

THERMOCHEMICAL AND SPECTROSCOPIC PROPERTIES FROM MAIN GROUP TO
HEAVY ELEMENT MOLECULES: A THEORETICAL PERSPECTIVE

By

Timothé Melin

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry – Doctor of Philosophy

2022

ABSTRACT

Thermodynamic properties such as enthalpies of formation, dissociation energies, and transition energies can be a challenge to determine for both experimental and theoretical chemistry. For main group elements, experimental thermodynamic data are widely available with low uncertainty with reproducibility with different experiments. However, the availability of such data is limited for the lower part of the periodic table. For heavy transition metals, lanthanide and actinide, it can be challenging to determine experimentally thermodynamic data due to the many low-lying states close to the ground state, relativistic effects, and stability of the elements. In many cases, the experimental properties are often extrapolated from other compounds and direct measurement does not exist or may lead to very large experimental uncertainties and/or inconsistencies. To aid in the investigation of such complex systems, computational chemistry can be utilized. Many *ab initio* and density functional theory methods among others, have been utilized to investigate heavy element complexes.

In this thesis, *ab initio* based methods called composite approaches are introduced and employed to investigate the enthalpy of formation of *5d* elements as well as the dissociation energies of lanthanide sulfide, selenide and halide species. Moreover, different density functional methods, the most widely used computational chemistry approach, were applied to investigate the *5d* thermodynamic data. However, as mentioned earlier, some of the heavy elements can show low-lying excited states and as such, single-reference wavefunction methods can struggle to accurately describe the correct behavior of the molecule. In these cases, multireference methods such as complete active space self-consistent field or multireference configuration interaction were utilized. Such methods have been applied to investigate the ground and excited state of the LuF molecule and used to calculate the spin-orbit coupling and determine the multireference character of a set of lanthanide diatomics.

Ab initio composite approaches, as well as density functional theory have also been utilized to study the enthalpy of formation of per- and polyfluoroalkyl substances (PFAS). These substances are often called "zombie chemicals" since they do not degrade in the environment due to their strong C-F bonds. They have been found in many different environments from soil samples, water, Antarctic ice and many other places around the globe. In this thesis, the PFAS gas phase enthalpies of formation

are investigated using the correlation consistent Composite Approach ccCA approach as well as density functional theory and coupled-cluster methods.

Finally, the interaction between light and matter can also be investigated using theoretical methods. In this case, the time-dependent Schrödinger equation needs to be solved. While many different methods exist, in this thesis, the time-dependent configuration interaction method is used to study the effect of a photon to the electron dynamics of small molecules, in particular for the ionization effects.

Copyright by
TIMOTHÉ MELIN
2022

Cette thèse est dédiée à mes parents, mes soeurs et mon frère ainsi qu'à ma femme Nicky.

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Dr. Angela K. Wilson for providing me with the opportunity to do my PhD in her group and for her guidance during the last five years. I would also like to thank all the people that have been present in the Wilson group. Thank you for all the discussions on research and otherwise that we had over the years. They helped me get through grad school and to finally defend my thesis. I would like to specifically thank Lucas and Zach as we shared the same office for a few years and for all the fun we had. Lucas, thank you for teaching me what you knew about quantum mechanics and for showing me how to be a good researcher. I would also like to thank Dr. Inga Ulusoy for her help during my first few years of grad school. I would like to thank my thesis committee: Dr. Kathy Hunt, Dr. James McCusker, Dr. Marcos Dantus and Dr. Piotr Piecuch and Dr. Warren Beck for all the help they provided during the years in both research and as a PhD student. Finally, I would like to thank the department staff and in particular Anna Osborn who truly helped me more than she can ever imagine.

Next, I would like to thank my friends in France with whom I have not spoken to much over the past years. I will always be thankful for all the good memories we had together and I am looking forward to seeing you all again once I moved back to France. I would like to thank Théo for all the conversations we had over the last years of grad school as well as the gaming nights that we both needed to take a break from the stressful life of being a graduate student.

To finish, I would like to thank my family. To my granddads Michel and DéDé as well as my grandma Marie-Ange who have all since passed on, thank you for everything and I miss you very much. To my other granddad Michel and my grandmas Thérèse and Françoise, I am really looking forward to seeing you more often once I am in France. To my family in France and my new family in the US, I am looking forward to spending more time with all of you. To my sisters Célia, Agathe and Clothilde and my brother Lilian, you are the best!! I am thankful for all of the time we spent together and I look forward for the future memories that we will create. To my parents, to whom I will always be so grateful for everything you have done for me and all the pep talks. Finally, to Nicky, my wife! You helped me survive grad school during COVID and even more. I am so ready to

spend the rest of my life with you :)

I would like to finish these acknowledgments with one of my favorite character quote that reminds me that sometimes, simpler is better.

Il en faut peu pour être heureux

Vraiment très peu pour être heureux

Il faut se satisfaire du nécessaire

TABLE OF CONTENTS

CHAPTER 1	OVERVIEW OF RESEARCH	1
CHAPTER 2	QUANTUM MECHANICS FOR COMPUTATIONAL CHEMISTRY	4
2.1	Introduction	4
2.2	Wavefunction-based methods	5
2.3	Density functional theory (DFT)	9
2.4	Approximation of the exchange-correlation term	11
2.5	Multireference methods	13
2.6	Time-dependent quantum mechanics	16
2.7	Basis sets	19
2.8	Relativity in chemistry	21
2.9	Calculation of thermodynamic properties	24
CHAPTER 3	RELATIVISTIC PSEUDOPOTENTIAL CORRELATION CONSISTENT COMPOSITE APPROACH (rp-ccCA) FOR THE 5d COMPOUNDS	31
3.1	Introduction	31
3.2	Methodology	35
3.3	Results	37
3.4	Conclusion	48
CHAPTER 4	MULTIREFERENCE CALCULATIONS ON THE GROUND AND EXCITED STATES AND DISSOCIATION ENERGY OF LuF	49
4.1	Introduction	49
4.2	Computational details	53
4.3	Results and discussion	56
4.4	Conclusion	78
CHAPTER 5	AB INITIO COMPOSITE STRATEGIES AND MULTIREFERENCE APPROACHES FOR LANTHANIDE SULFIDES AND SELENIDES	80
5.1	Introduction	80
5.2	Computational details	85
5.3	Results and discussion of results	87
5.4	Conclusion	103
CHAPTER 6	THERMOCHEMISTRY OF PER- AND POLYFLUOROALKYL SUBSTANCES (PFAS)	105
6.1	Introduction	105
6.2	Methodology	107
6.3	Results and discussion	111
6.4	Conclusions	125
CHAPTER 7	IONIZATION PROPERTIES OF DIATOMIC MOLECULE WITH TIME-DEPENDENT CONFIGURATION INTERACTION	127

7.1	Introduction	127
7.2	Ionization rate of H_2^+	128
7.3	Ionization rate of N_2	132
7.4	Conclusion	136
CHAPTER 8 FUTURE DIRECTIONS		138
BIBLIOGRAPHY		141
APPENDIX A	5 <i>d</i> DFT ENTHALPIES OF FORMATION	162
APPENDIX B	LuF MRCI CURVES AND 95% CONFIDENCE LIMIT	166
APPENDIX C	LnS AND LnSe <i>f</i> -ccCA ENERGIES AND MULTIREFERENCE DIAGNOSTICS	170
APPENDIX D	PFAS ENTHALPIES OF FORMATION, EXTRAPOLATED ENTHALPIES AND OPTIMIZED GEOMETRIES	174

CHAPTER 1

OVERVIEW OF RESEARCH

In the 1920s Schrödinger, Heisenberg and Heitler and London, respectively, published papers that became the foundation of quantum chemistry.¹⁻³ Schrödinger developed his famous equation, Heisenberg is uncertainty principle and Heitler and London, the first paper on quantum chemistry. These theories and those developed later (based on them) were initially applied to small systems (and often by hand) as the computational power was not sufficient.

Since the first computer, advances in electronic and computer engineering allowed more and more powerful computers to be built and be accessible to a growing scientific community. Computational chemistry has since then been used to understand and comprehend the world around us. While the computational hardware improved, many computational chemistry methods such as Hartree-Fock (HF), many body perturbation theory, coupled-cluster methods,⁴ configuration interaction methods,⁵ density functional theory,^{6,7} and many others, have been developed to take advantage of the available computational power. These methods are used to calculate many properties from thermochemical to spectroscopic properties.

In theory, experimental thermochemical data will always be used as a gauge of computational methods. For the early part of the periodic table, this is mostly true, and available experimental data have, for a large part, small uncertainties within different sources. However for the heavy elements, it becomes increasingly complicated to obtain reliable experimental data. Available data can have a high degree of uncertainty and/or be inconsistent between databases. The possible high uncertainty or inconsistency can often be attributed to the fact that the properties were not direct measurement but instead, extrapolated from other similar compounds. Thus, in some cases, it can become complicated to obtain reliable data and so computational chemistry can help in determining thermochemical properties.

This thesis focuses on the investigation of thermochemical properties for the *5d* and lanthanides complexes. Moreover, a part of the work has been done on understanding the stability of a particular class of organic molecules called per- and polyfluoroalkyl substances (PFAS). Finally, the ionization

properties of small organic diatomics has been investigated. Overall, the different studies presented in this thesis were motivated by the lack of experimental gas phase thermochemical properties for both the heavy elements and highly fluorinated compounds.

In Chapter 3, the relativistic pseudopotential correlation consistent composite approach (rp-ccCA) was used and the enthalpies of formation of seventeen *5d* compounds were determined. Moreover, density functional theory (DFT), one of the most used computational chemistry methods, was utilized to understand its reliability for the *5d* compounds. The enthalpy of formation was calculated using both the total atomization energy (TAE) approach and the isogyric type of approach. This study shows the importance of using the correct gauge for theoretical method and that some experimental enthalpies of formation might need to be revisited.

In Chapter 4, the spectroscopic constants of the ground and excited states of the LuF molecule were investigated employing the complete active space self-consistent field (CASSCF) and multireference configuration interaction (MRCI) methods. This part of the project was motivated by the fact that both experimental and previous theoretical studies disagree in the state ordering for the first few excited states. In the second part of this project, the dissociation energy (D_0) of the ground state was investigated since there is discrepancies between the experimental data available and theory. D_0 was calculated using *ab initio* methods as well as DFT. In both studies, the effect of different types of relativistic effects and electron correlation spaces were examined.

In Chapter 5, lanthanide sulfide (LnS) and lanthanide selenide (LnSe) were studied. The new experimental dissociation energies from Morse's group motivated this study, as they provided very accurate energies with uncertainties error of ~ 0.004 eV. Here, the *f*-ccCA approach is used where a new methodology is defined to correct for the molecular spin-orbit coupling as the initial *f*-ccCA method used only HF to correlate the electron for the spin-orbit correction. Moreover, two methods were employed to probe the multireference character of the molecules.

In Chapter 6, the enthalpy of formation of linear PFAS, alkanolic acids and perfluoroalkanes are calculated using DFT, the domain local pair natural orbital coupled-cluster single double and perturbative triple (DLPNO-CCSD(T)) and ccCA methods. The lack of experimental enthalpies

of formation in gas phase for the PFAS compounds leads to first, applied the methods on the alkanic acids which have well defined experimental data. Then, the methods were used on the perfluoroalkanes which have some experimental data. Finally, the methods were employed to determine for the first time the enthalpies of formation of the PFAS molecules as no experimental data are available.

In Chapter 7, the time-dependent configuration interaction (TDCI) method is employed to explore the electron dynamics of small diatomics when interacting with a laser pulse. The main goal of this project was to test the heuristic model, which is used to simulate ionization process. To this end, the ionization rate of H_2^+ and N_2 were investigated with the heuristic model.

CHAPTER 2

QUANTUM MECHANICS FOR COMPUTATIONAL CHEMISTRY

2.1 Introduction

The energy of an atom or a molecule is calculated by solving the time independent Schrödinger equation $\hat{H}\Psi = E\Psi$, where $\hat{H}$ is the Hamiltonian, Ψ is the wavefunction and E is the energy obtained. This equation can only be solved analytically for one-electron systems. However, for most chemical problems, more than one electron is present and no analytic solution exists. The Schrödinger equation can be solved numerically for a small number of electrons. Thus for larger systems, a number of approximations need to be made.

The first approximation often made is the Born-Oppenheimer approximation (BOA), which essentially treats the nuclei as static. This can be done because the electron is ~ 2000 lighter than the nuclei, and thus, the nuclei move on longer time scales than the electrons (in most cases), enabling the nuclei to be treated as static. The Schrödinger equation can then be expressed as:

$$\hat{H}^{el}(\vec{r}, \vec{R})\Psi^{el}(\vec{r}; \vec{R}) = E^{el}(\vec{R})\Psi^{el}(\vec{r}; \vec{R}) \quad (2.1)$$

where, $\vec{r}$ represents the electronic coordinates, $\vec{R}$ is the set of coordinates for the nuclear as a parameter and $\hat{H}^{el}$, E^{el} and Ψ^{el} are the electronic Hamiltonian, energy and wavefunction, respectively.

The bottleneck in this equation arises from the electron-electron repulsion term (two-electron integrals) which requires a four index transformation over all basis functions used in a basis set (see section 2.7). One way to simplify the two-electron term is to consider that each electron exists in the mean-field of the other electrons. In other words, instead of having to solve a many-electron problem, the equation becomes a one-electron problem where the electron-electron repulsion is treated in an average way. This approximation is called the Hartree-Fock (HF) approximation. In this method, the wavefunction is written as a Slater determinant which is a linear combination of Hartree product. A Hartree product corresponds to a wavefunction in which the electrons do not interact with each other and where each electron is described by a spin-orbital. A spin orbital depends on

the spatial and spin coordinates of the electron. The HF equation to solve is then:

$$\hat{F}(i)\chi(i) = \epsilon(i)\chi(i) \quad (2.2)$$

where $\hat{F}(i)$ is the Fock operator, $\chi(i)$ is the spin-orbital for the electron i and ϵ is the corresponding energy. The Fock operator of electron i is given by:

$$\hat{F}(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^N \frac{Z_A}{r_{iA}} + v^{HF}(i) \quad (2.3)$$

where Z_A is the atomic number of atom A and $v^{HF}(i)$ corresponds to the electron-electron repulsion replaced by the mean-field potential of the electron i . Equation 2.2 is solved iteratively until the energy ϵ no longer changes, this iterative method is called the self-consistent field (SCF) method. Finally, all of the spin-orbitals optimized by the SCF procedure are combined to give a Slater determinant Ψ_0 , from which the energy E_0 of the full system can be calculated. This method is frequently the starting point for any other wavefunction method since it often provides a good guess for the ground state wavefunction.

While the Hartree-Fock method accounts for some electron correlation (interaction between electrons), it is often not enough and thus two types of methods that include electron correlation beyond HF have been developed: those that use the wavefunction (i.e. many body perturbation theory, Møller-Plesset perturbation theory, coupled-cluster, configuration interaction, etc.) and others that do not (i.e. Kohn-Sham density functional theory (KS-DFT)).

2.2 Wavefunction-based methods

Different types of wavefunction based theory exists, among the most popular are many body perturbation theory (for example, Møller Plesset perturbation theory), coupled-cluster (CC) methods, and configuration interaction (CI) methods. These methods presents different advantages and disadvantages and will be discussed in the following sections. Among the key features of any method are: the computational cost, size-extensivity and size-consistency. The computational cost corresponds to the CPU time needed to complete a calculation, the disk space used to store the information or the required memory. The computational cost can also be represented by the scaling of the method

in terms of system size and is given as n^x (where n represents the system size and x an integer). Size-extensivity represents the scaling (linear) of a method as the number of electron increases. Finally, a method is size-consistent if the energy of two infinitely separated systems corresponds to two times the energy of one system. All these features are important when selecting a method to investigate a particular application.

Many body perturbation theory

In many body perturbation theory, the missing electron correlation (electrons interaction) energy from Hartree-Fock is obtained by using perturbation theory. To do so, the Rayleigh-Schrödinger theory starts from an unperturbed Hamiltonian $\hat{H}^0$ and adds a perturbation such that:

$$(\hat{H}^0 + \lambda \hat{V}) \Psi_0 = E_0 \Psi_0 \quad (2.4)$$

Ψ_0 is the wavefunction of the ground state, E_0 is the corresponding energy, $\hat{H}^0$ is the unperturbed Hamiltonian, and λ is a constant. This equation is then expanded in a Taylor serie to give:

$$\Psi_0 = \Psi_0^0 + \lambda \Psi_0^1 + \lambda^2 \Psi_0^2 + \dots \quad (2.5)$$

$$E_0 = E_0^0 + \lambda E_0^1 + \lambda^2 E_0^2 + \dots \quad (2.6)$$

with

$$\Psi_0^k = \frac{1}{k!} \left. \frac{\partial^k \Psi_0}{\partial \lambda^k} \right|_{\lambda=0} \quad (2.7)$$

$$E_0^k = \frac{1}{k!} \left. \frac{\partial^k E_0}{\partial \lambda^k} \right|_{\lambda=0} \quad (2.8)$$

where k represents the order of the expansion. At order 1 (when $k = 1$), equation 2.4 becomes:

$$\hat{H}^0 \Psi_0^1 + \hat{V} \Psi_0^0 = E_0^0 \Psi_0^1 + E_0^1 \Psi_0^0 \quad (2.9)$$

By solving this equation, the energy at order 1 is obtained such as:

$$E_0^1 = \langle \Psi_0^0 | \hat{V} | \Psi_0^0 \rangle \quad (2.10)$$

which corresponds to the perturbation energy over the unperturbed wavefunction. In the Rayleigh-Schrödinger perturbation theory, no assumption are made on the form of the unperturbed Hamiltonian $\hat{H}^0$.

Møller-Plesset perturbation theory

In 1934, Christian Møller and Milton S. Plesset published their paper on second order perturbation theory.⁸ The goal of this work was to add dynamic electron correlation (electrons interaction) to the Hartree-Fock method. To do so, they started from the Rayleigh-Schrödinger perturbation theory using equation 2.4 and defined the unperturbed Hamiltonian as a sum of Fock operator such as:

$$\hat{H}^0 = \sum_i \hat{F}(i) \quad (2.11)$$

where $\hat{F}(i)$ is given by equation 2.3. The energy shift: the expectation value of the perturbed Hamiltonian versus the non-perturbed state, for a given order can be determined. For the first order, the energy shift is:

$$E_0^1 = \langle \Psi_0^0 | V | \Psi_0^0 \rangle \quad (2.12)$$

For the second order:

$$E_0^2 = \sum_{m \neq n=0} \frac{\langle \Psi_n^0 | V | \Psi_m^0 \rangle \langle \Psi_m^0 | V | \Psi_n^0 \rangle}{E_n^0 - E_m^0} \quad (2.13)$$

To determine the total electronic energy, the energy shifted for a given order is added to the Hartree-Fock energy. However, it is important to note that this energy is not variational, meaning that it is possible to obtain an energy lower than the real ground state energy.

Technically, it is possible to go to any order in the perturbation. However, higher orders of perturbation theory become computationally costly, typically with limited possible improvement in energy;⁹ better improvements can be achieved by other methods for similar, or less cost. For example, MPPT(2) corresponds to the truncation of the energy expansion at the second order.

Coupled-cluster methods

Coupled-cluster (CC) theory was spread to chemistry because of the work done by Čížek in 1966.⁴ In the CC formalism, the wavefunction is written as an exponential ansatz constructed from a reference wavefunction (typically a HF determinants):

$$|\Psi_{CC}\rangle = e^{\hat{T}} |\Psi_0\rangle \quad (2.14)$$

where $|\Psi_{\text{CC}}\rangle$ is the coupled-cluster wavefunction, $|\Psi_0\rangle$ is the reference wavefunction and $\hat{T}$ is the cluster operator. $\hat{T}$ is written as the sum of excitation operator such as:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots + \hat{T}_N \quad (2.15)$$

where $\hat{T}_1$ represents the single excitation operator, $\hat{T}_2$ the double excitation operator and $\hat{T}_N$ the N-tuple excitation operator. Including only single and double excitations in the coupled-cluster theory gives the CCSD method and lead the following for $\hat{T}$:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 \longrightarrow \text{CCSD} \quad (2.16)$$

CCSD theory leads to qualitative results and thus, for direct comparison with experiment, higher excitation is usually needed (triples, quadruples, etc.). However, by including these excitations, the computational cost (in terms of CPU time, disk space) increases drastically, making computation less feasible. To improve the CCSD result without going to full triple excitations as in CCSDT:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 \longrightarrow \text{CCSDT} \quad (2.17)$$

the triple excitations can be included perturbatively at a reduced cost, leading to the CCSD(T) method. This method is one of the most popular single-reference method used in computational chemistry and is often called the "gold-standard". This popularity is due to the fact that CCSD(T) was able to achieve low error compared to experiment ($\sim 1\text{kcal}\cdot\text{mol}^{-1}$) for large sets of organic molecules.

However, CCSD(T) scales as n^7 (where n represents the system size) and thus, can become very expensive as the number of correlated electrons and basis function increases. Moreover, while CCSD(T) is very robust for typical single-reference systems (systems for which only one Slater determinant is needed to describe the wavefunction), it has been shown that this method can underperform or even fail for bond breaking, radical and high multireference character system.¹⁰ Thus, there have been development to improve both the computational time of CCSD(T) and its reliability for multireference systems.

Configuration interaction methods

Configuration interaction (CI) as all correlated methods correspond to theories in which the electrons from a reference wavefunction are excited.⁵ If all possible electrons excitation in all orbitals of a given basis set are included in CI, one obtains the so-called Full-CI. Full-CI is considered exact within a given basis set. While it is possible to use such methods, as the number of electrons to be excited and the number of unoccupied orbitals increases, the computational cost (CPU time, allocated memory, etc.) of Full-CI increases. Thus, to reduce the computational cost, the number of electron excitations can be reduced. The CI wavefunction can be represented as:

$$|\Psi_{\text{CI}}\rangle = c_0 |\Phi_0\rangle + \sum_{i,a} c_i^a |\Phi_i^a\rangle + \sum_{i<j,a<b} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle + \dots \quad (2.18)$$

where $|\Phi\rangle$ represents a single Slater determinant composed of spin-orbitals often taken from a Hartree-Fock calculation. In eq 2.18, c_0 and $|\Phi_0\rangle$ are the CI coefficient and Slater determinant for which no electrons are excited, respectively. The remaining terms correspond to the electron from spin-orbitals i, j , etc, being excited into the virtual spin-orbitals a, b , etc. The Full-CI method corresponds to a CI wavefunction where higher excitation rank corresponds to the total number of electrons. For example, for lithium, the Full-CI wavefunction will take the form:

$$|\Psi_{\text{CI}}\rangle = c_0 |\Phi_0\rangle + \sum_{i=1,2,3,a} c_i^a |\Phi_i^a\rangle + \sum_{1\leq i<j\leq 3,a<b} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle + \sum_{i=1,j=2,k=3,a<b<c} c_{ijk}^{abc} |\Phi_{ijk}^{abc}\rangle \quad (2.19)$$

If only the single and double excitations are of interest, the CI expansion is truncated at the second order to give CISD. However, the truncated CI suffers from being not size-consistent, which means that the energy of two infinitely separated systems does not correspond to two times the energy of one system. This leads to truncated-CI over/under-estimating dissociation energies and having issues representing potential energy curves. Any truncated CI methods is almost size-extensive, which means that one loses accuracy as the system size is made larger.

