the output.dat file and determine if it is finished or not. Note, some of the jobs might finish instantaneously, while others take a few minutes. Repeat the process by navigating to the other fresh folders with input files that have not yet been run.

Detailed installation and terminal instructions (Mac and Linux)

Open a new terminal. In this terminal, enter the full conda line provided in the gray box on the psi4 website. Please ensure that you are still selecting the correct OS, PSI4 version, and Python version. The program will begin to install, and eventually reach a point where the terminal asks if it should proceed. Enter "y" and continue. You can test Psi4 when it is finished installing by entering "psi4 –test" into the terminal and ensuring all of the required tests pass successfully.

Now you are ready to submit jobs of your own. We suggest leaving the terminal window open to actually initiate Psi4, and a standard file browser open to edit and read the input and output files. In the file browser, find an empty directory and copy in or make the input.dat file from a plain text editor file. (be sure that the file is not saved with the .txt ending.)

The job is now ready to submit, and we just need to navigate the terminal into the directory with the prepared input file and callPsi4. The terminal should start in your user home directory. You can see where you are at by entering "pwd". To see what other folders are in your current directory, type the command "-ls". To change directory, enter "cd target directory". To go back one layer from your current location, enter "cd ../". For example, considering the sample file structure shown above, if you are in the Project directory, you could enter "cd Molecule_2/cc-pV5Z" to enter a specific job folder and then "cd ../../" to go back to the project directory. Using these tools, navigate to the folder with the input file and enter "psi4 input.dat". The job should proceed and free up the terminal when it is finished. Use the open file browser to inspect the output.dat file and determine if it is finished or not. Note, some of the jobs might finish instantaneously, while others take a few minutes. Repeat the process by navigating to the other fresh folders with input files that have not yet been run.