2.3 Density functional theory (DFT)

For *ab initio* methods like CCSD(T), the computing cost (CPU time, disk space) increases greatly with the system size. Thus, alternative methods with reduced cost are of interest.

In 1927, the first non wavefunction approach to solve the Schrödinger equation was introduced in the Thomas-Fermi model.^{11,12} Here the ground state energy of a homogeneous electron gas can be calculated as a function of the electron density (ρ). The energy is separated into kinetic and potential parts: $E = T + V_{ee} + V_{eN}$, where V_{ee} and V_{eN} are the electron-electron repulsion potential and nuclei-electron potential, respectively. Moreover, by writing the energy as a function of the electron density (ρ) instead of the wavefunction, it simplifies the calculations. This is due to the fact that the electron density only depends on three coordinates instead of the $3N$ dimension in the wavefunction (where N is the number of electrons).

The energy can be written as:

$$E^{\text{TF}}[\rho] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho^{5/3}(\vec{r})d\vec{r} - Z \int \frac{\rho(\vec{r})}{r}d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}}d\vec{r}_1d\vec{r}_2 \quad (2.20)$$

where the first term is the kinetic term associated with a system of non-interacting electrons in a homogeneous electron gas and the second and third terms are V_{eN} and V_{ee} , respectively. The last term is approximated by the classical Coulomb repulsion. The density used is obtained by the variational principle under the constraint: $\int \rho(\vec{r})d\vec{r} = N$, which means that the number of electrons is given by the integral of the density. However, this model was problematic because of the very inaccurate expression of the kinetic energy (non-interacting electron approximation) and the lack of an exchange interaction term since the electron-electron interaction is treated classically.

The first steps to improvement were introduced in 1964 by Hohenberg and Kohn.⁶ Hohenberg and Kohn theory is based on two theorems:

- A given electron density maps to a unique Hamiltonian
- If the density corresponds to the ground state density, then that density will give the lowest energy.

The functional for the energy can be written as:

$$E[\rho] = T_e[\rho] + \int V_{eN}(r) \cdot \rho(r)dr + J[\rho] + E_{xc}[\rho] \quad (2.21)$$

where, $E[\rho]$ is the energy from the density functional, $T_e[\rho]$ is the kinetic energy functional, $V_{eN}(r)$ is the potential energy between the electron and the nuclei, $\rho(r)$ is the N-electron density, $J[\rho]$ is the classical Coulomb functional, and $E_{xc}[\rho]$ is the exchange-correlation energy. The last term is included to compensate for the fact that in the classic electron-electron interaction, all of the electron pairs are treated whereas in quantum mechanics, the electrons are indistinguishable and, thus, the classic interaction overestimates the energy. This theory is called Hohenberg-Kohn-DFT (HK-DFT) and is exact because no approximations are made. However, it is impracticable since it does not give the expression of the kinetic energy and exchange-correlation terms. Thus, approximations of $T_e[\rho]$ and $E_{xc}[\rho]$ need to be made.

In 1965, Kohn and Sham⁷ provided an expression for both terms. As in the Thomas-Fermi model, the kinetic term is approximated from a non-interacting system of electrons. Thus, $T_e[\rho]$ takes the form of:

$$T_e[\rho] = \sum_{i=1}^N \left\langle \Phi_i \left| -\frac{1}{2} \nabla_i^2 \right| \Phi_i \right\rangle \quad (2.22)$$

where Φ_i is the i^{th} spin-orbital of a non-interacting system and ∇_i^2 is the Laplacien operator for electron i . However, by introducing the orbital in this scheme, the method no longer depends on the three variables (x,y,z) of the density but now on $3N$ variables which increases the cost, but it is still faster than correlated methods. This formulation is called Kohn-Sham DFT (KS-DFT) and is typically used for DFT based approaches. By giving this expression for $T_e[\rho]$, only the approximation of $E_{xc}[\rho]$ needs to be determined.

2.4 Approximation of the exchange-correlation term

There are many different approximations used for the exchange term and the most common different categories are given in Figure 2.1.

The simplest is the local density approximation (LDA) and assumes that the density can be approximated as a homogeneous electron gas. In LDA, the electrons do not interact, therefore, the exchange functional only depends on the density at each point. This approximation results in errors of $\sim 10\text{--}20\%$ for ionization energies of atoms and dissociation energies of molecules.¹³

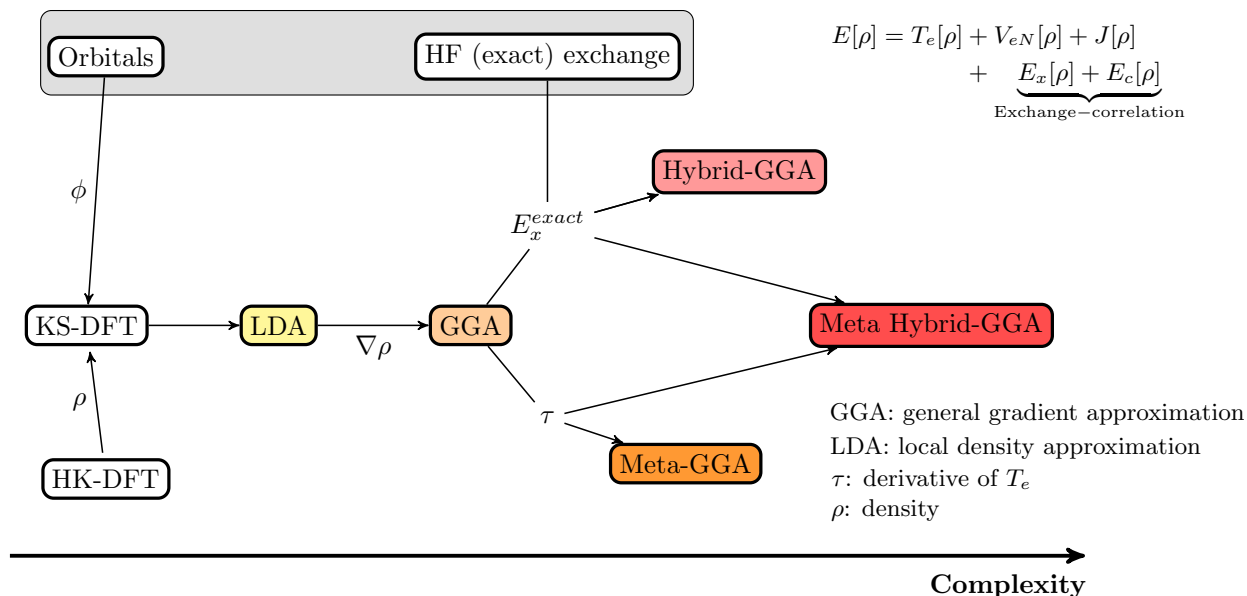


Figure 2.1: Toward the approximation of the exchange-correlation in DFT.

The next approximation is called the generalized gradient approximation (GGA). The GGA is built from the LDA and includes the gradient of the density ($\nabla\rho$). This approximation reduces the same error of LDA by a factor of 2–5, but because in atoms and molecules the density decays exponentially with respect to the nuclei-electron distances, the GGA functionals were constructed by calculating the gradient close to the nuclei. While the GGA uses the first derivative of the density, the next approximation, the meta-GGA (M-GGA), includes the second derivative of the density. This second derivative comes from the fact that the kinetic energy density is included in the functional. This approximation allows a better description of properties such as the atomization energy and dissociation energy.¹⁴ The next exchange-correlation approximation is called hybrid-GGA (H-GGA), where the exchange-correlation functional is approximated by GGA and a certain percentage of exact exchange, analogous to the Hartree-Fock exchange correction, is included. Since the results (bond dissociation, bond length or atomization energy) obtained with H-GGAs have shown very good accuracy in comparison with experimental data for dissociation and ionization energies for small organic molecules, the H-GGA functional became the most spread out in computational chemistry with B3LYP^{15,16} as the most known and used functional (see Figure 1 in Sousa¹⁷). The two last types of functionals (M-GGA and H-GGA) which are improvements upon GGA and LDA, are combined

into a new type of functional including both meta and hybrid components called meta-hybrid-GGA (MH-GGA). However, it is important to note that functionals are fit with parameters from theory and/or experimental data, thus some functionals might be more appropriate for certain applications. Moreover, due to the fitting, DFT is not systematically improvable and so different functionals may be better for different molecular systems.

2.5 Multireference methods

Until this point, all the methods described were based on a single Slater determinant to describe the wavefunction. However, while these single-reference methods can properly describe the ground state of most compounds, they can struggle to accurately describe potential energy curves and molecules that have a strong multireference character. Here, a molecule with multireference character refers to a system for which more than one Slater determinant is needed to provide a good zeroth-order approximation for the wavefunction. To study bond-breaking reactions or even molecules with nearly degenerate ground states, methods that can handle multireference system need to be used. Some solutions to this problem have been given in Kowalski and Piecuch paper¹⁸ under the form of CR-CCSD(T), however, while it improves CCSD(T) results for multireference systems, it is sometimes not enough. Instead, multireference methods can be used.

Many types of multireference methods are available, such as multiconfigurational self-consistent field,¹⁹ complete active space self-consistent field, multireference coupled-cluster or multireference configuration interaction.

Multiconfigurational Self-Consistent Field (MCSCF)

The multiconfigurational self-consistent field or MCSCF is a method in which the wavefunction (Ψ_{MCSCF}) is described by more than one Slater determinant and which corresponds to a linear combination of multiple configuration (determinants). MCSCF corresponds to a multireference Hartree-Fock. Ψ_{MCSCF} is defined as:

$$\Psi_{\text{MCSCF}} = \sum_i^{\text{states}} C_i \Phi_i \quad (2.23)$$

where Φ_i is a Slater determinant and C_i its corresponding coefficient which are both optimized. In this method, each determinants or configuration state functions (CSF) are chosen individually based on chemical intuition. A CSF corresponds to a linear combination of Slater determinants. Moreover, from MCSCF, a set of optimized orbital are obtained that can then be used for further calculations.

Complete Active Space Self-Consistent Field (CASSCF)

The complete active space self-consistent field or CASSCF method is based on MCSCF, however instead of manually choosing determinants, they are generated from a partitioning of the molecular orbitals into an "active space". The selection of these orbitals is made as Figure 2.2 shows:

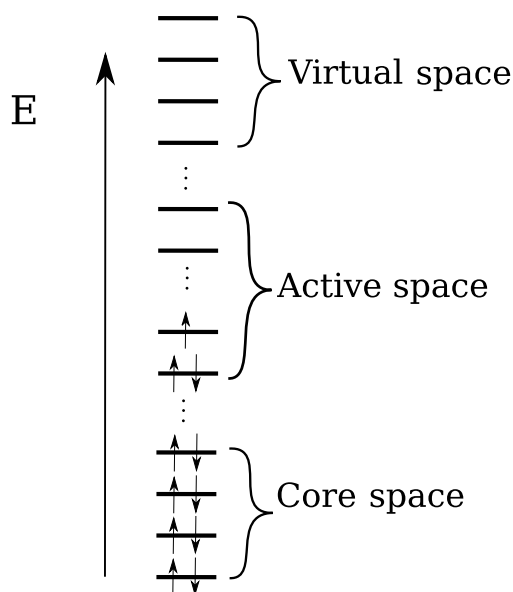


Figure 2.2: The molecular orbitals are divided in three spaces: the core space, the active space and the virtual space.

The core space corresponds to doubly occupied orbitals in which electrons cannot be excited to unoccupied orbitals. The active space represents orbitals that can be doubly, singly or unoccupied. In this space, the electrons from the occupied orbitals can be excited to the unoccupied one. Generally, the sub-valence ($n-1$) and/or valence (n) orbitals with the first virtual orbital ($n+1$) (with n representing the principal quantum number) are included in the active space.

However, it is important to note that the manual selection of the active space can become very difficult requiring chemical intuition (and at times, serendipity). Active space methods are often referred to as an "art" and less as an "exact science". Finally, the virtual space is only composed of

unoccupied orbitals and do not participate in the calculation. In the CASSCF method, all possible electron excitations are included in the active space which correspond to a Full-CI representation in that space. However, the computational cost increases steeply with the number of orbitals and electrons present in the active space. A common limitation for the CASSCF method is 20 electrons into 20 orbitals in the active space, referred to as CASSCF(20,20). One way to reduce the cost is to restrict the active space, which lead to the restrictive active space self-consistent field (RASSCF).²⁰ Moreover, both MCSCF and CASSCF lack a good description of dynamic correlation effect which can be very important for energetic and spectroscopic properties. Often a more rigorous inclusion of the dynamic correlation is done by doing multireference configuration interaction (MRCI) calculations, multireference coupled-cluster (MRCC) or multireference many body perturbation theory (MRMBPT) on top of the configuration obtained at a CASSCF or MCSCF level.

Restrictive Active Space self-consistent Field

The restrictive active space self-consistent field (RASSCF) method consists of limiting the number of possible excitations and thus reducing the number of determinants needed. In order to achieve this cost reduction, the active space from CASSCF is divided into 3 spaces: RAS1, RAS2 and RAS3.

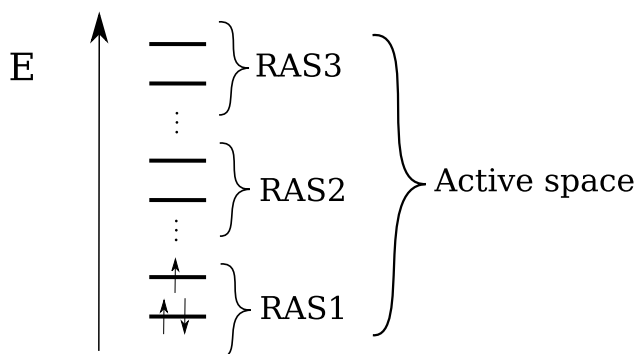


Figure 2.3: Restrictive active space.

RAS1 and RAS3 are the restricted spaces where a maximum of holes in RAS1 and a maximum of electrons in RAS3 are given. On the other hand, no restrictions are imposed on RAS2 and thus, this space can have no electrons/holes or have any number of holes and electrons (with a maximum corresponding to the number of orbitals in RAS2). The main advantage of the RASSCF method

compared to CASSCF is its flexibility and ability to include more configurations at a lower cost. In general, RASSCF methods can use a larger active space than the corresponding CASSCF. Moreover with RASSCF, it is possible to obtain a single reference method if no electrons and holes are given in RAS2. If any electrons or holes are added in RAS2, then the method becomes multireference. As for the CASSCF and MCSCF methods, a MRCI calculation is run using the reference configuration obtained from RASSCF to recover most of the dynamic correlation.

2.6 Time-dependent quantum mechanics

In the standard form of the Schrödinger equation the system is studied in a vacuum and at 0K, which is the standard condition in molecular electronic structure theory. However, to investigate properties such as ionization process, UV-visible, IR and X-ray absorption or fluorescence process, it is possible to introduce an electromagnetic field (often a laser pulse) to interact with the system. To simulate such interaction, electron dynamic methods can be employed and in particular the time-dependent electronic Schrödinger equation:

$$i\hbar \frac{\partial \Psi_e(t)}{\partial t} = \hat{H}_e(t) \Psi_e(t) \quad (2.24)$$

where i is the complex number, $\hbar$ the Planck constant and $\Psi_e(t)$ the electronic wavefunction. $\hat{H}_e$ is the electronic Hamiltonian and can be taken from different theory such as HF, MCSCF, CASSCF, CC or CI to give the corresponding time-dependent (TD) methods: TDHF,²¹ MCTDHF,²² TD-CASSCF,²³ TDCC,²⁴ TDCI²⁵ and many others. In those methods, the interaction with the light is included by modifying the time independent Hamiltonian $\hat{\mathcal{H}}_e$ such as:

$$\hat{H}_e(t) = \hat{\mathcal{H}}_e + \hat{Q}(t) \quad (2.25)$$

where $\hat{Q}(t)$ represents the interaction between the system and the external electric field. Technically, the full multipole (dipole electric, quadrupole electric, dipole magnetic, etc) must be introduced in $\hat{Q}(t)$.²⁶ However, for most application the multipole expansion can be truncated, in particular when the wavelength of the light is larger than the size of the molecule. In these cases, the dipole approximation is used, which means that only the dipole operator $\hat{\mu}$ is used in the $\hat{Q}(t)$. This

approximation is only valid when the wavelength of the light interacting with the system is longer than the size of the system. In the dipole approximation, the most common gauges that can describe the interaction between the field and the system, are: the length and velocity gauge. In the length (or electric field) gauge, $\hat{Q}(t)$ take the form:

$$\hat{Q}(t) = -\hat{\mu}E(t) \quad (2.26)$$

When using the length gauge, equation 2.25 becomes:

$$\hat{H}_e(t) = \hat{\mathcal{H}}_e - \hat{\mu}E_p(t) \quad (2.27)$$

where the electric field is often described as a pulse and is polarized along the x, y and z axis. The field is often treated classically and the envelope is described by a $\cos^2$, $\cos \sin$ or $\sin^2$ function. For example, Ulusoy and co-workers used an electric pulse described as:²⁵

$$E_p(t) = E_{p,0} \cos^2 \left(\frac{\pi}{2\sigma}(t - t_0) \right) \cos \omega(t - t_0) \quad (2.28)$$

where $E_{p,0}$ is the maximum field strength, ω is the carrier frequency, t_0 is the time at which the pulse is maximal and σ the pulse width.

Time-dependent configuration interaction (TDCI)

In TDCI,^{25,27} the wavefunction $\Psi_{\text{CI}}(t)$ is described as the linear combination of each CI state such as:

$$\Psi_{\text{TDCI}}(t) = \sum_i C_i(t) \Psi_i \quad (2.29)$$

where only the coefficients $C_i(t)$ are time-dependent and Ψ_i is given by equation 2.18. The coefficients $C_i(t)$ are written as:

$$C_i(t) = e^{-i\hat{H}(t)} \cdot C_i(0) \quad (2.30)$$

where $C_i(0)$ are the CI coefficients at $t = 0$ and $\hat{H}$ is the CI Hamiltonian taken from eq. 2.25 with $\hat{Q}(t)$ being either the length or the velocity gauge. To solve this equation in a presence of an

electric field, there exist different propagator such as: split-operator,^{28,29} Adams,³⁰ or Runge-Kutta methods.³¹ If no external field are present, only equation 2.30 is needed.

Here the split-operator approximation method is used which lead to the following expression:

$$e^{-i\hat{H}(t+\Delta t)}C_i(t) \approx \left(e^{-i\hat{H}_0(t+\Delta t/2)}\mathbf{U}^\dagger e^{-i\hat{D}(t)(t+\Delta t)}\mathbf{U} e^{-i\hat{H}(t+\Delta t/2)} \right) C_i(t) \quad (2.31)$$

where $\mathbf{U}$ is a unitary matrix that transform the eigenfunction of $\hat{H}_0$ (field-free system) into the eigenfunction of $\hat{H}(t)$ which is the system-field interaction and vice versa. Here, both $\hat{H}_0$ and $\hat{D}(t)$ are diagonal matrices. This approximation is correct as long as the time step (Δt), which is dependent on the field strength, is small. It is important to note that TDCI is gauge variant and thus depending on how the interaction system-field is represented (length or velocity) different results can be obtained for transition dipoles.²⁵

Ionization process within TDCI

There exists two types of ionization processes, below and above ionization threshold. In the former, the ionization happens through the tunneling effect and is observed when the laser pulse energy is lower than the ionization potential (IP). Above threshold ionization processes are present when the laser pulse energy is greater than the IP. In a typical TDCI calculation, Gaussian type orbitals (GTO) are used to model the space in which an electron moves (see section 2.7). However, the use of this finite basis set leads to the wrong description of the continuum states (states higher in energy than IP) due to the fact that GTO decay exponentially as the distance nuclei-electron increases. This fast decay allows the description of above ionization process. To solve this issue, an approximation called a heuristic model has been proposed in the literature, which consist of introducing a decay-time to the C_i states with energies above the IP.^{29,32} In this model, the energy of a CI state j (E_j) is replaced by a complex energy written as:

$$E_j \rightarrow E_j - \frac{i}{2}\Gamma_j \quad (2.32)$$

where i is the complex number and Γ_j is the ionization rate of the CI state j . In this model, the ionization potential is defined as the absolute value of the Hatree-Fock HOMO energy (Koopman's

theorem). The ionization rate for the CIS method is then defined as:

$$\Gamma_j^{\text{CIS}} = \begin{cases} 0 & \text{if } E_j < \text{IP} \\ \sum_a \sum_r |D_{a,j}^r|^2 \frac{1}{d} \sqrt{\epsilon_r} & \text{if } E_j \geq \text{IP and } \epsilon_r > 0 \end{cases} \quad (2.33)$$

where a and r are molecular orbitals (MO) indices, $D_{a,j}^r$ is the probability of the electron being excited from MO a to MO r in the CI state j and ϵ_r is the energy of the corresponding MO r . In eq. 2.33, d is an empirical parameter that represents the electron escape length. It is important to note that the heuristic model is only valid when above ionization threshold processes are investigated. Moreover, as shown by Klinkusch *et. al.*, this model does not include a lifetime for states below IP which can lead to some errors, especially when the states below IP interact greatly with the one above the IP.³²

2.7 Basis sets

For quantum calculations, the space mapped by the electrons must have a mathematical representation represented as a probability distribution. This comes in the form of a basis set and is used to solve the basic HF, KS-DFT or MCSCF equations. A linear combination of the basis set functions ϕ_i times a coefficient can be used to represent the wavefunction Ψ , in the following way,

$$\Psi = \sum_{i=1}^N c_i \phi_i \quad (2.34)$$

where c_i are the expansion coefficients and N the number of basis functions.

Each basis function needs to properly represent the real behavior of an electron in order to reduce the number of basis functions used to build the molecular orbital. Moreover, a basis set is defined complete when all basis functions completely span the space. Thus, a basis is complete if adding a basis function no longer changes the electronic energy, this process (reaching the completeness) is called the complete basis set limit. Finally, the basis set should provide a good balance between desired accuracy and the computational cost.

For a hydrogen-like (one-electron) system, the exact solution of the Schrödinger equation is known and thus its orbitals. Therefore, a good choice for a basis set would be one composed of

atomic hydrogen-like orbitals. For these orbitals, the radial wavefunction has a form as shown in Figure 2.4:

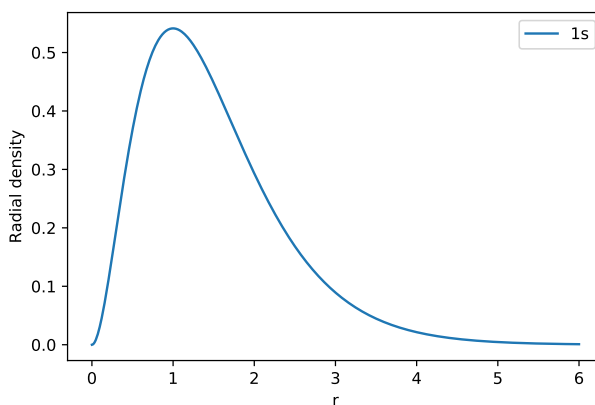


Figure 2.4: Representation of the radial distribution of the hydrogen 1s orbital.

Two main types of basis functions have been developed: Slater type orbitals (STO)³³ and Gaussian type orbitals (GTO).³⁴ When a molecule is studied, the molecular orbitals are obtained by linear combination of atomic orbital (LCAO). However, linear combinations of STOs are very difficult to compute, the two-electron integrals have to be solved numerically and are costly in terms of CPU time and memory allocated. Because of that, most quantum software uses GTOs. Many Gaussian functions can be combined to describe the hydrogen-like orbitals to give a STO. However, GTOs do not represent the behavior of the electron very well in the region close to the nucleus and they fall too slowly compared to the STOs. Moreover, the linear combination of GTOs gives a GTO and thus makes the GTOs suitable for numerical calculations of the two-electron integrals. When multiple Gaussian functions are combined, the function obtained is called contracted GTO.

There are two different schemes of contraction: the segmented³⁵ and the general³⁶ contraction. Many types of segmented basis sets were developed such as the STO- n G from Pople group³⁷ where n Gaussians are contracted in a Slater type orbital or the split-valence basis set such as 6-31G, 6-31G*.³⁸ In terms of general contracted basis sets, Dunning introduced the correlation consistent (cc) basis sets³⁹ for electron correlation calculations, Roos and co-workers developed the atomic natural orbital (ANO)⁴⁰ which were obtained by using the coefficient from a natural orbital Complete Active Space Perturbation Theory (CASPT2) calculation.

2.8 Relativity in chemistry

When describing molecules relativity needs to be treated, and as the atomic number increases, the effect of relativity on the properties increases. There are two main effects in relativity: the spin-orbit coupling (SO) and the scalar relativistic (SR) effects. SO coupling is a result of the magnetic field created by an electron's ability to interact with the magnetic field created by the same or another electron moving around the nucleus. In other words, the spin of an electron interacts with its own angular momentum. SO causes splitting of orbitals with the same angular momentum. For example, in the non-relativistic case, there are three degenerate p orbitals (p_x , p_y and p_z) whereas employing relativistic theory, the p 's are split energetically into two non-degenerate levels, the $p_{\frac{1}{2}}$ and $p_{\frac{3}{2}}$ (Figure 2.5).

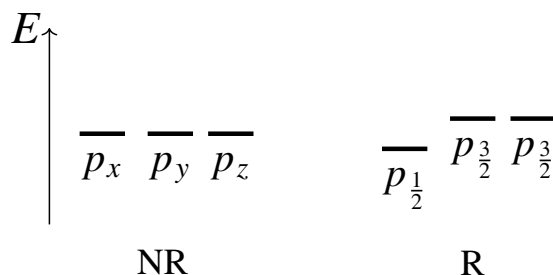


Figure 2.5: Spin-orbit splitting of the p orbitals in relativistic theory.

The SO coupling mainly affects properties such as the dissociation and ionization energies. The second main effect, the scalar relativistic (SR) effect, consists of a direct and an indirect part. The direct effect corresponds to a contraction and stabilization of all of the s and most of the $p_{\frac{1}{2}}$ orbitals. In the innermost shell, the speed of the electron reaches a fraction of the speed of light ($>0.1c$). From the theory of relativity, the law of physics must be the same in all inertial reference frames, which implies that the speed of light must be a constant. This leads to the length contraction. When the speed of an object increases, its mass increases too. The radius of an electron depends on its mass thus, when the mass increases, the radius decreases. This implies the contraction of the s and $p_{\frac{1}{2}}$ orbitals. While these two types of orbitals stabilize, the d and f destabilize. This indirect effect is due to a screening of the nucleus charge by the s and $p_{\frac{1}{2}}$ electrons. Because the distance between

those electrons and the nucleus changes, the charge experienced by the d and f change as well. Thus, shielding some of the nuclear charge, allows the electron density to expand.

There are many ways to include the relativistic effects, but two will be considered herein. The first is by using a pseudopotential to represent the core electrons and the appropriate basis set, and the second is by using an all-electron basis set and a relativistic Hamiltonian. The advantage of using a relativistic Hamiltonian is that it reaches a higher accuracy as compared to experimental data but at a high computational cost. On the other hand, the pseudopotential method is much cheaper but does not provide the same level of accuracy.

Relativistic Hamiltonian

In 1928, Paul Dirac⁴¹ introduced a one-electron relativistic Hamiltonian:

$$\hat{H}^R = \beta mc^2 + c(\boldsymbol{\alpha} \cdot \mathbf{p}) \quad (2.35)$$

This Hamiltonian depends on the rest mass (m), the speed of light c , the momentum $\mathbf{p}$ and two 4×4 matrices β and $\boldsymbol{\alpha}$, with $\boldsymbol{\alpha}$ being build upon the Pauli's spin matrices. This Hamiltonian implies a four-component (4c) wavefunction. The wavefunction has large and small components (ψ^L and ψ^S) each having a different spin variable. When the free-particle Dirac equation is solved, it leads to four solutions (one for each component). In this case, the solution of the upper two components (energies) are much larger than the two lower ones. Moreover, the energy obtained from the large component (upper) is positive whereas the small component leads to a negative energy. However, this Hamiltonian cannot be used directly for many-electron systems as it was developed for a one electron system.

The Dirac Hamiltonian leads to the electronic (positive) and positronic (negative) energy, but only the electronic energy is needed to solve chemical problems. Thus the Hamiltonian can be reduced to show only the electronic spectrum, giving a two-component (2c) Hamiltonian. One way to do that is to treat perturbatively the Dirac Hamiltonian. This leads to the well-known Pauli

Hamiltonian:⁴²

$$\hat{H}^P = \hat{H}^{NR} - \underbrace{\frac{\mathbf{p}^2 V}{8m^2 c^2}}_{\text{Darwin}} - \underbrace{\frac{\mathbf{p}^4}{8m^3 c^2}}_{\text{Mass-velocity}} + \underbrace{\frac{i\boldsymbol{\sigma} \cdot (\mathbf{p}V) \times \mathbf{p}}{4m^2 c^2}}_{\text{Spin-Orbit}} \quad (2.36)$$

where $\hat{H}^{NR}$ corresponds to the non-relativistic Hamiltonian. The mass-velocity term comes from the expansion of the classical relativistic Hamiltonian and corresponds to the change of mass with the velocity. The Darwin term corresponds to a perturbation in the trajectory of the electron. However, this Hamiltonian was developed from perturbation theory and thus has to be used only in this theory. Moreover, it was developed for the one-electron system only and thus can not be used for a many-electron system.

Other two-component (2c) Hamiltonians exist such as: the zero order regular approximation (ZORA),⁴³ the exact two component one electron (X2C)⁴⁴ and the Douglas-Kroll-Hess (DKH) Hamiltonians⁴⁵ as well as many others. Among these methods, DKH is the most widely used among the computational chemistry community. The idea behind the DKH method is to reduce the Dirac Hamiltonian to two components by repeating several unitary transformations and thus eliminate the positronic energy solutions. The advantage of this technique is that it allows variational stability. However, the operators obtained are complicated and the matrix elements can only be calculated by numerical methods.

Non-relativistic Hamiltonian and pseudopotentials

Another means to include relativity in a less expensive way is to use a pseudopotential. Because the core electrons do not play a large role in most bonding interactions, they can essentially be replaced by a parameterized one-electron operator. This means that the core and valence electrons are treated separately and thus the core-valence correlation is not accounted for. In relativistic chemistry, the effective core pseudopotentials (ECP) are used since they are parameterized to include relativistic effects and are cheaper compared to 4c and 2c Hamiltonian while maintaining accuracy compared to the experimental data and the results obtained from 4c and 2c methods. The ECP needs to have two characteristics: to represent the core electrons with a nodeless pseudo-orbital and at the same time

to accurately describe the valence wavefunction. Thus, the pseudopotential is often represented as:

$$\Psi(\mathbf{r}) \rightarrow \tilde{\Psi}(\mathbf{r}) = \begin{cases} \Psi(\mathbf{r}) & r \geq r_C \text{ (original valence orbital)} \\ f(\mathbf{r}) & r < r_C \text{ (pseudo-orbital)} \end{cases} \quad (2.37)$$

where r_C represents the distance between the nucleus and the farther core electron.

Using a pseudopotential has many advantages in terms of computation. It allows a smaller basis set than an all electron (AE) basis set to be used. Moreover, there are many types of ECPs characterized by the number of core electrons. For example, for the f -elements, there are small and large core pseudopotentials. However, the choice of the ECP needs to be carefully made; in general the number of core electrons will play a key role in the description of the molecule properties.⁴⁶ Finally, each ECPs is provided with a corresponding basis set which means that not all basis sets will work for a given pseudopotential. For example, for the $5d$, the ECP and the basis set were published in the same paper.⁴⁷

2.9 Calculation of thermodynamic properties

Thermodynamic properties are critical to chemistry. They help to understand the behavior of the compound at different conditions (temperature, pressure, etc.). In calculations, many effects may need to be addressed (electron correlation, relativistic effects for heavy elements, etc.) to correctly predict thermodynamic properties. An ideal approach would be to do a Full-CI calculation including all electrons with a 4c Hamiltonian and using a large basis set near completeness, however this is not computationally feasible for most systems due to the high costs of configuration interaction calculation and calculation using 4c Hamiltonian. Thus, many methodologies have been developed to reduce the cost with minimal losses of accuracy, and among the most widely used approaches are composite approaches.

Composite approaches

The goal of a composite approach is to approximate the results of high-level, computationally expensive, though reliable methods with a series of less expensive and less accurate methods. To do

so, these distinct calculations are done at different levels of theories and basis sets. By combining all of the pieces, the targeted accuracy is approached.

Many composite approaches have been developed such as the Gaussian-n method,^{48,49} the Peterson style CBS model,⁵⁰ the Weizman-n,^{51,52} (W-n) the High-accuracy extrapolated *ab initio* thermochemistry⁵³ (HEAT), the Feller-Peterson-Dixon procedure⁵⁴(FPD) and the correlation consistent composite approach⁵⁵⁻⁵⁷ (ccCA).

Correlation consistent composite approach (ccCA)

The correlation consistent composite approach or ccCA was initially developed in 2006 by De Yonker *et. al.*⁵⁵ The main goal was to propose an alternative to the Gaussian-2 method developed by Pople and co-workers in 1989 which include empirical parameters.⁴⁸ The fact that no empirical parameters are present in the ccCA methodology, allow the method to be extended for any element in the periodic table. Although the name is the same, the current version of ccCA differs from the original ccCA methodology. Nowadays, the ccCA energy E_{ccCA} is expressed as:

$$E_{\text{ccCA}} = E_{\text{ref}} + \Delta E_{\text{CV}} + \Delta E_{\text{CC}} + \Delta E_{\text{SR}} + \Delta E_{\text{SO}} \quad (2.38)$$

where E_{ref} is the reference energy, ΔE_{CV} is the electron correlation between the valence and the sub-valence electrons, ΔE_{CC} is the correlation energy beyond MP2, ΔE_{SR} is the scalar relativistic contribution to the energy and ΔE_{SO} is the relativistic spin-orbit coupling correction energy. In this equation, Δ represents the energy difference between different basis set or methods as show in the following equations. The reference energy is composed of two terms: a Hartree-Fock and a MP2 correlation energy term. The Hartree-Fock term is obtained with a two points extrapolation scheme:^{58,59}

$$E(n) = E_{\text{CBS}} + Ae^{-1.63n} \quad (2.39)$$

where $n = \text{T, Q}$ represent the basis set level at the aug-cc-pVnZ level and A is a constant. The constant 1.63 was found to be optimal for the extrapolation of HF energy. The MP2 correlation energy is

calculated using a three-point extrapolation scheme by Peterson *et. al.*:⁶⁰

$$E_n = E_{\text{CBS}} + B e^{-(n-1)} + C e^{-(n-1)^2} \quad (2.40)$$

where B and C are two constants and $n=D,T,Q$ is the basis set level at the aug-cc-pV n Z level. The CBS terms are here to compensate for the fact that it is impossible to use an infinite basis set.

The core-valence correction ΔE_{CV} is calculated such as:

$$\Delta E_{\text{CV}} = E_{\text{MP2}(FC1)/\text{aug-cc-pCVTZ}} - E_{\text{MP2}/\text{aug-cc-pVTZ}} \quad (2.41)$$

where $\text{MP2}(FC1)$ is a MP2 calculation where the subvalence electrons are added in the correlation space. In order to account for more electron correlation, the ΔE_{CC} term is added to the reference energy. It is computed using CCSD(T) and MP2 such as:

$$\Delta E_{\text{CC}} = E_{\text{CCSD(T)}/\text{cc-pVTZ}} - E_{\text{MP2}/\text{cc-pVTZ}} \quad (2.42)$$

Finally, two terms are added to account for relativistic effects. The first term, ΔE_{SR} is computed using the Douglas-Kroll-Hess Hamiltonian:

$$\Delta E_{\text{SR}} = E_{\text{MP2}/\text{cc-pVTZ-DK}} - E_{\text{MP2}/\text{cc-pVTZ}} \quad (2.43)$$

and the second, ΔE_{SO} , is calculated by taking an average over the available spin multiplet atomic energies for the lowest multiplets using the expressions:

$$\Delta E_{\text{SO}} = \frac{\sum_J (2J+1) \Delta E_J}{\sum_J (2J+1)} \quad (2.44)$$

where, J is the total angular momentum ($J = L + S$) and ΔE_J is the corresponding energy level. The energy levels are taken from the experimental data.

This formulation of ccCA can be used for main group molecules as well as the $2p$ and $3p$ blocks and is the starting point for all other types of ccCA method.

Relativistic pseudopotential correlation consistent composite approach (rp-ccCA)

The rp-ccCA method is the ccCA method adapted for transition metal complexes. It was developed for $4d$ transition metal complexes in 2011 by Laury and coworkers.⁶¹ This approach is

mainly based on MP2 and CCSD(T) calculations and includes a relativistic effect through the use of a pseudopotential. The rp-ccCA energy is computed through the following expression:

$$E_{\text{rp-ccCA}} = E_{\text{ref}} + \Delta E_{\text{CC}} + \Delta E_{\text{CV}} + \Delta E_{\text{SO}} \quad (2.45)$$

where E_{ref} is the reference energy and calculated at the complete basis set limit, ΔE_{CC} corresponds to the correlation effect, ΔE_{CV} is the core valence correlation and ΔE_{SO} is the spin-orbit term.

This E_{ref} term corresponds to a MP2 and Hartree-Fock complete basis set limit calculations. The former uses a three point extrapolation scheme:⁶⁰

$$E_n = E_{\text{CBS}} + B e^{-(n-1)} + C e^{-(n-1)^2} \quad (2.46)$$

where B and C are two constants, n is the basis set level (double-, triple- and quadruple- ζ). The Hartree-Fock energy is obtained with a two points extrapolation scheme:^{58,59}

$$E(n) = E_{\text{CBS}} + A e^{-1.63n} \quad (2.47)$$

where $n = \text{T, Q}$ represent the basis set level at the aug-cc-pVnZ level and A is a constant and the 1.63 being optimal for the HF energy extrapolation.

However, even if MP2 does include some correlation, other methods such as coupled-cluster can recover more correlation. Thus, correlation beyond the MP2 level is computed through ΔE_{CC} by the following equation:

$$\Delta E_{\text{CC}} = E_{\text{CCSD(T)/cc-pVTZ-PP}} - E_{\text{MP2/cc-pVTZ-PP}} \quad (2.48)$$

where $E_{\text{CCSD(T)}}$ and E_{MP2} are energies calculated at the CCSD(T) and MP2 levels respectively, with the cc-pVTZ-PP basis set.

Since the valence electrons are at the same time responsible for the bonding but also are dependent on the core electrons, it is necessary to include the correlation between the core and valence electrons. This is done in the ΔE_{CV} correction term through the following calculation:

$$\Delta E_{\text{CV}} = E_{\text{CCSD(T,FC1)/aug-cc-pCVTZ-PP}} - E_{\text{CCSD(T)/aug-cc-pVTZ-PP}} \quad (2.49)$$

where, FC1 corresponds to a correlation space where the sub-valence electrons are included. For example, the $1s$, $2s$ and $2p$ electrons of chlorine are frozen in each step, but for the CV term, these electrons are included in the calculations.

Finally, the ΔE_{SO} term corresponds to the correction to the atomic spin-orbit. It is calculated by taking an average over the available spin multiplet energies for the lowest multiplets using the expressions:

$$\Delta E_{SO} = \frac{\sum_J (2J+1) \Delta E_J}{\sum_J (2J+1)} \quad (2.50)$$

where, J is the total angular momentum ($J = L + S$) and ΔE_J is the corresponding energy level. The energy levels are taken from the experimental data.

***f* element correlation consistent composite approach (*f*-ccCA)**

The *f*-ccCA methodology is based on the standard ccCA and was developed for the *f* elements.^{62,63} While ccCA only needed a correction to the scalar relativistic and spin-orbit coupling effects, the *f*-ccCA methods includes the scalar relativistic effect at each step through the use of the Douglas-Kroll-Hess Hamiltonian at the 3rd order. The *f*-ccCA energy is defined as:

$$E_{f\text{-ccCA}} = E_{\text{ref}} + \Delta E_{CC} + \Delta E_{CV} + \Delta E_{SO} \quad (2.51)$$

where ΔE_{ref} is the reference energy, ΔE_{CV} is the electron correlation between the valence and the sub-valence electrons, ΔE_{CC} is the correlation energy beyond the second order perturbation and ΔE_{SO} is the relativistic spin-orbit coupling correction energy.

As for ccCA, the reference energy is composed of two terms: a Hartree-Fock and a MP2 correlation energy term, both obtained with an extrapolation scheme. The Hartree-Fock term is calculated at the complete basis set limit with a two-points extrapolation scheme:^{58,59}

$$E(n) = E_{\text{CBS}} + A e^{-1.63n} \quad (2.52)$$

where $n = T, Q$ represent the basis set level at the aug-cc-pVnZ-DK level and A is a constant. The MP2 correlation energy is calculated using a three-point extrapolation scheme by Peterson *et. al.*:⁶⁰

$$E_n = E_{\text{CBS}} + B e^{-(n-1)} + C e^{-(n-1)^2} \quad (2.53)$$

where B and C are two constants and $n=D,T,Q$ is the basis set level at the aug-cc-pVnZ-DK level. The core-valence correction ΔE_{CV} is calculated such as:

$$\Delta E_{CV} = E_{CCSD(T,FC1)/aug-cc-pCVTZ-DK} - E_{CCSD(T)/aug-cc-pCVTZ-DK} \quad (2.54)$$

where $CCSD(T, FC1)$ is a CCSD(T) calculation where the sub-valence electrons are added in the correlation space. In order to account for more electron correlation, the ΔE_{CC} term is added to the reference energy. It is computed using CCSD(T) and MP2 such as:

$$\Delta E_{CC} = E_{CCSD(T)/cc-pVTZ-DK} - E_{MP2/cc-pVTZ-DK} \quad (2.55)$$

To include the spin-orbit coupling, the ΔE_{SO} term is calculated using the following equation:

$$\Delta E_{SO} = SO(M) - \sum_i SO(X_i) \quad (2.56)$$

where $SO(M)$ is the spin-orbit coupling obtained at the multireference configuration interaction or at the Hartree-Fock average of configuration and $SO(X_i)$ is the atomic spin-orbit coupling obtained from experiment. The atomic spin-orbit is calculated using the same J average equation as described for ccCA and rp-ccCA.

Thermodynamic properties

The thermodynamic properties can be determined by two main routes: those that calculate the total atomization energy (TAE) and the others called isogyric approach based on an actual chemical equation.

The general scheme of the TAE approach is the following:

$$\begin{aligned} \Delta H_f^{298\text{ K}}(M) = & \sum_i^{atoms} \nu_i \Delta H_f^{0\text{ K}}(X_i) - \sum_i^{atoms} \nu_i [H_{298}(X_i) - H_0(X_i)] \\ & - \sum D_0^{0\text{ K}} + [H_{298}(M) - H_0(M)] \end{aligned} \quad (2.57)$$

where $\Delta H_f^{298\text{ K}}(M)$ is the enthalpy of formation at room temperature and gas state of the molecule, ν_i is the stoichiometric coefficient, $D_0^{0\text{ K}}$ is the atomization energy, $[H_{298}(M) - H_0(M)]$ is the enthalpy correction of the molecule (zero-point vibrational energy (ZPVE)), $\Delta H_f^{0\text{ K}}$ is the enthalpy

of formation at 0 Kelvin of the atoms and $[H_{298}(X_i) - H_0(X_i)]$ is the enthalpy corrections for the atoms. The two last terms are taken from experiment while the others are computed. $D_0^{0\text{ K}}$ is calculated through the composite method or with other methods, while the ZPVE term is obtained using the harmonic oscillator approximation.

The isogyric approach was developed by Pople^{64,65} in the 70's and can be applied directly to any molecule. It was originally applied to hydrocarbon compounds but can be generalized for any type of compounds. Many types of isogyric approaches exist but they all rely on the same principle, that the number of electron pairs or bond types is the same on both sides of the chemical equation. The enthalpy of formation is calculated by using the following equation:

$$\Delta H_r^{298\text{ K}} = \sum_i^{\text{products}} \Delta H_f^{298\text{ K}}(i) - \sum_j^{\text{reactants}} \Delta H_f^{298\text{ K}}(j) \quad (2.58)$$

where $\Delta H_r^{298\text{ K}}$ is the enthalpy of reaction and is calculated with a composite approach or with other methods and $\Delta H_f^{298\text{ K}}$ is the experimental enthalpy of formation of each molecules involved in the reaction.

CHAPTER 3

RELATIVISTIC PSEUDOPOTENTIAL CORRELATION CONSISTENT COMPOSITE APPROACH (rp-ccCA) FOR THE 5d COMPOUNDS

This chapter will be submitted as a paper in the journal *Chemical Physics Letters*. All the geometries were optimized by Dr. Lucas Aebersold. The enthalpy of formation with the atomization and isogyric approaches with rp-ccCA and DFT were calculated by Timoth   Melin. The original draft was written by Timoth   Melin. The finalized paper was developed by Timoth   Melin and Dr. Angela K. Wilson.

3.1 Introduction

Since the beginning of the twentieth century, 5d heavy elements have been used in many different industries due to their unique properties. The large resistance of 5d transition metals to extreme conditions, for example, have made them vital in aerospace engineering. The addition of HfB₂ to a SiC/C ceramic, for example, is used to coat rocket engines to avoid high temperature oxidation, increasing the durability of the rocket.⁶⁶ Another broad use of 5d compounds is as catalysts, as the large range of oxidation states (from I to VI) that they exhibit make them particularly useful in roles such as for the production of H₂.⁶⁷

Thermodynamic properties are key to understanding potential reactivity and interactions of 5d species, and capitalizing upon their current and potential uses. A number of compilations of thermodynamic data which date back to the 1950s have been reported for 5d species, including the NIST-JANAF,⁶⁸ Yungman,⁶⁹ Krasnov,⁷⁰ and Gurvich handbooks.⁷¹ However, there are significant limitations to this. For early main group species, a well-established body of experimental energetic data (i.e., enthalpies of formation, ionization potentials, electron affinities) with small uncertainties (< 1 kcal·mol⁻¹) is available, but for transition metal species, this is not the case. Of the data that is available for transition metal species, very little of it has such small uncertainties. For example, the gas phase enthalpies of formation at room temperature, $\Delta H_f^{298\text{ K}}$, for WOF₄ and WCl₂ have experimental uncertainties of 22 and 25 kcal·mol⁻¹, respectively.⁶⁸ Moreover, the reported enthalpies of formation of 5d species can differ substantially. For example, from Krasnov,⁷⁰ the enthalpy of

formation of HfI_4 is $-86.3 \pm 1.8 \text{ kcal}\cdot\text{mol}^{-1}$ whereas from Filippenko,⁷² it is $-113 \text{ kcal}\cdot\text{mol}^{-1}$. For some molecules, the rationale for the differences is clear, with new, improved experiments over time, but, unfortunately, this is not easily resolved for all species.

The most common experimental method used to determine the enthalpy of formation is via a calorimetry, where $\Delta H_f^{298 \text{ K}}$ is obtained by measuring the heat of combustion. This route has been widely used for small organic molecules where the composition of combustion products is easier to analyze than for inorganic compounds. Moreover, for such small organic molecules, the results tend to have very low uncertainties ($\pm 1 \text{ kcal}\cdot\text{mol}^{-1}$) for the experimental properties. However, for inorganic molecules, the calorimetric measurement can be difficult since the reaction products are often more complex than for small organic molecules. Furthermore, because the analysis of the product is more complicated, this can lead to high inaccuracies and low reproducibility for the enthalpy of formation. Other experimental routes have been utilized such as via vapor pressure at one or more temperatures. (Further details about the routes are provided in the appendix (see appendix A).) These vapor pressure-related approaches are reported in some of the best known compilations of thermodynamics data including the Gurvich⁷¹ and Yungman⁶⁹ handbooks and the NIST-JANAF tables.⁶⁸ For some of the species listed in these resources, the experimental enthalpies of formation were not obtained through experiment, per se, but instead, were extrapolated using experimental data available for other species.^{68,69,72} To illustrate, the NIST-JANAF tables indicate that the WCl_2 gas phase $\Delta H_f^{298 \text{ K}}$ was computed by using the enthalpy of formation of the WCl_2 crystal and the enthalpy of sublimation obtained via an extrapolation based upon data for a number of other $3d$ and $4d$ dichloride compounds.⁶⁸

To aid in gauging the utility of methodologies for $5d$ species, new experimental methods are emerging such as the resonant two-photon ionization process from the Morse group, resulting in dissociation energies with unprecedented accuracies, for transition metal and heavy element species ($\sim 30 \text{ cm}^{-1}$ experimental uncertainties were obtained).^{73,74} This approach provides an important new gauge for theoretical methods. The focus of these experiments has been on diatomic molecules, which provides theoretically accessible targets for quantitative assessment.

To determine $\Delta H_f^{298\text{ K}}$, a number of pathways can be considered. A widely used route is based upon the total atomization energy (TAE),⁷⁵ requiring theoretical methods that can describe both atomic and molecular energies well, typically requiring high level *ab initio* methods in combination with large basis sets, towards complete basis set (CBS) limit energy predictions. A second route is via the isodesmic family of approaches (i.e., isogyric, homodesmotic, hypohomodesmotic).⁶⁴

Isodesmic approaches are based on reactions in which the compounds involved in a reaction have similar structure. Wheeler describes many types of these reactions, including (1) the number of electrons paired is the same for the product and reactant (isogyric), (2) the number of similar bond types (C–C, C=C and C≡C) are the same on both sides of the reaction (isodesmic).⁷⁶ Such isodesmic approaches are useful as the predicted energies are less dependent on the method chosen than occurs for the TAE approach. Having similar structures on both sides of the reaction enables a greater cancellation of error arising from the chosen computational methodologies. An important key for isodesmic approaches to be useful is having sufficient high-quality enthalpies from experiment with small uncertainties. For heavier elements, unfortunately, as noted, there is limited experimental data that would enable the utilization of an isodesmic approach, and, thus, an atomization approach is typically necessary.

Transition metal and heavy element species are some of the most challenging complexes to study, with increasingly complex electronic manifolds, often requiring methodology such as CCSD(T) and a large basis set for thermochemical prediction, becoming computationally expensive, and, easily, prohibitively so. Moreover, due to the high atomic numbers for these species, relativistic effects such as scalar effects and spin-orbit coupling effects should also be addressed. With these challenges, theoretical strategies to enable reliable predictions, but at reduced computational cost, are of interest. While there are many approaches that can be considered, here the focus is upon *ab initio* composite approaches. In composite approaches, the goal is to achieve the accuracy in energetic predictions that can be achieved with a high-level *ab initio* method in combination with a large basis set, but with a combination of less costly, though less accurate calculations. Many different composite approaches have been developed including the Gaussian-n method,^{48,77} the

Peterson style CBS model,⁵⁰ the Weizman-n (W-n),^{51,78–80} the High-accuracy extrapolated *ab initio* thermochemistry (HEAT),^{53,81,82} the Feller-Peterson-Dixon procedure⁵⁴(FPD) and the correlation consistent composite approach (ccCA) which was developed in our group.^{55,57,61,83}

In prior work the 5*d* transition metal complexes have been studied theoretically.^{84–90} Minenkov and co-workers predicted the enthalpy of formation of a number of 5*d* and other transition metal complexes using an approach based on the FDP composite approach and DLPNO-CCSD(T) and an isogyric scheme.⁸⁵ For some of the hafnium and tungsten compounds, very large differences (up to ± 50 kcal·mol⁻¹ for HfCl₂) were obtained between the calculated enthalpies of formation and the experimental enthalpy. Schimmelpfennig *et. al.* investigated the gas phase structure and properties of different tungsten chloride compounds by employing the B3LYP functional with a 14-electron quasi relativistic effective core potential (ECP) for W and a seven-electron ECP for Cl.⁸⁷ The enthalpies of formation of WCl₄ and WCl₃ were determined using an isogyric reaction and the experimental data of WCl₅ and WCl₆ obtained from NIST-JANAF.⁶⁸ Finally, Fang and co-workers employed a composite scheme based on the FPD procedure to determine the enthalpies of formation of hafnium and tungsten complexes.⁹⁰ They predicted the enthalpies of formation using several different density functionals (B3LYP, M06, BP86, PW91, PBE and SVWN5) as well, considering both the Hartree-Fock and Brueckner orbitals. In these prior studies, only the hafnium and tungsten metals were investigated. Large errors (~ 9 , ~ 10 , ~ 50 and ~ 50 kcal·mol⁻¹ with respect to the experimental enthalpies of formation) were found for a number of the compounds (WO₃, HfF₄, HfCl₂ and HfI₂, respectively). The choice of Hartree-Fock or Bruckner orbitals made only a small variation (~ 1 kcal·mol⁻¹) in the enthalpies of formation.

Previously, the ccCA method has been successfully applied to 3*d* and 4*d* transition metal species to calculate their enthalpies of formation.^{55,56,61,83,91,92} Moreover, Jiang *et. al.* showed that for a set of ~ 200 3*d* transition metal species – a set of 3*d* species with the gas-phase enthalpies from experiment – ccCA-TM, the transition metal version of ccCA, led to an overall mean absolute deviation (MAD) of 4.34 kcal·mol⁻¹ as compared with experiment.⁹¹ While this error is on par with the average experimental uncertainty for these species of 3 kcal·mol⁻¹, referred to as "transition

metal chemical accuracy".⁵⁶ For a subset (~70) of the ~200 species, ccCA-TM achieved transition metal chemical accuracy of 3 kcal·mol⁻¹ relative to experiment. The overall performance of ccCA for these species is far better than that of density functional approaches, which are typically broadly used as a first choice of method, even for transition metal species, due to their much lower computational cost, relative to *ab initio* methods.^{61,83}

These prior efforts utilized an atomization approach. As noted, isodesmic approaches can reduce deviations from experiment, particularly for methods that are less reliable via atomization approaches. To illustrate, for B3LYP, Jorgensen *et. al.* demonstrated that the isogyric approach reduced the MAD by 15 kcal·mol⁻¹ when compared to the atomization MAD showing much better overall results when an isodesmic approach is selected.⁹³ However, in contrast, ccCA is not dependent upon thermochemical pathway. Both atomization and isodesmic approaches result in similar energies. ccCA has been used by Wilson *et. al.*⁹⁴ and Jorgensen *et. al.*⁹³ to study the enthalpy of formation of hydrocarbons and organosulfur compounds, respectively. In both studies, using an isodesmic approach reduced the MAD of ccCA by only ~0.5 to ~1 kcal·mol⁻¹ when compared to the atomization MAD, demonstrating the strength, reliability, consistency, and robustness of the method, regardless of thermochemical route. It is this stability that is capitalized upon in the present study.

With the successes of ccCA for earlier transition metal species, in the current study, the enthalpies of formation of 5d metal complexes are examined. As the thermochemical data available from experiment is limited, and, of this limited data, a number of species have very large experimental uncertainties, here, the ccCA strategies are first utilized for comparison with experiment, but, then are utilized to examine the prior experimental data, using theoretical data and thermochemical pathways. DFT methods are also considered in this study to gain insight about their utility for 5d species.

3.2 Methodology

In this study, several DFT and rp-ccCA⁶¹ approaches are used to determine the $\Delta H_f^{298\text{ K}}$ for a set of 5d transition metal species including HfF₄, HfCl, HfCl₂, HfCl₄, HfBr₂, HfI₂, TaCl₅, WCl,

WCl₂, WCl₄, WCl₅, HgCl, HgCl₂, HgBr, HgBr₂, HgI and HgI₂. The molecules include species with a span of experimental uncertainties and several species with experimental enthalpies that were extrapolated using enthalpies of other transition metal species. The mercury compounds (except HgBr) were selected for their well-established experimental data; WCl, WCl₂, WCl₄, WCl₅ and HgBr were selected for their large experimental uncertainties (from ~7 to ~25 kcal·mol⁻¹); HfCl, HfCl₂, HfBr₂ and HfI₂ were selected as they have experimental $\Delta H_f^{298\text{ K}}$ s that have been extrapolated from other transition metal enthalpies of formation.

The geometries of the 5d species were optimized with the meta-generalized gradient approximation (meta-GGA) density functional TPSS.⁹⁵ In prior work, Aebersold *et. al.*⁹⁶ showed that both TPSS and PBE0⁹⁷ provided geometries in good agreement with experiment for 54 lanthanide compounds. Moreover, a smaller mean signed deviation (MSD) and mean absolute deviation (MAD) were obtained with TPSS than with PBE0 for the enthalpies of formation, thus, TPSS was used in the present study for the geometry optimization. To determine the best ground state for each molecule, possible multiplicities were investigated and the one that resulted in the lowest energy and the smallest percentage spin contamination was used for the subsequent single point calculations.

The aug-cc-pVTZ and aug-cc-pV(T+d)Z³⁹ basis sets were used for F and Cl, respectively, incorporating the revised (and recommended) correlation consistent basis set for chlorine. For the Br, I, and all metals, small-core ECP10MDF,⁹⁸ ECP28MDF⁹⁸ and ECP60MDF⁴⁷ pseudopotentials were used with their respective pseudopotential (PP), aug-cc-pVTZ-PP, basis sets.

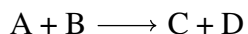
A number of density functionals were used for single point energy calculations on the optimized structures, spanning the multiple tiers (functional classes) of Jacobs ladder of complexity,⁹⁹ including (a) the local density approximation (LDA): SVWN;¹⁰⁰ (b) generalized gradient approximation (GGA): BP86,¹⁵ BLYP,¹⁰¹ PW91¹⁰² and PBE;¹⁰³ (c) meta-GGA (m-GGA): TPSS⁹⁵ and M06-L;¹⁰⁴ (d) hybrid-GGA (h-GGA): X3LYP,¹⁰⁵ B3P86,¹⁵ B97-1,¹⁰⁶ B3LYP,¹⁵ PBE0⁹⁷ and BHLYP;¹⁵ and (e) mh-GGA: TPSSH.¹⁰⁷ These functionals were selected due to their performance in prior studies of transition metal species, as well as to sample representation from different functional class characteristics.

The ccCA methodology used in this study is based on the rp-ccCA formulation by Laury and co-workers used for 4d species,⁶¹ but with a modification of the calculation for the reference energy. Instead of using both Hartree-Fock and second order Møller Plesset (MP2) theory to calculate the reference energy, only the MP2 energy is extrapolated to the complete basis set limit. This is because the extrapolation of the Hartree-Fock (HF) energy did not bring any benefit for these 5d species, with changes of less than 1 kcal·mol⁻¹ with the extrapolation. (It is noted that this could become important for larger species, and, in those cases, the HF extrapolation should be included). Thus, the rp-ccCA energy becomes:

$$E_{\text{rp-ccCA}} = \Delta E_{\text{ref}} + \Delta E_{\text{CC}} + \Delta E_{\text{CV}} + \Delta E_{\text{SO}} \quad (3.1)$$

where the only difference with the method presented in section 2.9, is that ΔE_{ref} correspond to a three point extrapolation scheme of the MP2 energy.⁶⁰

To calculate the enthalpy of formation, two approaches were used: the atomization energy approach and the isogyric reaction approach, which is one of the isodesmic schemes. The isogyric reaction corresponds to a reaction in which the number of paired electrons is the same on both sides of the reaction. If the enthalpy of formation of a molecule A is calculated using the following reaction,



the enthalpies of formation of B, C and D need to be known *a priori* (from experiment). In this study, the experimental data needed for this scheme were taken from either NIST-JANAF or from a study by Krasnov *et. al.*^{68,72} Furthermore, to consider the reliability of the experimental data used for the 5d compounds in this scheme, theoretical data obtained from literature were also used.

The DFT calculations were done with NWChem6.8¹⁰⁸ and the rp-ccCA calculations were done using both NWChem6.8 and Molpro2015.¹⁰⁹

3.3 Results

The optimized geometry obtained with TPSS is compared to structures from prior experiment and calculations. As shown in Table 3.1, the bond length of HfCl₄ obtained with TPSS is close

(0.004 Å) to the experimental result and slightly better than the optimized bond distance obtained by Fang and Dixon using CCSD(T)/aug-cc-pVTZ (0.017 Å).⁸⁹ The same observation is made for HfF₄, where the bond distance is closer (0.02 Å) to the experimental result than the results obtained by Thanthiriwatte at the CCSD(T)/aug-cc-pVTZ level, differing by 0.03 Å.⁸⁸

Table 3.1: Computed and experimental bond lengths (M-X) in Å between the metal (M) and the ligand (X).

Molecule	CCSD(T)	TPSS	Exp.
HfCl ₄	2.333 ⁸⁹	2.320	2.316(5) ¹¹⁰
HfF ₄	1.935 ⁸⁸	1.897	1.89 ⁷⁰
HgCl ₂	2.261 ¹¹¹	2.276	2.29 ± 0.02 ⁷⁰

Overall, the DFT (TPSS) bond-distances are in agreement with both experimental and other theoretical results, and due to the low computational cost of TPSS, the TPSS structure were used for the geometry optimization of the 5*d* compounds.

Enthalpies of formation using rp-ccCA

The enthalpies of formation calculated with rp-ccCA as well as the experimental enthalpies of formation are given in Table 3.2. Other theoretical enthalpies are also provided, where available.

Table 3.2: Experimental and theoretical $\Delta H_f^{298\text{ K}}$ in kcal·mol⁻¹ computed with rp-ccCA using a total atomization energy (TAE) approach.

Class ^d	Molecule	Multiplicity	$\Delta H_f^{298\text{ K}}$			
			Lit.	rp-ccCA	Exp.	Exp. Ref.
I	HgCl	2		22.21	18.75 ± 2.30	Ref ⁶⁸
	HgCl ₂	1		-34.35	-34.96 ± 1.50	Ref ⁶⁸
	HgBr ₂	1		-20.36	-20.50 ± 3.00	Ref ⁶⁸
	HgI	2		33.34	31.90 ± 1.00	Ref ⁶⁸
	HgI ₂	1		-0.82	-3.86 ± 0.5	Ref ⁶⁸
	HfCl ₄	1	-217.0 ^{88,a}	-210.18	-212.90 ± 0.30	Ref ¹¹²
	TaCl ₅	1		-172.57	-182.90 ± 2.00	Ref ⁶⁸
II	HfF ₄	1	-407.8 ^{88,a}	-451.88	-399.10	Ref ¹¹³
III	WCl	6		144.63	132.30 ± 10.00	Ref ⁶⁸
	WCl ₄	3	-70.268 ^{87,c}	-67.53	-80.3 ± 7.9	Ref ⁶⁸
	WCl ₅	2		-104.12	-98.60 ± 7.88	Ref ⁶⁸
	HgBr	2		27.33	24.49 ± 9	Ref ⁶⁸
IV	HfCl	2		73.48	7.00	Ref ⁶⁹
	HfCl ₂	1	-27.1 ± 3.3 ^{85,b}	-27.00	-76.00	Ref ⁷²
	HfBr ₂	1	4.2 ± 5.8 ^{85,b}	-0.98	-6.90	Ref ⁷²
	HfI ₂	1	40.5 ± 4.9 ^{85,b}	38.22	-8.90	Ref ⁷²
	WCl ₂	5		67.72	-3.00 ± 25	Ref ⁶⁸

^a CCSD(T) calculation following the FPD composite approach.

^b Homodesmotic approach and DLPNO-CCSD(T).

^c B3LYP/DZ+P.

^d Class I represents molecules with experimental uncertainties of < 3 kcal·mol⁻¹. Class II, no uncertainties are given. Class III, experimental uncertainties of > 3 kcal·mol⁻¹. Class IV, experimental data were determined via extrapolation.

Table 3.2 shows that the errors between the experimental and theoretical $\Delta H_f^{298\text{ K}}$ for the mercury compounds (Class I) are relatively small, with HgCl having the largest error (3.46 kcal·mol⁻¹). As the experimental uncertainties for the Hg compounds are small (with the exception of HgBr), these compounds provide a useful gauge for rp-ccCA. The mean unsigned error for the mercury species is 1.95 kcal·mol⁻¹, which is lower than the average experimental uncertainty of 2.71 kcal·mol⁻¹ showing that the rp-ccCA methodology is capable of describing mercury compounds. It is worth noting that the calculated ($\Delta H_f^{298\text{ K}}$) are all higher than their respective enthalpies of formation. However, no atomic spin-orbit term contribution was included at the molecular level beyond the spin-orbit contribution defined within the ECP.

For the tungsten compounds, the calculated enthalpies of formation are higher than the experimental values, with the exception for WCl_5 . The errors between experiment and theory vary substantially, ranging from $\sim 70 \text{ kcal}\cdot\text{mol}^{-1}$ for WCl_2 to only $-5.40 \text{ kcal}\cdot\text{mol}^{-1}$ for WCl_5 . For WCl_2 , the extrapolated experimental enthalpy of formation reported by NIST-JANAF⁶⁸ has an error of $25 \text{ kcal}\cdot\text{mol}^{-1}$ and was obtained from the enthalpy of formation of the crystal and the heat of sublimation, the later which was obtained by an extrapolation of the heats of sublimation for a series of $3d$ and $4d$ dichloride compounds. To the best of our knowledge, no direct experimental $\Delta H_f^{298 \text{ K}}$ for WCl_2 has been reported in the literature. The enthalpy of formation for WCl_4 ($-67.41 \text{ kcal}\cdot\text{mol}^{-1}$) using ccCA was found to be relatively close to the enthalpy proposed by Schimmelpfenning of $-70.268 \text{ kcal}\cdot\text{mol}^{-1}$,⁸⁷ using B3LYP/DZ+P, which differs substantially from experiment. For HfCl_4 , the enthalpy of formation obtained with rp-ccCA is $-210.18 \text{ kcal}\cdot\text{mol}^{-1}$ which is near the experimental value of $-212.90 \text{ kcal}\cdot\text{mol}^{-1}$; but further from $-217.0 \text{ kcal}\cdot\text{mol}^{-1}$ obtained by Dixon and co-workers using a method based on the FPD scheme.⁸⁸ For HfCl_2 , the difference between the rp-ccCA enthalpy and the experimental data is quite large ($49 \text{ kcal}\cdot\text{mol}^{-1}$). However, the rp-ccCA $\Delta H_f^{298 \text{ K}}$ is in better agreement with prior theoretical studies. The result from Minenkov and co-workers⁸⁵ obtained using a homodesmotic scheme and a DLPNO-CCSD(T) approach is very close to the result obtained in this study, with a difference of only $0.1 \text{ kcal}\cdot\text{mol}^{-1}$, while the result from Dixon obtained with a TAE approach and the FPD composite scheme has a difference of $7 \text{ kcal}\cdot\text{mol}^{-1}$ from experiment.⁸⁸ It should be noted that the reported experimental data is an estimate and is not taken directly from experiment.

For the enthalpy of formation of HfI_2 , The enthalpy reported by Minenkov⁸⁵ is $40.5 \text{ kcal}\cdot\text{mol}^{-1}$ and the result obtained from rp-ccCA is $38.22 \text{ kcal}\cdot\text{mol}^{-1}$, whereas the experimental enthalpy from Filippenko is $-8.90 \text{ kcal}\cdot\text{mol}^{-1}$.⁷² Just as for HfCl_2 , the experimental data for HfI_2 is an estimate. Finally, for HfBr_2 , the enthalpy obtained with rp-ccCA ($-0.98 \text{ kcal}\cdot\text{mol}^{-1}$) is slightly closer to the estimated experimental result ($-6.90 \text{ kcal}\cdot\text{mol}^{-1}$) than the enthalpy ($4.20 \text{ kcal}\cdot\text{mol}^{-1}$) from Minenkov *et. al.*

Overall, a very low error is obtained for the mercury compounds and HfCl_4 , while much larger

discrepancies occur for the other metals.

DFT enthalpies of formation

While clearly, there are substantial differences between theory and experiment for a number of the $5d$ species, there are also some – particularly those with smaller experimental error bars – for which much better agreement between theory (ccCA) and experiment is achieved. However, DFT is often a first choice due to efficiency for the transition metals and beyond, here, some insight about the utility of density functional approaches is also useful. As a total atomization energy thermochemical approach is often necessary for transition metal species, that is the approach utilized here. The impact of spin-orbit DFT calculations is also considered.

In Figure 3.1, a comparison between the mean unsigned error of the DFT and the SO-DFT enthalpies is presented, relative to experiment.

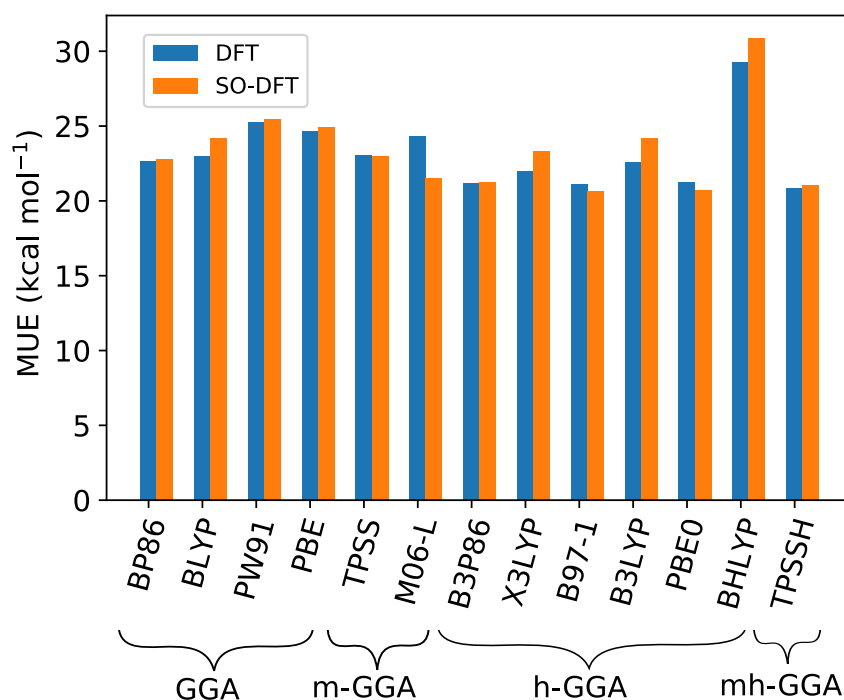


Figure 3.1: Comparison between the mean unsigned error (MUE) in kcal·mol⁻¹ for each density with and without spin-orbit correction.

The overall difference in the mean unsigned error (MUE) in the enthalpy between DFT and SO-DFT is very small, with the largest error being 3 kcal·mol⁻¹ for the M06-L functional. The MUE obtained with SVWN for both DFT and SO-DFT were over 40 kcal·mol⁻¹, at least 10

$\text{kcal}\cdot\text{mol}^{-1}$ larger than the MUE of the other functionals. Overall, the errors obtained with the hybrid-GGA functionals (B3P86 to PBE0) are lower than the errors obtained with the GGA and m-GGA functionals. For all of the hybrid-GGA functionals, the percentage of exact exchange from Hartree-Fock is between 20 to 25 and for the hybrid-GGA; for BHLYP, the percentage is 50. When higher percentages of exact exchange are used, the enthalpies calculated have the largest errors.

For the $5d$ species, the interaction between the sub-valence ($5s^2$ and $5p^6$) electrons and the valence electrons ($6s^2 5d^n$) play an important role in the energy description. Consequently, using a functional that includes a high percentage of exchange such as BHLYP will not lead to improved enthalpies of formation as compared with hybrid-GGA functionals.

In Figure 3.2.b for rp-ccCA and in Figure 3.2.a for each density functional, the MUE in the energies for each metal type are provided. The errors in the energies as compared to experiment vary significantly, depending upon on the metal.

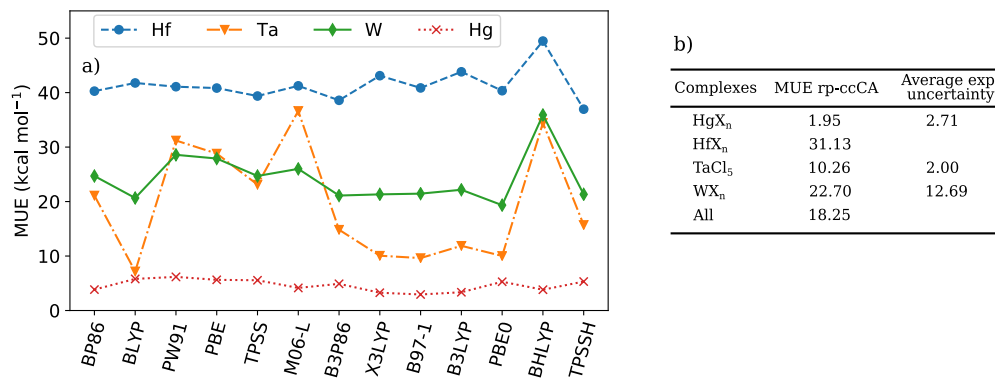


Figure 3.2: a) DFT MUE in $\text{kcal}\cdot\text{mol}^{-1}$ for each functional and metal center. b) rp-ccCA MUE and average experimental uncertainty in $\text{kcal}\cdot\text{mol}^{-1}$ for each metallic center.

The DFT comparison is provided per metal type for simplicity. (Note, there is just one Ta compound. Also, a table of results (Table A.1 and Table A.2) for each molecule and functional is provided in the appendix A.)

For the mercury complexes, the MUE ranges from 2 to 5 $\text{kcal}\cdot\text{mol}^{-1}$ for most functionals, whereas for hafnium, their MUE varies from 37 to 50 $\text{kcal}\cdot\text{mol}^{-1}$ for all functionals. The large MUEs from DFT for the hafnium compounds also occurred for rp-ccCA. As well, large errors were obtained for HfBr₂, HfCl₂ and HfI₂ (up to ~ 36 , ~ 60 and ~ 40 $\text{kcal}\cdot\text{mol}^{-1}$, respectively) – all are compounds

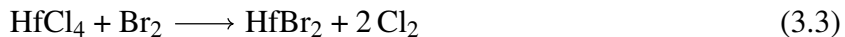
for which the experimental $\Delta H_f^{298\text{ K}}$ was extrapolated. For HfCl_4 , however, the DFT MUE is smaller ($\sim 12\text{ kcal}\cdot\text{mol}^{-1}$). The rp-ccCA MUE for the hafnium complexes is $31.13\text{ kcal}\cdot\text{mol}^{-1}$ which is better by $\sim 10\text{ kcal}\cdot\text{mol}^{-1}$ as compared to DFT MUE. For the tungsten compounds, their MUEs range between $20\text{ kcal}\cdot\text{mol}^{-1}$ with BLYP and PBE0 to $50\text{ kcal}\cdot\text{mol}^{-1}$ with SVWN while the rp-ccCA MUE is $\sim 23\text{ kcal}\cdot\text{mol}^{-1}$. Overall, the h-GGA functionals with $\sim 20\%$ exact exchange and BLYP result in the smallest error for the tungsten compounds, the mercury compounds, and for TaCl_5 (~ 20 , ~ 4 and $\sim 10\text{ kcal}\cdot\text{mol}^{-1}$, respectively). Furthermore, as shown in Figure 3.2.a, BHLYP results in the largest error for TaCl_5 , the tungsten compounds, and hafnium compounds (~ 35 , ~ 38 and $\sim 50\text{ kcal}\cdot\text{mol}^{-1}$, respectively).

Isogyric reaction

As shown for rp-ccCA and DFT, a total atomization approach can lead to large errors in the enthalpies, relative to reported experimental data. In fact, only for the mercury compounds did both DFT and rp-ccCA result in $\Delta H_f^{298\text{ K}}$ within reasonable agreement (error of ~ 2 to $\sim 4\text{ kcal}\cdot\text{mol}^{-1}$) of experiment. For the other metals, there was significant variance in the predictions, from agreement to significant disagreement between theory and experiment. As ccCA has been demonstrated to be robust, regardless of thermochemical pathway and a number of the experiments resulted from extrapolation, rather than direct measurement, here, isogyric reactions are examined to gain greater insight about the theoretical predictions and experimental extrapolations.

Both hafnium and mercury complexes were studied using the isogyric approach, mercury, as a well-behaved system, and hafnium, due to the significant differences between calculation and experiment for some of the hafnium species. The reactions involving hafnium were based on HfCl_4 since many prior theoretical studies are in agreement with the experimental enthalpy of formation of $-212.9 \pm 0.3\text{ kcal}\cdot\text{mol}^{-1}$. For mercury, HgI will be central to the thermochemical pathway, as it has the lowest experimental uncertainty among the mercury compounds studied. Moreover, the energy for each molecule involved in the reaction was calculated with rp-ccCA. In Table 3.2, a summary of the enthalpy obtained with rp-ccCA as well as the experimental and prior theoretical data for each compound is given. For the complexes based on hafnium, three chemical reactions were considered

in this study:



For each of the reactions, the number of paired electrons is the same on both sides of the equation. For example, in reaction 3.2, to determine the enthalpy of formation of HfCl_2 , the following expression is used:

$$\Delta H_f^{298 \text{ K}}(\text{HfCl}_2) = \Delta H_f^{298 \text{ K}}(\text{HfCl}_4) - E(\text{HfCl}_4) + E(\text{HfCl}_2) - \Delta H_f^{298 \text{ K}}(\text{Cl}_2) + E(\text{Cl}_2) \quad (3.5)$$

where E represents the energy calculated for each molecule using rp-ccCA and $\Delta H_f^{298 \text{ K}}$ represents the enthalpies of formation of HfCl_4 and Cl_2 taken from experiment.

In Figure 3.3, first, the error in the enthalpy of formation is shown for the total atomization energy approach for each of the hafnium species. Each of the isogyric reactions (Equations 3.2, 3.3, and 3.4) was utilized to predict the enthalpy of formation for HfCl_4 . As well, an isogyric approach was used, employing the respective equations (Equation 3.2 for HfCl_2 , Equation 3.3 for HfBr_2 , and Equation 3.4 for HfI_2) for the remaining enthalpies shown in the figure. For the enthalpies determined for Figure 3.3, the isogyric approaches only utilized enthalpies from experiment

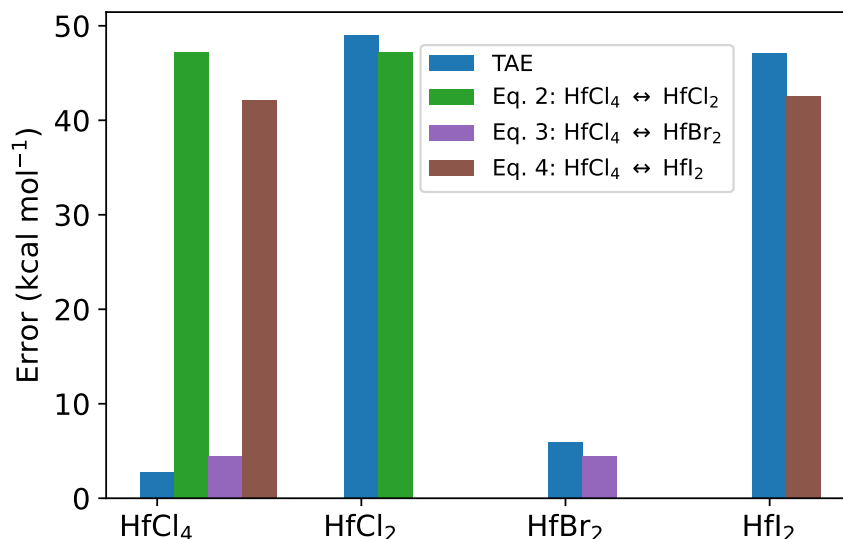


Figure 3.3: Comparison between enthalpies of formation in $\text{kcal}\cdot\text{mol}^{-1}$ obtained with the atomization approach (blue) and the isogyric reaction (3.2–3.4) using experimental data and rp-ccCA to calculate the individual energies.

For HfCl_4 , there are significant differences in the predictions made via the atomization energy approach and the Equation 3.2 and 3.4 isogyric approaches, whereas Equation 3.3 results in somewhat similar enthalpies as for the atomization energy approach. The $\Delta H_f^{298\text{ K}}$ of HfCl_4 obtained with reaction 3.2 and 3.4 lead to errors that are ~ 47 and $\sim 42\text{ kcal}\cdot\text{mol}^{-1}$, respectively, from experiment, while for HfBr_2 (reaction 3.3), the enthalpy of formation obtained for HfCl_4 is only $\sim 5\text{ kcal}\cdot\text{mol}^{-1}$ from experiment. As noted earlier, rp-ccCA typically results in similar energies, regardless of thermochemical pathway, so these significant differences suggest a possible issue with the reported experimental data (which were estimates) for some of the molecules. In considering the isogyric and atomization energy approaches for each of the other three molecules, the isogyric and atomization energies are similar. However, the errors from experiment are very large for HfCl_2 and HfI_2 .

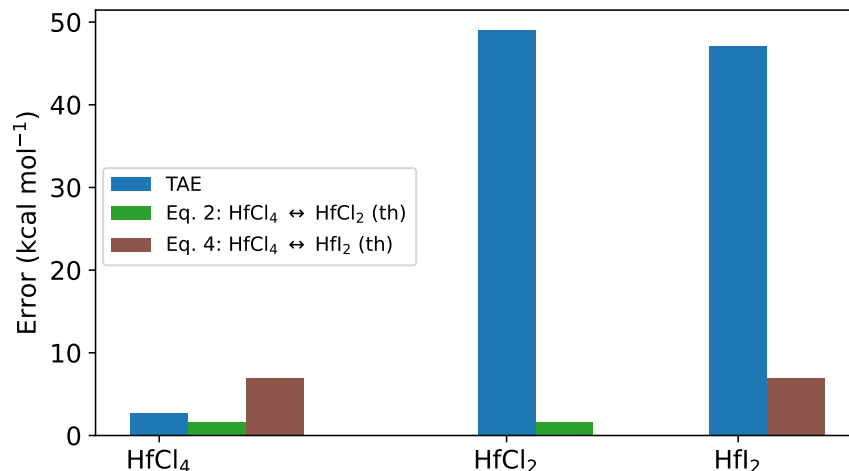


Figure 3.4: Comparison between enthalpies of formation in $\text{kcal}\cdot\text{mol}^{-1}$ obtained with the atomization approach (blue) and the isogyric reaction (3.2 and 3.4) using theoretical data and rp-ccCA to calculate the individual energies.

To gain additional perspective and further consider the validity of the experimental enthalpies of formation of HfCl_2 and HfI_2 , the isogyric reactions now utilize theoretical data (while utilizing the same total atomization energy approach as described for Figure 3.3, and included in the figure to provide comparison.) More specifically, for Figure 3.4, the isogyric enthalpies for HfCl_4 were based on theoretical enthalpies for HfCl_2 (Equation 3.2) and HfI_2 (Equation 3.4), from Minenkov and co-worker (results noted "a" in the Table A.3 in the appendix A). And, for the other two molecules, the experimental value is utilized for HfCl_4 , and the isogyric enthalpies of HfCl_2 and HfI_2 are determined, and compared to results from theory. HfBr_2 has not been reevaluated, as the enthalpies are similar from the isogyric and atomization approaches. (Note that the atomization energy-based enthalpies are compared with experiment.) The error has been driven down substantially, as shown for HfCl_4 , HfCl_2 , and HfI_2 using the isogyric Equations 3.2 and 3.4. These results emphasize that in order to obtain the correct enthalpy of formation for HfCl_4 with Reactions 3.2 and 3.4, the available experimental data for HfCl_2 and HfI_2 cannot be used in the isogyric approach, and that the experiments should be revisited to provide measurements, rather than extrapolations. A table (Table A.3) describing the detailed analysis is provided in the appendix A.

For the mercury compounds, two reactions were studied:



As theoretical enthalpies of formation for HgCl, HgBr and HgI were not readily available in the literature, only the experimental enthalpies of formation was used to determine the enthalpies of formation.

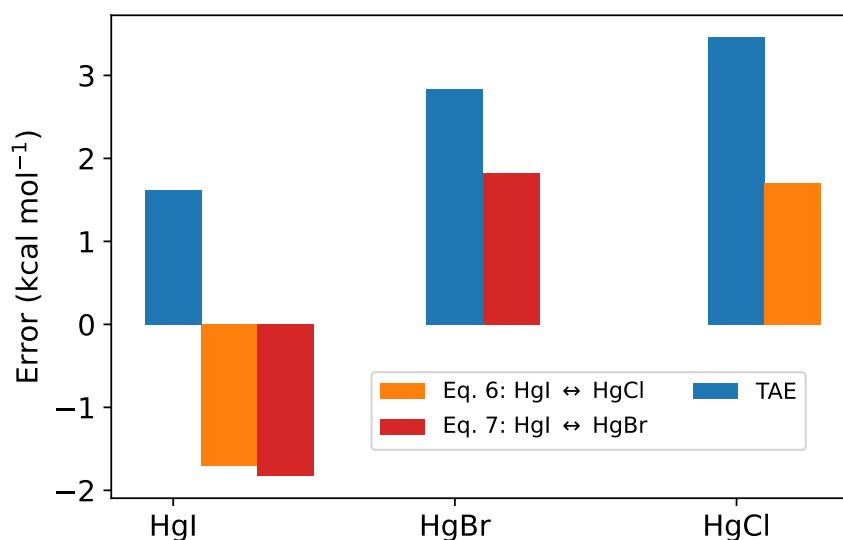


Figure 3.5: Comparison between enthalpies of formation in kcal·mol⁻¹ with the atomization approach (blue) and the isogyric approach using experimental data and rp-ccCA to calculate the individual energies.

As depicted in Figure 3.5, the errors in the enthalpies for the mercury compounds are much smaller than for the hafnium compounds. When the isogyric reaction is used, the errors obtained are slightly lower than those with the TAE approach. Indeed, the enthalpy obtained for HgBr with the experimental enthalpy of HgI gives a smaller error than the rp-ccCA/TAE results. This smaller error from the isogyric approach also occurs for HgCl. These results demonstrate the viability of an isogyric approach with well-established experimental data. Moreover, it also further shows the utility of the rp-ccCA method in the determination of the enthalpy of formation for these compounds.

3.4 Conclusion

In this study, rp-ccCA was utilized to calculate enthalpies of formation for a set of 17 *5d* complexes. Density functional theory was also used in order to further gain a greater understanding about its performance for *5d* compounds. Thermochemical pathways can be useful in gauging the viability of both theoretical and experimental methods. Here, the viability of rp-ccCA for *5d* compounds was demonstrated. An analysis of rp-ccCA, previous theoretical data, and experiment, suggests the need to revisit a number of *5d* species with new experimental measurements.

CHAPTER 4

MULTIREFERENCE CALCULATIONS ON THE GROUND AND EXCITED STATES AND DISSOCIATION ENERGY OF LuF

Reproduced from N. M. S. Almeida, T. R. L. Melin and A. K. Wilson, J. Chem. Phys. **154**, 244304 (2021),¹¹⁴ with the permission of AIP publishing. The dissociation energy calculations and most of the data analysis for multireference calculations were done by Timoth  Melin. Dr. Nuno Almeida did most of the multireference calculations and plotted the potential energy curves. Timoth  Melin and Dr. Nuno Almeida contributed equally to the writing of the paper.

4.1 Introduction

The accurate description of ground and excited state properties of lanthanides provides a route towards understanding their fundamental chemical reactivity. The high density of states and partially filled $4f$ and $5d$ orbitals are hurdles that need to be properly addressed in order to achieve such predictions. The use of multireference methods in lanthanide electronic structure calculations is of paramount importance and allows for an accurate description of static and dynamic correlation. As well, an appropriate choice of methods to account for correlation and spin-orbit effects is necessary for both the ground and excited states.

Lutetium, the last element in the lanthanide series is also generally regarded as the first element of the sixth period transition metals, due to its full $4f$ and partially filled $5d$ orbitals. Recently, interest in lutetium has grown, with one of its main applications in the radiopharmaceutical industry, more specifically with the use of ^{177}Lu as a radionuclide.¹¹⁵ Small molecules such as peptides and steroids have been radiolabeled with ^{177}Lu in the treatment of a number of diseases. For example, ^{177}Lu -labeled DOTA-Tyr³-octreotate, which is a somatostatin analogue peptide is currently being used to treat neuroendocrine tumors.¹¹⁵ Lutetium also has been linked to astrophysics. It has been discovered in the composition of the metal-poor stars CS 31062-050 and CS 22892-052, and in the enriched star BD +17 3248.^{116–119} The Lu^+ spectra has been investigated by Hartog and co-workers revealing the presence of an excited state at 28503.16 cm^{-1} , which corresponds to a $6s\ 6p,\ ^3\text{P}_1$ configuration.¹²⁰ Lanthanide species, in general, are also being used and considered in a broad range

of applications such as in electrodes and optical telecommunications (i.e., NaLuF₄). With such a wide range of applications, it is important to better understand lutetium at a fundamental level, and the methodologies needed to describe its complex electron manifold.

In considering the ground and excited state properties of LuF, the available experimental data are from the 1960s, 1970s and 1980s. In 1968, Zmbov extrapolated the dissociation energy of LuF from other lanthanide monofluorides by means of mass spectroscopy, and obtained 136 ± 12 kcal·mol⁻¹.¹²¹ The authors estimated the dissociation energy of lutetium fluoride using both the heats of sublimation and the enthalpies of other lanthanide fluorides. Their estimation came from fluorine-exchange reactions of Sm, Eu, Gd, Dy and Er. Kaledin and *et al.* predicted the dissociation energy of LuF to be 124 kcal·mol⁻¹. The authors used ligand field theory and extrapolated the dissociation energy, utilizing a fitting model and experimentally determined ionization potentials for other lanthanide fluorides.¹²² Since the 1970s, several experimental studies have targeted the vibrational and rotational spectrum of lutetium fluoride.^{123–126} DIncan *et al.* and Effantin *et al.* reported dissociation energies for LuF (105 kcal·mol⁻¹) and assigned the lowest lying electronic excited states for LuF. The symmetry and spin were labelled either ¹Σ or ¹Π for all the excited states.^{123,124,126} These results were later compiled by Huber and Herberg in an extensive review of molecular spectra.¹²⁵ In the 1980s Rajamanickam and Narasimhamurthy and Reddy *et al.* obtained experimental dissociation energies of 96.0 ± 2.4 and 79 kcal·mol⁻¹, respectively.^{127,128} These authors used the experimental spectroscopic constants of the ground state (ω_e , $\omega_e\chi_e$, etc, obtained from the work of Effantin *et al.*¹²⁴), calculated the vibrational potential energy curve (PEC), fitted it with different empirical formulas, and calculated the dissociation energy.

Theoretical studies are useful in describing the spectroscopic properties of lanthanides. There are a number of recent studies on lanthanide monohalides (LnX, X= F, Cl, Br, I).^{129–136} In the 1990s, a number of theoretical studies focused on the spectroscopic properties of lanthanides and actinides. Wang *et al.* and Küchle *et al.* studied diatomics, lanthanide, and lanthanide and actinide contractions and were the first to use density functional theory (DFT) along with coupled cluster (CC) methods to calculate ground state properties and bond lengths for some of these molecules.^{137,138} Cooke

et al. investigated the rotation spectra of LuF and used DFT to compare with their ground state experimental values. Their theoretical prediction of the dissociation energy of $96.6 \text{ kcal}\cdot\text{mol}^{-1}$ was based upon a statistical average of orbital potentials.¹³⁹ Density functional theory with scalar-relativistic ZORA and Douglas-Kroll-Hess approaches have been used by Hong *et al.* to calculate the dissociation energy of LuF. The authors obtained values in the range of $167\text{-}176 \text{ kcal}\cdot\text{mol}^{-1}$.¹⁴⁰ In more recent work (2016), Grimm *et al.* determined for the Ln54 set, a set of 54 enthalpies of formation and bond dissociation energies of small lanthanides, using 22 different DFT functionals and employing the Douglas-Kroll-Hess Hamiltonian in combination with a triple- ζ level basis set [Sapporo-DouglasKrollHess third order Hamiltonian (DKH3)-TZP-2012 for Ln and cc-pVTZ-DK or cc-pV(T+d)Z-DK for the ligands], resulting in average overall energy errors for the set on the order of 1 eV, even with the most popular and well-utilized functionals for the lanthanides.¹⁴¹ Aebbersold *et al.* reexamined the energies of the Ln54 set using the same functionals employed by Grimm and co-workers, considering the several impacts including the introduction of effective core potential (ECP) and DKH3 approaches.^{96,141} In terms of *ab initio* studies, the equation of motion completely renormalized coupled-cluster single, double, and perturbative triple excitations [CCSD(T)] [EOM-CR-CCSD(T)] was used in a study of NdF and LuF.¹⁴² The authors reported that the use of a full valence shell rather than the traditional frozen core approximation can result in a dramatic change in the dissociation energy of LuF (a change of $\sim 35 \text{ kcal}\cdot\text{mol}^{-1}$). *Ab initio* composite methods have also been employed in the prediction of ground state properties of lanthanides. Solomonik and Smirnov calculated the bond dissociation of LuF as $169.7 \text{ kcal}\cdot\text{mol}^{-1}$ and Qing computed the same as $172.4 \text{ kcal}\cdot\text{mol}^{-1}$,^{143,144} which are near to our recent prediction of $170.2 \text{ kcal}\cdot\text{mol}^{-1}$ in a large scale study of lanthanides.⁶² In considering the prior experimental and theoretical studies, as overviewed, there are substantial differences in the predictions. It is important to note that the dissociation energies reported from experiments are not direct measurements, but are instead, based on empirical models.^{121–125}

In terms of excited states, a complete understanding of the potential energy surface of LuF and its bonding patterns, allows for the probing of possible chemical reactivity routes using excited

state dissociation channels. Towards this goal, in 2009, Hamed *et al.*¹⁴⁵ used CASSCF (complete active space self-consistent field) and MRCI (multireference configuration interaction), for the first low-lying excited states of LuF, using a pseudopotential for lutetium of 60 electrons. The authors determined 26 electronic states including the spectroscopic constants and bond lengths for each state; however, these calculations did not account for spin-orbit effects. The authors assigned the first and second excited states as $^3\Pi$ and $^3\Delta$, instead of the $^1\Sigma$ and $^1\Pi$ states, respectively, previously assigned in the literature.^{123–125} In 2019, Assaf *et al.* used multireference methods (CASSCF and MRCI+Q) to calculate spectroscopic constants and bond lengths for ground and excited states.¹⁴⁶ The authors considered a 28 electron pseudopotential (ECP28MWB), which allowed for a more accurate treatment of electron correlation. In addition, sub-valence electrons (4f) were also correlated, though not included in the active space. The latter step enables the prediction of bond lengths within 0.1 Å of experiment. The active space utilized in this study did not include the bonding orbitals of fluorine, which are important in the construction of the full potential energy curves. However, spin-orbit effects were considered, and spectroscopic constants were calculated for the low-lying excited states using the Breit-Pauli Hamiltonian.

Although there have been a number of studies on lutetium fluoride, detailed insight about its dissociation channels and binding patterns have not yet been provided. For this work, 132 states were investigated using multireference methods and double-, triple-, and quadruple- ζ level basis sets. The results herein provide important insight about the higher energy channels that play a role on the excited state surface of LuF. MRCI calculations were performed to recover dynamic correlation of the system beyond what CASSCF can obtain. Valence, sub-valence and inner core levels of correlation were probed, detailing their effects on the energetics of the ground and excited states. The second part of this work focuses on the dissociation energy (D_0) using a range of DFT functionals, and also *ab initio* methods, including coupled-cluster and CASSCF. Complete basis set (CBS) extrapolation was also considered for the *ab initio* methods.

4.2 Computational details

Multireference calculations were performed using MOLPRO 2018.¹⁴⁷ As MOLPRO does not use full linear molecule symmetries, the C_{2v} point group symmetry was utilized and the molecular orbitals were optimized using CASSCF. For this step, the active space used was composed of eight electrons and fifteen orbitals (8,15). The 15 orbitals correspond to 6 a_1 [$5d_{z^2}$, $5d_{x^2-y^2}$, $6s$, $6p_z$ (Lu), $2p_z$, $3p_z$ (F)], 4 b_1 [$5d_{xz}$, $6p_x$ (Lu), $2p_x$, $3p_x$ (F)], 4 b_2 [$5d_{yz}$, $6p_x$ (Lu), $2p_x$, $3p_x$ (F)], and 1 a_2 [$5d_{xy}$ (Lu)], which correspond to the $6s$ and $5d$ of lutetium, and to $2p$, and $3p$ of fluorine. The inclusion of the additional $3p$ orbitals of fluorine was deemed necessary to obtain smooth potential energy curves (PECs).

MRCI and MRCI+Q were employed to calculate spectroscopic constants.^{148–151} Harmonic vibrational frequencies, anharmonicities, and $\Delta G_{1/2}$ values were calculated solving the rovibrational Schrödinger equation numerically using the Dunham approach.¹⁵² Due to the computational cost, the $2p$ and $3p$ orbitals of fluorine were not included in the active space, and, thus were not optimized at the CASSCF level, within the MRCI calculations. The active space for MRCI consists of the following orbitals: 4 a_1 [$5d_{z^2}$, $5d_{x^2-y^2}$, $6s$, $6p_z$ (Lu)], 2 b_1 [$5d_{xz}$, $6p_x$ (Lu)], 2 b_2 [$5d_{yz}$, $6p_x$ (Lu)], and 1 a_2 [$5d_{xy}$ (Lu)]. However, the $2p$ orbitals of fluorine were included in the MRCI calculations as core (per MOLPRO 2018), by allowing the electrons to be promoted to the active and virtual spaces, through single and double excitations. Considering the CI vectors, for the equilibrium bond region there are not significant contributions that correspond to the promotion of electrons from the $2p$ orbitals of fluorine. In addition, for the MRCI calculations, sub-valence correlation effects were also described by including the $4f^{14}$ orbitals of Lu by also allowing single and double excitations to the active and virtual spaces. Since a pseudopotential was considered for the metal (see next paragraph), the remaining 52 electrons (9 from the fluorine and 43 of lutetium) were also correlated for MRCI calculations. The Davidson correction, or MRCI+Q as implemented within MOLPRO was used to account for size extensivity issues.^{148–151} To account for spinorbit coupling, the Breit-Pauli Hamiltonian was diagonalized in the basis of the MRCI wavefunction. For this step, two levels of correlation were considered for inclusion in the core: $4f^{14}$ (Lu) and $2p^5$ (F) orbitals, and $4d^{10} 5s^2$

$5p^6 4f^{14}$ (Lu) and $2s^2 2p^5$ of (F) orbitals. The latter describe the effects of inner-shell correlation.

For CASSCF calculations, a segmented contracted basis set along with a pseudopotential (ECP28MWB) developed by Cao and Dolg were employed (triple- ζ level).^{153,154} For fluorine, the aug-cc-pVTZ basis set was utilized.¹⁵⁵ For MRCI and spin-orbit calculations, the def2-QZVPP basis set was employed for lutetium with a pseudopotential (ECP28MWB), while fluorine was described with aug-cc-pVQZ.^{154–157}

For the second part of this work, the geometry optimization step was carried out with CCSD(T) in combination with a contracted basis set by Cao and Dolg, which was used for lutetium, and the aug-cc-pVTZ basis set for fluorine.^{153–155} The frequency was also obtained at the same level to ensure a minimum at the potential energy surface. The geometry was then used to evaluate dissociation energies at different levels of theory and the energy was corrected for the zero-point vibrational energy (ZPE). CCSD(T) and the completely renormalized (CR-CCSD(T)) approach with DKH3 in combination with Sapporo double-, triple- and quadruple- ζ basis set for lutetium and fluorine have been utilized.¹⁵⁸ The effect of a four-component Hamiltonian on the dissociation energy was also probed with CCSD(T), using a Dirac-Coulomb Hamiltonian. In addition, the Perdew-Burke-Ernzerhof (PBE),¹⁰³ the Becke, 3-parameter, Lee -Yang -Parr (B3LYP),^{15,16} the Minnesota 2006 local functional (M06-L)¹⁰⁴ and the Tao, Perdew, Staroverov, Scuseria (TPSS)⁹⁵ functionals were utilized to predict dissociation energies, employing a DKH3 Hamiltonian. These functionals were chosen as they are either widely utilized or were among the better functionals for the prediction of enthalpy of formation and dissociation energies for lanthanide complexes.^{96,141}

Moreover, these functionals will provide some level of comparison between the generalized gradient approximation (GGA): PBE; meta-GGA: TPSS, M06-L; and hybrid-GGA: B3LYP on the prediction of the dissociation energy. The double-, triple-, and quadruple- ζ level Sapporo basis sets for lutetium and fluorine were used (noted Sap- nz) and the Dyal augmented double-, triple- and quadruple- ζ (noted Dyal- nz) for the Dirac-Coulomb Hamiltonian where $n = D, T, Q$.¹⁵⁹

The dissociation energy was calculated using the methods described above and at each level of basis set as well. Extrapolations of the total energies to the complete basis set limit, using a mixed

exponential/gaussian three points scheme developed by Peterson,⁶⁰

$$E_n = E_{CBS} + Be^{-(n-1)} + Ce^{-(n-1)^2} \quad (4.1)$$

where B and C are constants determined in the scheme, and n is the basis set level ($n = D, T, Q$), E_n represents the energy for each basis set level, and E_{CBS} represents the energy at the CBS limit. Unfortunately, it was not possible to obtain values at a quadruple- ζ basis set for CCSD(T) and MP2 with the Dirac Coulomb Hamiltonian due to the very high computational cost. Thus, the complete basis set limit using the following two-point extrapolation (Dyall.dz and Dyall.tz) scheme by Martin¹⁶⁰ was used:

$$E = E_{CBS} + \frac{B}{(n + 0.5)^4} \quad (4.2)$$

This scheme has been shown to provide reliable extrapolated energies for molecules containing lighter elements when compared to experiment.^{92,161} The final dissociation energy is calculated by adding the zero-point vibrational energy to the final energy. The 95% confidence limit has been investigated and results (Table B.1) are given in the appendix B. In addition to evaluating the 95% confidence intervals, the error from basis set superposition (BSSE) was calculated utilizing Boys and Bernardis counterpoise correction approach (section 4.3).¹⁶²

Due to the large number of electrons, it is important to consider different frozen-core spaces, i.e., the number of electrons explicitly correlated. Thus, two frozen-core spaces have been considered: FC-val and FC-subval. FC-val corresponds to a space where only the valence electrons (6s and 5d of Lu and 2s and 2p of F) are treated and the rest is frozen. The FC-subval describes the space where the valence and sub-valence electrons are explicitly treated (5s, 5p of Lu). All calculations using the DKH3 Hamiltonian were performed with NwChem 6.1,¹⁰⁸ while the Dirac-Coulomb calculation were done using DIRAC18.¹⁶³

4.3 Results and discussion

Electronic structure calculations (CASSCF, MRCI, MRCI+Q)

The PECs calculated at the CASSCF level are displayed in Figure 4.1 and 4.2. The former portrays the $\text{Lu } (^2\text{D}; 5d^1 6s^2) + \text{F}(^2\text{P})$ and $\text{Lu } (^2\text{P}; 6s^2 5p^1) + \text{F}(^2\text{P})$ channels, and the latter shows the two upper binding $\text{Lu } (^4\text{F}; 5d^1 6s^2 5p^1) + \text{F}(^2\text{P})$ channels. In Fig. 4.3, MRCI+Q curves are provided with selected states spanning the equilibrium bond region. The zero of the energy scale in Figure 4.1-4.3 is taken as the energy of the lowest energy asymptote $\text{Lu } (^2\text{D}) + \text{F}(^2\text{P})$. In Table 4.1, detailed spectroscopic information of the ground and 22 excited states is shown, which includes, spectroscopic constants, harmonic vibrational frequencies, $\omega_e\chi_e$, $\Delta G_{1/2}$ and T_e (excitation energies).

Table 4.1: Computational method, bond length R_{eq} (Å), harmonic vibrational frequencies ω_e (cm^{-1}), anharmonicity $\omega_e \chi_e$ (cm^{-1}), $\Delta G_{1/2}$ (cm^{-1}) values, and excitation energy T_e (cm^{-1}) for the lowest electronic excited states of ^{175}Lu ^{19}F . MRCI, MRCI+Q and CCSD(T) calculations were performed using an ECP28MWB/Def2-QZPP for Lu and aug-cc-pVQZ for F and CASSCF with ECP28MWB/ANO-TZ for Lu and aug-cc-pVTZ for F.

States	Methodology	R_{eq}	ω_e	$\omega_e \chi_e$	$\Delta G_{1/2}$	T_e
$X^1\Sigma^+$	Exp. ¹²⁵	1.9171	611.79	2.54	-	0
	Exp. ¹²⁶	1.9165	611.79	2.54		0
	CASSCF	-	-	-	-	0
	MRCI	1.916	613.9	2.67	608.6	0
	MRCI+Q	1.914	611.7	2.82	606.1	0
	CCSD(T)	1.917	610.4	2.51	605.4	0
	CCSD(T)	1.917	610.8	-	-	0
	MRCI+Q ¹⁴⁶	1.913	618.9	2.5	-	0
	MRCI+Q ¹⁴⁵	1.922	606.6	3.3	-	0
$1^3\Delta$	Exp. ¹²⁵	1.9319	587.95	2.58	-	16165
	Exp. ¹²⁶	1.9313	587.95	2.58		16153
	CASSCF	-	-	-	-	18000
	MRCI	1.947	573.4	2.54	568.3	14917
	MRCI+Q	1.945	570.5	2.45	565.6	14676
	MRCI+Q ¹⁴⁶	1.947	576	2.7	-	14927
	MRCI+Q ¹⁴⁵	1.952	596.2	3	-	17904
$1^3\Pi$	Exp. ¹²⁵	1.9361	576.08	2.5	-	16800
	Exp. ¹²⁶	1.933	581.3	2.6		16785
	CASSCF	-	-	-	-	17155
	MRCI	1.928	570.8	3.88	563	15630
	MRCI+Q	1.93	570	3.75	562.5	15805
	CCSD(T) def2	1.943	574.8	2.5	569.6	18528
	MRCI+Q ¹⁴⁶	1.933	579.2	2.7	-	15959
	MRCI+Q ¹⁴⁵	1.923	567.1	2.6	-	16165
$1^3\Sigma^+$	Exp. ^{125,126}		605.5	2.5	-	18894
	CASSCF	-	-	-	-	19900
	MRCI	1.957	600.4	2.47	595.4	17947
	MRCI+Q	1.958	590.5	2.53	585.4	18181
	MRCI+Q ¹⁴⁶	1.961	559.6	2.5	-	18856
	MRCI+Q ¹⁴⁵	1.953	567.1	2.6	-	19131

Table 4.1: Table (continued)

States	Methodology	R_{eq}	ω_e	$\omega_e \chi_e$	$\Delta G_{1/2}$	T_e
$1^1\Delta$	Exp. ¹²⁵	1.948	569.7	2.5		20048
	Exp. ¹²⁶	-	-	2.6		20027
	CASSCF	-	-	-	-	21612
	MRCI	1.954	567	2.14	562.7	19392
	MRCI+Q	1.953	564.4	2.11	560.2	19060
	MRCI+Q ¹⁴⁶	1.955	567.7	2.8	-	19471
	MRCI+Q ¹⁴⁵	1.956	555	2.5	-	21634
$1^1\Pi$	Exp. ¹²⁵	1.9584	543.42	2.28		24474
	Exp. ¹²⁶	1.9584	543.42	2.28		24440
	CASSCF	-	-	-	-	27049
	MRCI	1.966	554	2.42	549.1	23371
	MRCI+Q	1.969	546.7	2.53	541.7	23065
	MRCI+Q ¹⁴⁶	1.972	525.3	2.2	-	23708
	MRCI+Q ¹⁴⁵	1.945	544.7	2.6	-	25538
$2^1\Sigma^+$	Exp. ¹²⁵	1.952	555.59	2.6		25832
	Exp. ¹²⁶	1.9514	560.8	2.6		25806
	CASSCF	-	-	-	-	29240
	MRCI	1.959	548.5	4.41	539.7	25628
	MRCI+Q	1.957	543.2	3.82	535.5	25292
	MRCI+Q ¹⁴⁶	1.959	553	2.5		25932
	MRCI+Q ¹⁴⁵	1.947	563.8	2.8		26524
$2^3\Pi$	CASSCF	-	-	-	-	34583
	MRCI	1.983	570.9	2.9	565.1	29091
	MRCI+Q	1.978	559.9	1.67	556.5	28870
	MRCI+Q ¹⁴⁶	1.981	577.7	2.3		29354
	MRCI+Q ¹⁴⁵	1.995	525.7	3.4		30681
$2^1\Pi$	Exp. ¹²⁵	1.951	599.1	2.6	-	33226
	CASSCF	-	-	-	-	38511
	MRCI	1.948	593.9	3.09	587.7	32809
	MRCI+Q	1.944	606.4	3.12	600.1	32517
	MRCI+Q ¹⁴⁶	1.951	614.7	2.9		32968
	MRCI+Q ¹⁴⁵	1.961	579.3	2.5		33378

Table 4.1: Table (continued)

States	Methodology	R_{eq}	ω_e	$\omega_e \chi_e$	$\Delta G_{1/2}$	T_e
$1^3\Phi$	CASSCF	-	-	-	-	38846
	MRCI	1.944	565.2	-0.33	565.9	33566
	MRCI+Q	1.944	571.2	0.28	570.7	33499
	MRCI+Q ¹⁴⁶	1.944	570.5	2.7		34248
	MRCI+Q ¹⁴⁵	1.942	570.8	3.2		36401
$3^3\Pi$	CASSCF	-	-	-	-	42188
	MRCI	1.96	554.9	2.54	549.8	36422
	MRCI+Q	1.959	543.7	2.63	538.4	36123
	MRCI+Q ¹⁴⁶	1.956	545	2.8	-	36896
	MRCI+Q ¹⁴⁵	1.957	552.4	3.2	-	39048
$2^3\Delta$	CASSCF	-	-	-	-	42211
	MRCI	1.974	573.3	5.5	562.3	36674
	MRCI+Q	1.974	592.2	7.18	577.8	36323
	MRCI+Q ¹⁴⁶	1.976	540.8	3	-	37162
	MRCI+Q ¹⁴⁵	1.969	541.8	2.3	-	39569
$1^3\Sigma$	CASSCF	-	-	-	-	41126
	MRCI	1.974	534.3	0.34	533.6	36683
	MRCI+Q	1.974	522.2	0.24	521.8	36338
	MRCI+Q ¹⁴⁶	1.973	544	2.6	-	37338
	MRCI+Q ¹⁴⁵	1.949	551.3	3.6	-	39216
$2^1\Delta$	CASSCF	-	-	-	-	46419
	MRCI	1.955	567	1.85	563.3	40151
	MRCI+Q	1.956	557.6	1.47	554.6	39524
	MRCI+Q ¹⁴⁶	1.956	558.5	2.6		40954
	MRCI+Q ¹⁴⁵	1.946	566.6	3.3		45661
$3^1\Sigma^+$	CASSCF	-	-	-	-	-
	MRCI	-	-	-	-	-
	MRCI+Q	-	-	-	-	-
	MRCI+Q ¹⁴⁶	1.942	550.1	3	-	42847
	MRCI+Q ¹⁴⁵	1.917	588.9	2.8	-	42763
$1^1\Sigma^-$	CASSCF	-	-	-	-	46100
	MRCI	1.953	565.3	2.4	560.5	43049
	MRCI+Q	1.959	557.9	2.33	553.2	41310

Table 4.1: Table (continued)

States	Methodology	R_{eq}	ω_e	$\omega_e \chi_e$	$\Delta G_{1/2}$	T_e
$1^1\Phi$	CASSCF	-	-	-	-	51158
	MRCI	1.952	566.9	2.41	562.1	43048
	MRCI+Q	1.951	562.5	2.41	557.7	41767
	MRCI+Q ¹⁴⁶	1.942	564.2	2.7	-	43231
	MRCI+Q ¹⁴⁵	1.95	567.7	2.4	-	45152
$2^3\Sigma^-$	CASSCF	-	-	-	-	-
	MRCI	1.983	522.9	2.12	518.6	42275
	MRCI+Q	1.983	510.5	2.16	506.2	41714
	MRCI+Q ¹⁴⁶	-	-	-	-	-
	MRCI+Q ¹⁴⁵	-	-	-	-	-
$3^1\Pi$	CASSCF	-	-	-	-	52481
	MRCI	1.963	600.1	2.79	594.5	44083
	MRCI+Q	1.955	555.6	2.04	551.5	42790
	MRCI+Q ¹⁴⁶	1.941	550.4	2.9	-	44678
	MRCI+Q ¹⁴⁵	1.944	574.2	2.8	-	45319
$4^3\Pi$	CASSCF	-	-	-	-	49935
	MRCI	1.968	553.4	4.25	544.9	44648
	MRCI+Q	1.972	545.9	4.06	537.8	44453
	MRCI+Q ¹⁴⁶	1.972	553.5	3	-	44849
	MRCI+Q ¹⁴⁵	1.957	553.4	3.2	-	45454
$3^1\Delta$	CASSCF	-	-	-	-	53789
	MRCI	1.996	517.3	-3.92	525.1	45578
	MRCI+Q	1.982	525.5	2.79	519.9	44774
	MRCI+Q ¹⁴⁶	1.975	546.6	3.3	-	43806
	MRCI+Q ¹⁴⁵	1.965	540.3	2.1	-	47006
$2^1\Sigma^-$	CASSCF	-	-	-	-	-
	MRCI	1.98	529.5	2.38	524.6	45660
	MRCI+Q	1995	500.6	-6.47	513.59	45461
$2^3\Sigma^+$	CASSCF	-	-	-	-	-
	MRCI	-	-	-	-	-
	MRCI+Q	-	-	-	-	-
	MRCI+Q ¹⁴⁶	1.9	556.4	2.7	-	47316
	MRCI+Q ¹⁴⁵	1.871	664.6	2.1	-	43031

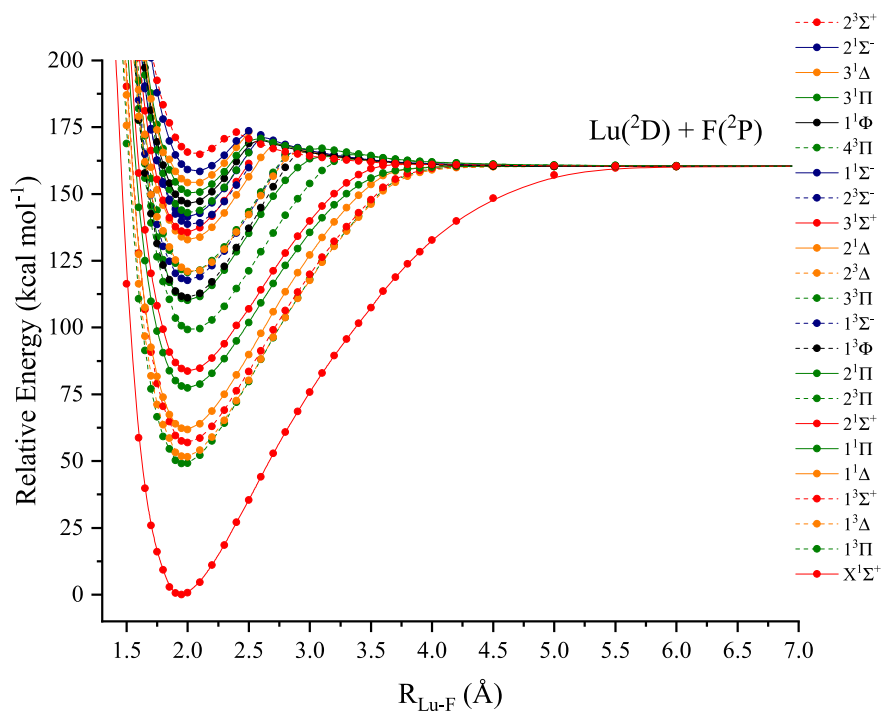


Figure 4.1: CASSCF PECs of LuF with respect to the Lu-F distance.

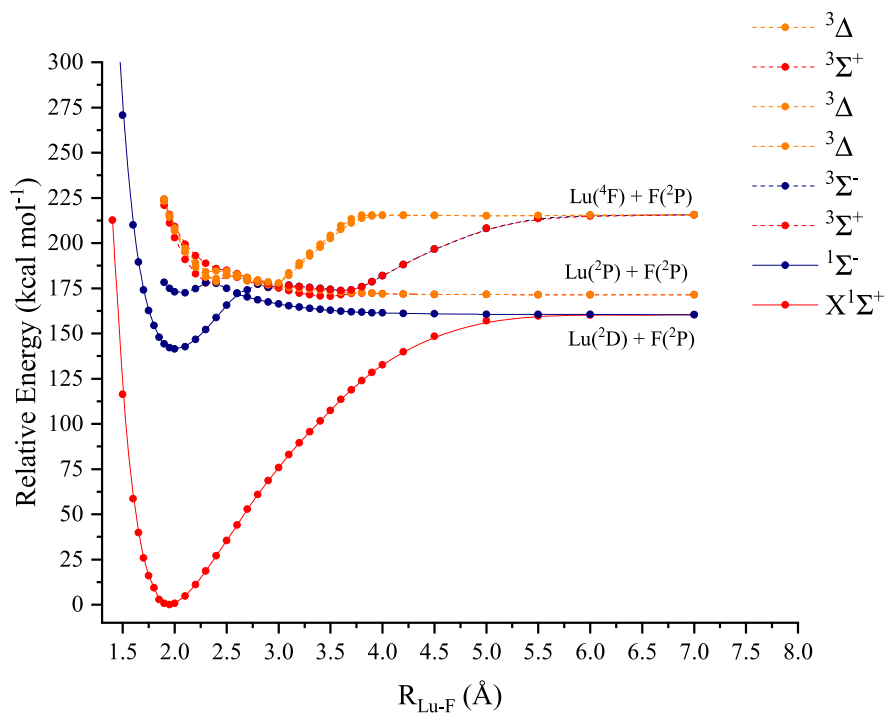


Figure 4.2: Example of intersystem crossing from upper dissociation channels of LuF at the CASSCF level.

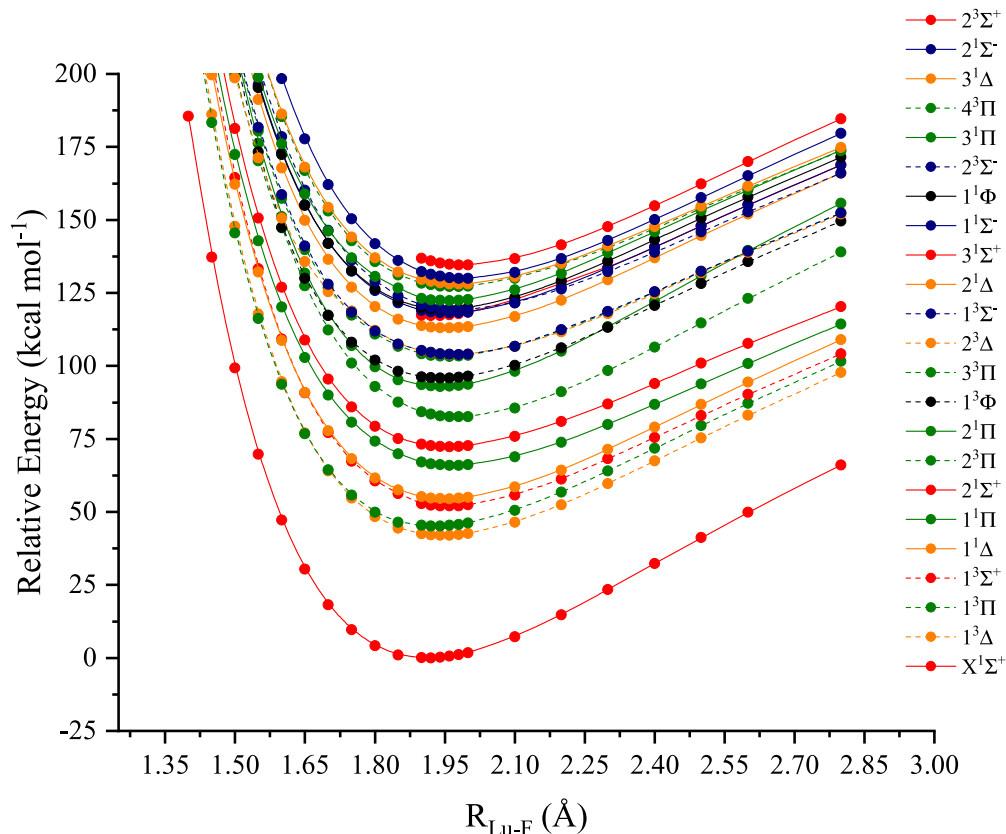


Figure 4.3: MRCI+Q PECs of LuF with respect to the Lu-F distance.

In Table 4.2, the CI vectors of the studied ground and excited states are shown. For the calculations, CASSCF, MRCI and MRCI+Q were used, and for states that were deemed to be single reference in nature, CCSD(T) was employed. For the first part of this work, state averaged CASSCF was used for the 132 states, which aids in describing intersystem crossings that come from upper channels and merge with the Lu (2D) + F(2P) channel. This is the first time such level of detail is considered for LuF, providing insight on how the dissociation channels are formed, and describing some of the higher energy, upper channel intersystem crossings. In addition, no evidence of the presence of the ionic channel ($Lu^+ + F^-$) was found in the MCSCF calculations. Also, from the 132 states studied none of them converged to $Lu^+ + F^-$ at infinity, demonstrated by its CI vectors. The orbital pictures included in the active space at equilibrium bond length (1.92 Å) and at 6 Å are shown in Figures 4.4 and 4.5 respectively.

Table 4.2: CI vectors at equilibrium bond length for LuF were obtained through CASSCF using ECP28MWB/ANO-TZ for Lu and aug-cc-pVTZ for F.

States	Coeff	$1\pi_z$	1σ	$1\delta_{x^2-y^2}$	$1\delta_{z^2}$	$2\pi_z$	$3\pi_z$	$1\pi_x$	$1\delta_{xz}$	$2\pi_x$	$3\pi_x$	$1\pi_y$	$1\delta_{yz}$	$2\pi_y$	$3\pi_y$	$1\delta_{xy}$
X $^1\Sigma^+$	0.91	2	2	0	0	0	0	2	0	0	0	2	0	0	0	0
1 $^3\Delta$	0.94	2	α	α	0	0	0	2	0	0	0	2	0	0	0	0
1 $^3\Pi$	0.91	2	α	0	0	0	0	2	α	0	0	2	0	0	0	0
1 $^3\Sigma^+$	0.96	2	α	0	α	0	0	2	0	0	0	2	0	0	0	0
1 $^1\Delta$	0.59	2	β	α	0	0	0	2	0	0	0	2	0	0	0	0
1 $^1\Pi$	-0.59	2	α	β	0	0	0	2	0	0	0	2	0	0	0	0
	-0.35	2	β	0	0	0	0	2	α	0	0	2	0	0	0	0
	0.35	2	α	0	0	0	0	2	β	0	0	2	0	0	0	0
	-0.50	2	β	0	0	0	0	2	0	α	0	2	0	0	0	0
	0.50	2	α	0	0	0	0	2	0	β	0	2	0	0	0	0
2 $^1\Sigma^+$	-0.65	2	β	0	α	0	0	2	0	0	0	2	0	0	0	0
	0.65	2	α	0	β	0	0	2	0	0	0	2	0	0	0	0
2 $^3\Pi$	0.86	2	α	0	0	0	0	2	0	α	0	2	0	0	0	0
2 $^1\Pi$	-0.50	2	α	0	β	0	0	2	0	0	0	2	0	0	0	0
	0.50	2	β	0	α	0	0	2	0	0	0	2	0	0	0	0
1 $^3\Phi$	0.66	2	0	α	0	0	0	2	α	0	0	2	0	0	0	0
	0.66	2	0	0	0	0	0	2	0	0	0	2	α	0	0	α
3 $^3\Pi$	-0.54	2	0	0	0	0	0	2	0	0	0	2	α	0	0	α
	0.54	2	0	α	0	0	0	2	α	0	0	2	0	0	0	0
2 $^3\Delta$	0.94	2	0	α	α	0	0	2	0	0	0	2	0	0	0	0
1 $^3\Sigma^-$	0.64	2	0	α	0	0	0	2	0	0	0	2	0	0	0	α
	-0.65	2	0	0	0	0	0	2	α	0	0	2	α	0	0	0
2 $^1\Delta$	0.59	2	0	β	α	0	0	2	0	0	0	2	0	0	0	0
	-0.59	2	0	α	β	0	0	2	0	0	0	2	0	0	0	0

Table 4.2: Table (continued)

States	Coeff	$1\pi_z$	1σ	$1\delta_{x^2-y^2}$	$1\delta_{z^2}$	$2\pi_z$	$3\pi_z$	$1\pi_x$	$1\delta_{xz}$	$2\pi_x$	$3\pi_x$	$1\pi_y$	$1\delta_{yz}$	$2\pi_y$	$3\pi_y$	$1\delta_{xy}$
$3^1\Sigma^+$	0.56	2	0	2	0	0	0	2	0	0	0	2	0	0	0	0
	0.57	2	0	0	0	0	0	2	0	0	0	2	0	0	0	2
	-0.30	2	0	0	0	0	0	2	0	0	0	2	2	0	0	0
	-0.30	2	0	0	0	0	0	2	2	0	0	2	0	0	0	0
$1^1\Phi$	0.43	2	0	α	0	0	0	2	β	0	0	2	0	0	0	0
	-0.43	2	0	β	0	0	0	2	α	0	0	2	0	0	0	0
	0.43	2	0	0	0	0	0	2	0	0	0	2	α	0	0	β
	-0.43	2	0	0	0	0	0	2	0	0	0	2	β	0	0	α
$2^3\Sigma^-$	0.71	2	0	α	0	0	0	2	0	0	0	2	0	0	0	α
	-0.34	2	0	0	0	0	0	2	0	α	0	2	α	0	0	0
	-0.34	2	0	0	0	0	0	2	α	0	0	2	0	α	0	0
	0.45	2	0	0	0	0	0	2	α	0	0	2	α	0	0	0
$3^1\Pi$	-0.30	2	0	0	β	0	0	2	α	0	0	2	0	0	0	0
	0.30	2	0	0	α	0	0	2	β	0	0	2	0	0	0	0
$4^3\Pi$	0.80	2	0	0	α	0	0	2	α	0	0	2	0	0	0	0
	0.30	2	0	α	0	0	0	2	α	0	0	2	0	0	0	0
	-0.30	2	0	0	0	0	0	2	0	0	0	2	α	0	0	α
$3^1\Delta$	0.49	2	0	0	0	0	0	2	α	0	0	2	β	0	0	0
	-0.49	2	0	0	0	0	0	2	β	0	0	2	α	0	0	0
$2^1\Sigma^-$	0.48	2	0	0	0	0	0	2	α	0	0	2	0	β	0	0
	-0.48	2	0	0	0	0	0	2	β	0	0	2	0	α	0	0
	-0.48	2	0	0	0	0	0	2	0	α	0	2	β	0	0	0
	0.48	2	0	0	0	0	0	2	0	β	0	2	α	0	0	0
$2^3\Sigma^+$	0.94	2	α	0	0	α	0	2	0	0	0	2	0	0	0	0

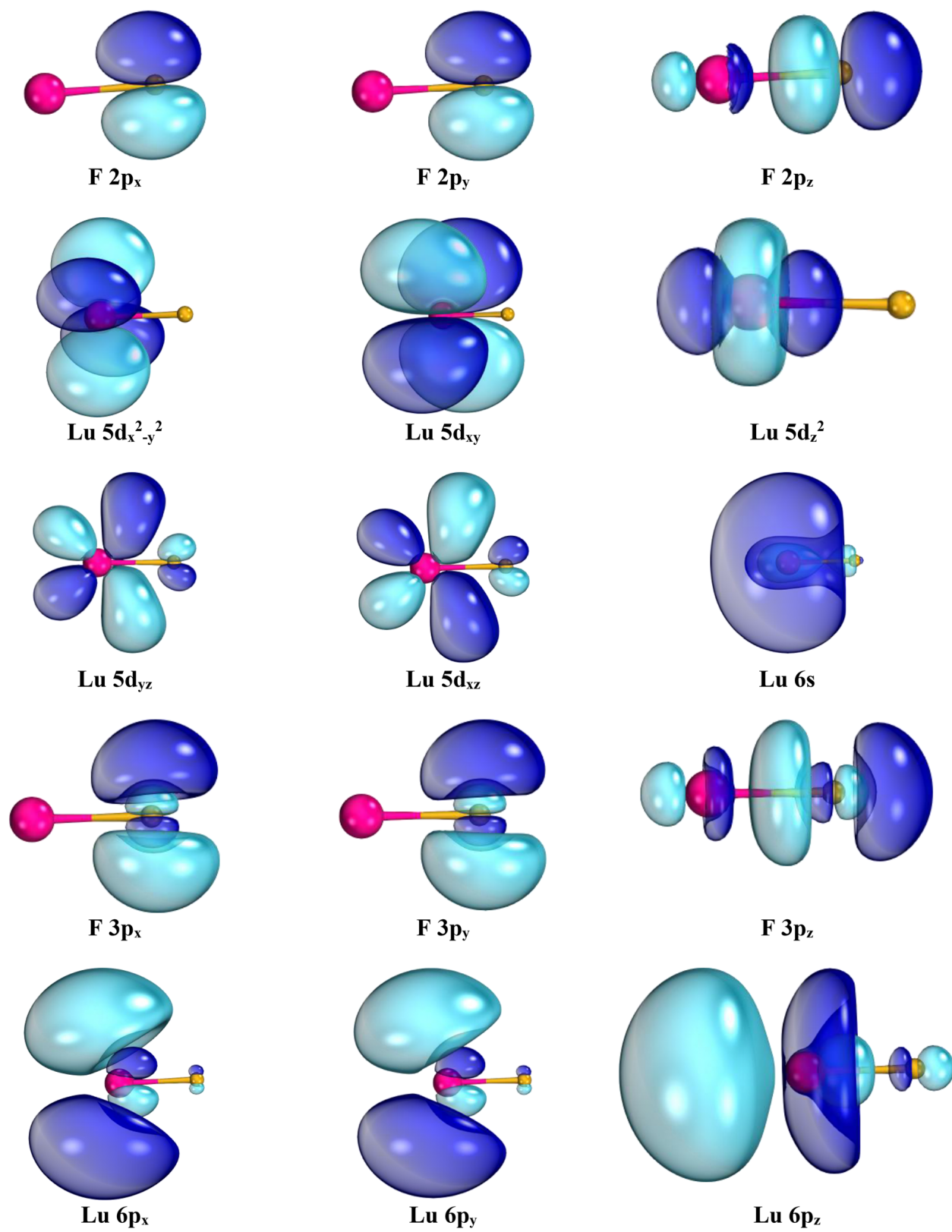


Figure 4.4: Molecular orbitals for LuF at 1.92 Å.

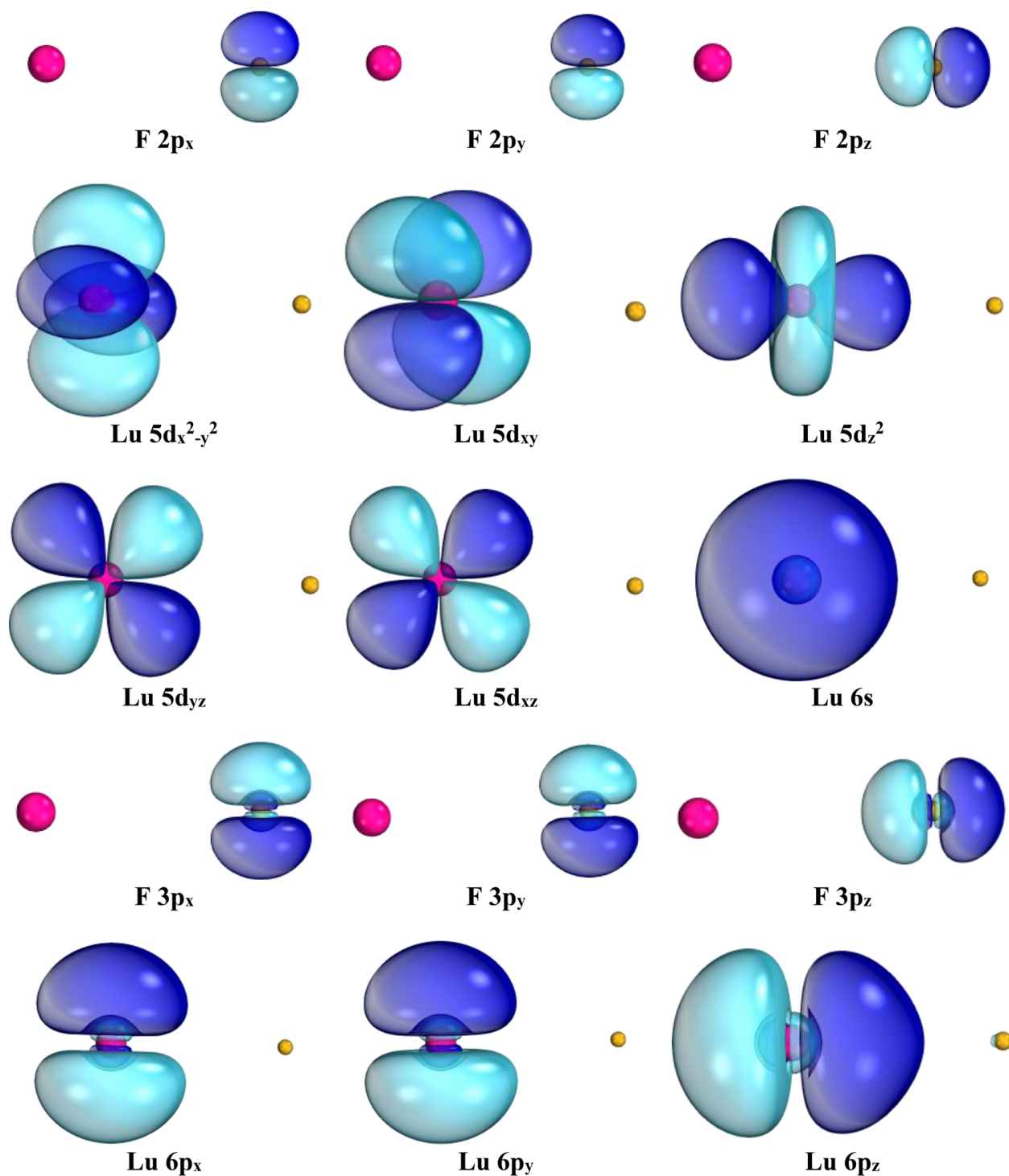


Figure 4.5: Molecular orbitals for LuF at 6.0 Å.

At 6 Å, the orbitals resemble atomic ones, with no mixing between fluorine and lutetium, providing insight about dissociation. The radial distribution using CR-CCSD(T) is plotted in Figure 4.6

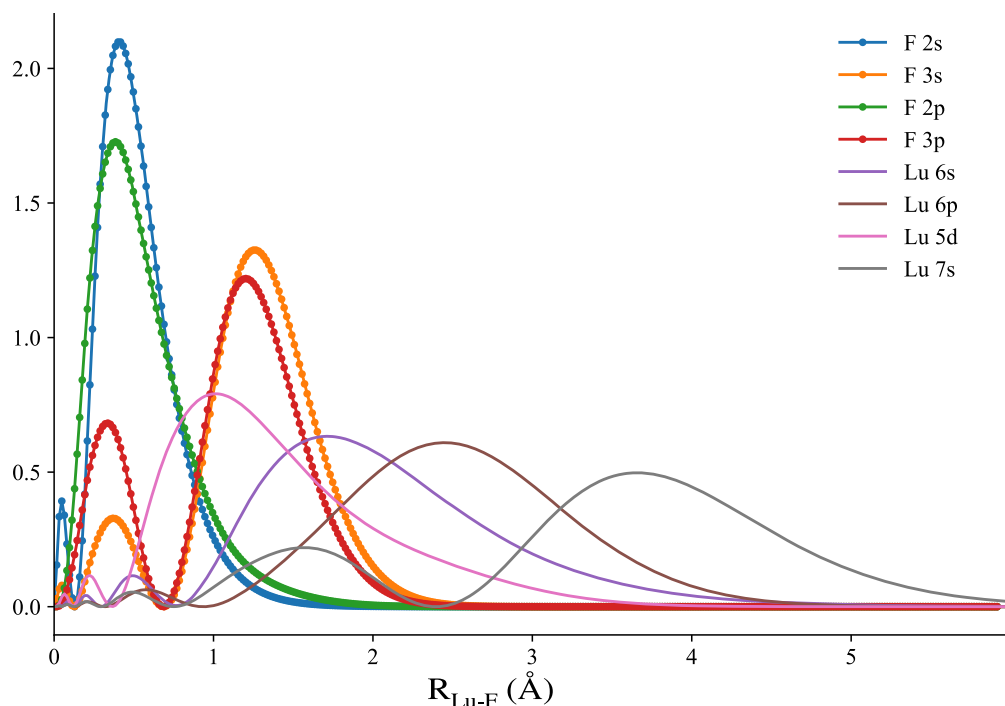


Figure 4.6: Radial distribution functions at the CR-CCSD(T) level.

The large orbital overlap near the equilibrium bond length (1.92 Å) shows that the $3p$ orbitals need to be included at the CASSCF level to describe the full dissociation channels from infinity to equilibrium smoothly. An active space with 15 orbitals in the calculation of full potential energy curves for LuF was deemed necessary to obtain smooth curves. Accounting for the irreducible representation for each spin generates hundreds of thousands of configuration state functions (CSFs), increasing both the complexity of the calculations, as well as the computational time.

According to the Witmer-Wigner angular momentum coupling rules, the four channels generate the following manifolds of states:

first — Lu (2D) + F(2P): $^{1,3}[\Sigma^+(2), \Pi(3), \Delta(2), \Phi, \Sigma^-]$;

second — Lu (2P) + F(2P): $^{1,3}[\Sigma^+(2), \Pi(2), \Delta, \Sigma^-]$;

third and fourth — Lu (4F) + F(2P): $^{3,5}[\Sigma^+(2), \Pi(3), \Delta(3), \Phi(2), \Gamma, \Sigma^-]$;

The calculations show that the ground state is a well separated $^1\Sigma^+$, a closed shell singlet, which is in agreement with experiment.^{123–125} In the ground state, the unpaired $5d^1$ (Lu, at infinity) electron

couples with the unpaired electron on the $2p_z$ orbital of fluorine (see Table 4.2). The spectroscopic constants calculated with CCSD(T) and MRCI/MRCI+Q are all within 1 cm^{-1} of experiment. The next two states were assigned as either $^3\Pi$ or $^3\Delta$ in previous literature. Hamade *et al.* predicted the $^3\Pi$ to be the first excited state and Assaf *et al.* predicted the $^3\Delta$ as the first excited state.^{145,146} According to our calculations, for CASSCF, $1^3\Pi$ is followed by $1^3\Delta$ and their separation is 845 cm^{-1} . However, for MRCI and MRCI+Q, the $^3\Delta$ is the first excited state followed by the $^3\Pi$. The separation of states for MRCI is 713 cm^{-1} and MRCI+Q, 1129 cm^{-1} (see Table 4.2). Both states are a product of electron promotion from the lutetium $6s$ (at infinite separation) to its $5d$ orbitals (see Figures 4.4 and 4.5). In order to generate the $1^3\Delta$ state, an electron populates the $5d_{x^2-y^2}$ (Lu), while for $^3\Pi$, it occupies the $5d_{xz}$ (Lu). These two states are very close in energy and both were assigned a different spin and symmetry in previous experimental data. In the present work, the two experimental values from the literature were assigned to $1^3\Delta$ and $1^3\Pi$.¹²³⁻¹²⁶ Previous theoretical data, from Hamade *et al.* and Assaf *et al.* does not compare the first experimental excited state energy with their first calculated excited state.^{125,126,145,146} Assaf *et al.* assigns their second excited state to A and B from the literature, $^1\Sigma^+$ and $^1\Pi$ respectively.^{125,126,146} The $1^3\Pi$ is in good agreement with experiment for bond lengths and spectroscopic constants, but the $^3\Delta$ is $\sim 1000\text{ cm}^{-1}$ below the experimental value. However, when both $1^3\Delta$ and $1^3\Pi$ are corrected for spin-orbit effects (see Section III.B.), the range of Ω -state energies spans over 3000 cm^{-1} (Table 4.3). The next excited state is $1^3\Sigma^+$, which corresponds to a promotion of an electron from the $6s$ of lutetium to the $5d_{z^2}$. In fact, electronic excitations from $6s \rightarrow 5d$ orbitals occur until $\sim 33000\text{ cm}^{-1}$, as per Table 4.2. States, $1^1\Delta$ and $1^1\Pi$ are the corresponding open-shell singlets of $1^3\Delta$ and $1^3\Pi$, respectively, and are 4384 and 7260 cm^{-1} above the aforementioned, according to MRCI+Q.

$1^1\Delta$ and $1^1\Pi$ are also 1000 cm^{-1} below their assigned experimental states, but their bond length is within $0.01\text{ \AA}$ from experiment. The next three states, $2^1\Sigma^+$, $2^3\Pi$ and $2^1\Pi$, also correspond to the promotion of an electron from the $6s$ (Lu) into the $5d$ orbitals (Lu). $1^3\Phi$ is 33566 and 33499 cm^{-1} above the ground state according to MRCI and MRCI+Q, respectively, and it is the first excited state that has two electrons promoted from the $6s$ (Lu) into $5d$ and $6p$ (Lu) orbitals. There is a $\sim$

3000 cm^{-1} gap in which there are no populated states, but in the 36000 cm^{-1} region, there are three excited states within 200 cm^{-1} of one another according to MRCI+Q ($3^3\Pi$, $2^3\Delta$, and $1^3\Sigma^-$). From 36000 to 50000 cm^{-1} there is a large agglomeration of states, which show mixing from the first two dissociation channels. In this 14000 cm^{-1} or 30 $\text{kcal}\cdot\text{mol}^{-1}$ region, nine states overlap each other. The first state in this region is $2^1\Delta$, followed by $3^1\Sigma^+$. The latter belongs to the next binding channel, Lu ($^2\text{P}; 6s^2 5p^1$) + F(^2P) (see Figure 4.2). This channel is not displayed in Figure 4.1, due to the very large mix of states from upper channels, so only the binding region (2.7Å 1.4Å) is plotted. The other states displayed in Figure 4.1, which belong to the Lu ($^2\text{P}; 6s^2 5p^1$) + F(^2P) channel are: $2^3\Sigma^-$, $4^3\Pi$, $3^1\Delta$ and $2^1\Sigma^-$. The last states that belong to the first binding channel are: $1^1\Sigma^-$, $1^1\Phi$ and $2^3\Sigma^+$. The first $1^1\Sigma^-$ state undergoes intersystem crossings as shown in Figure 4.2 There is a range of singlet and triplet states that couple together after 45000 cm^{-1} ($\sim 175 \text{ kcal}\cdot\text{mol}^{-1}$) from three different dissociation channels, which originate multiple avoided and intersystem crossings.

Spin-orbit calculations

Spin-orbit calculations were performed on the ground state and the first eight excited states of LuF, which covers a region of $\sim 100 \text{ kcal}\cdot\text{mol}^{-1}$ or $\sim 36,000 \text{ cm}^{-1}$. The first nine $^{2S+1}\Lambda$ states split into the Ω -states as follows: $X^1\Sigma^+ \rightarrow X^1\Sigma_{0+}^+$; $1^3\Delta \rightarrow ^3\Delta_1, ^3\Delta_2, ^3\Delta_3$; $1^3\Pi \rightarrow ^3\Pi_{0-}, ^3\Pi_{0+}, ^3\Pi_1, ^3\Pi_2$; $1^3\Sigma^+ \rightarrow ^3\Sigma_{0+}^+, ^3\Sigma_1^+$; $1^1\Delta \rightarrow ^1\Delta_2$; $1^1\Pi \rightarrow ^1\Pi_1$; $2^1\Sigma^+ \rightarrow 2^1\Sigma_{0+}^+$, $2^3\Pi \rightarrow ^3\Pi_{0-}, ^3\Pi_{0+}, ^3\Pi_1, ^3\Pi_2$, $2^1\Pi \rightarrow ^1\Pi_1$, and $1^3\Phi \rightarrow ^3\Phi_2, ^3\Phi_3, ^3\Phi_4$. For singlet states the $\Lambda=0$ is expected to be minimal. The C-MRCI spin-orbit PECs are depicted in Figure 4.7 (spin-orbit states with the same Ω value have the same color) and MRCI spin-orbit are depicted in the appendix (Figure B.1 and Table B.2). The bond lengths and spectroscopy constants are included in Table 4.3, and the decomposition of the spin-orbit states is included in Table 4.4.

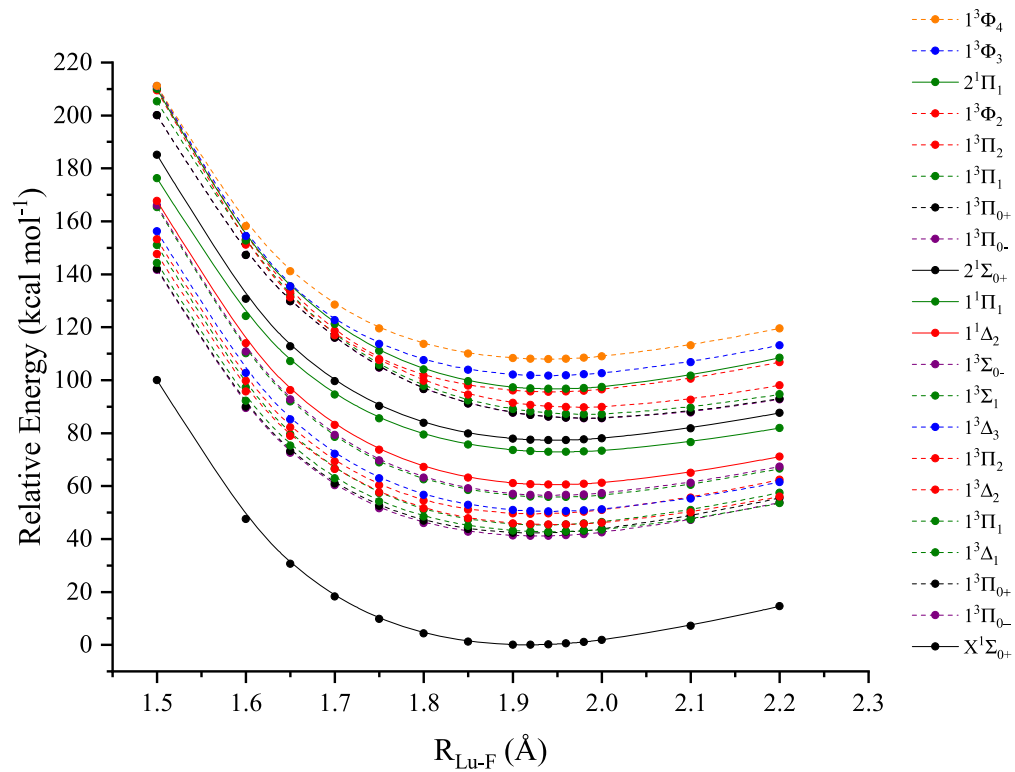


Figure 4.7: Core-spin-orbit MRCI (C-MRCI) PECs of LuF with respect to the Lu-F distance.

Table 4.3: Computational method, bond length R_{eq} (Å), harmonic vibrational frequencies ω_e (cm^{-1}), anharmonicity $\omega_e \chi_e$ (cm^{-1}), $\Delta G_{1/2}$ (cm^{-1}) values, and excitation energy T_e (cm^{-1}) for the lowest electronic excited states of $^{175}\text{Lu } ^{19}\text{F}$ at the spin-orbit level. The state are ordered according to C-MRCI energetics. MRCI and C-MRCI calculations were performed using an ECP28MWB/Def2-QZPP for Lu and aug-cc-pVQZ for F.

States	Methodology	R_{eq}	ω_e	$\omega_e \chi_e$	$\Delta G_{1/2}$	T_e
$X^1\Sigma_{0+}^+$	MRCI	1.917	614.90	2.24	610.4	0
	C-MRCI	1.913	618.20	2.52	613.2	0
	MRCI ¹⁴⁶	1.914	619.40	2.53	-	0
$1^3\Pi_{0-}$	MRCI	1.933	584.70	2.72	579.3	13831
	C-MRCI	1.924	595.60	2.59	590.4	14377
	MRCI ¹⁴⁶	1.938	573.30	2.58	-	14629
$1^3\Pi_{0+}$	MRCI	1.928	590.00	2.62	584.9	14270
	C-MRCI	1.919	601.30	2.57	596.2	14788
	MRCI ¹⁴⁶	1.935	577.30	2.69	-	15003
$1^3\Delta_1$	MRCI	1.952	567.20	2.81	561.6	13866
	C-MRCI	1.939	565.00	1.62	561.8	14943
	MRCI ¹⁴⁶	1.949	572.40	2.57	-	13513
$1^3\Pi_1$	MRCI	1.935	585.90	1.76	582.4	15142
	C-MRCI	1.932	601.30	3.29	594.7	15844
	MRCI ¹⁴⁶	1.938	571.30	2.59	-	15600
$1^3\Delta_2$	MRCI	1.953	569.30	2.3	564.7	14781
	C-MRCI	1.943	571.30	2.88	565.6	15890
	MRCI ¹⁴⁶	1.949	572.70	2.82	-	14435
$1^3\Pi_2$	MRCI	1.929	589.80	2.48	584.8	16774
	C-MRCI	1.922	601.00	2.62	595.7	17313
	MRCI ¹⁴⁶	1.931	580.50	2.66	-	16884
$1^3\Delta_3$	MRCI	1.947	576.40	2.43	571.5	16748
	C-MRCI	1.941	579.70	2.2	575.2	17641
	MRCI ¹⁴⁶	1.946	576.40	2.53	-	16170
$1^+\Sigma_1^+$	MRCI	1.952	574.60	2.51	569.5	18955
	C-MRCI	1.945	581.60	2.52	576.5	19520
	MRCI ¹⁴⁶	1.956	567.70	2.8	-	19101

Table 4.3: Table (continued)

States	Methodology	R_{eq}	ω_e	$\omega_e \chi_e$	$\Delta G_{1/2}$	T_e
$1^3\Sigma_{0-}^+$	MRCI	1.952	575.90	2.74	570.4	19238
	C-MRCI	1.945	581.60	2.48	576.7	19782
	MRCI ¹⁴⁶	1.955	567.90	2.77	-	19352
$1^1\Delta_2$	MRCI	1.953	570.60	2.07	566.5	19902
	C-MRCI	1.946	576.80	2.38	572	21180
	MRCI ¹⁴⁶	1.954	570.20	2.54	-	19702
$1^1\Pi_1$	MRCI	1.967	534.10	2	530.1	24403
	C-MRCI	1.954	544.10	2.03	540.1	25493
	MRCI ¹⁴⁶	1.969	528.65	2.31	-	23839
$2^1\Sigma_{0-}^+$	MRCI	1.955	562.70	2.71	557.3	26211
	C-MRCI	1.946	572.50	2.82	566.9	27042
	MRCI ¹⁴⁶	1.959	553.10	2.6	-	26037
$2^3\Pi_{0-}$	MRCI	1.993	532.10	-	556.6	28782
	C-MRCI	1.991	523.80	-	566.08	29920
	MRCI ¹⁴⁶	1.984	573.60	2.61	-	28744
$2^3\Pi_{0+}$	MRCI	1.991	536.80	-	557.5	28818
	C-MRCI	1.989	523.40	-	568.3	29959
	MRCI ¹⁴⁶	1.984	574.50	2.58	-	28730
$2^3\Pi_1$	MRCI	1.988	570.20	-	556.8	29369
	MRCI	1.986	567.80	-	556	30461
	C-MRCI ¹⁴⁶	1.982	577.50	2.64	-	29291
$2^3\Pi_2$	MRCI	1.981	656.30	-	586.9	30345
	C-MRCI	1.979	611.70	-	564.2	31373
	MRCI ¹⁴⁶	1.978	583.20	2.66	-	30095
$1^3\Phi_2$	MRCI	1.948	512.80	-	536.9	31774
	C-MRCI	1.939	531.60	-	553.4	33444
	MRCI ¹⁴⁶	1.951	560.60	2.41	-	31877
$2^1\Pi_1$	MRCI	1.949	577.15	-	579.3	32891
	C-MRCI	1.946	584.10	-	590.5	33812
	MRCI ¹⁴⁶	1.957	594.72	2.35	-	32921
$1^3\Phi_3$	MRCI	1.946	597.70	-	586.9	34013
	C-MRCI	1.937	608.90	-	595.5	35587
	MRCI ¹⁴⁶	1.946	565.90	2.41	-	33965
$1^3\Phi_4$	MRCI	1.942	571.80	-	568.7	36287
	C-MRCI	1.934	583.00	-	578.7	37762
	MRCI ¹⁴⁶	1.949	494.93	2.53	-	36218

Table 4.4: Spin-orbit composition at the C-MRCI level (1.92 Å) for the lowest excited state of $^{175}\text{Lu}^{19}\text{F}$.

State	Composition
$X^1\Sigma^+_{0+}$	99.87% $X^1\Sigma^+$, 0.06% $1^3\Pi$, 0.08% $2^3\Pi$
$1^3\Pi_{0-}$	89.24% $1^3\Pi$, 10.76% $1^3\Sigma^+$
$1^3\Pi_{0+}$	98.44% $1^3\Pi$, 1.49% $2^1\Sigma^+$, 0.06% $X^1\Sigma^+$
$1^3\Delta_1$	42.49% $1^3\Delta$, 53.19% $1^3\Pi$, 3.81% $1^3\Sigma^+$, 0.26% $1^1\Pi$, 0.16% $2^1\Pi$, 0.08% $2^3\Pi$
$1^3\Pi_1$	45.12% $1^3\Pi$, 46.13% $1^3\Delta$, 7.32% $1^3\Sigma^+$, 1.34% $1^1\Pi$, 0.07% $2^3\Pi$, 0.02% $2^1\Pi$
$1^3\Delta_2$	69.16% $1^3\Delta$, 25.20% $1^3\Pi$, 5.30% $1^1\Delta$, 0.24% $1^3\Phi$, 0.08% $2^3\Pi$
$1^3\Pi_2$	74.0% $1^3\Pi$, 25.83% $1^3\Delta$, 0.12% $1^1\Delta$, 0.02% $2^3\Pi$, 0.02% $1^3\Phi$
$1^3\Delta_3$	99.95% $1^3\Delta$, 0.05% $1^3\Phi$
$1^3\Sigma^+_1$	86.83% $1^3\Sigma^+$, 11.78% $1^3\Pi$, 1.14% $1^1\Pi$, 0.24% $1^3\Delta$, 0.01% $2^3\Delta$
$1^1\Sigma^+_{0-}$	89.22% $1^3\Sigma^+$, 10.76% $1^3\Pi$, 0.02% $2^3\Pi$
$1^1\Delta_2$	93.78% $1^1\Delta$, 4.84% $1^3\Delta$, 0.78% $1^3\Pi$, 0.36% $2^3\Pi$, 0.24% $1^3\Phi$
$1^1\Pi_1$	97.21% $1^1\Pi$, 2.03% $1^3\Sigma^+$, 0.57% $1^3\Pi$, 0.14% $1^3\Delta$, 0.06% $2^3\Pi$
$2^1\Sigma^+_{0+}$	98.32% $2^1\Sigma^+$, 1.50% $1^3\Pi$, 0.18% $2^3\Pi$
$2^3\Pi_{0-}$	99.98% $2^3\Pi$, 0.02% $1^3\Pi$, 0.01% $1^3\Sigma^+$
$2^3\Pi_{0+}$	99.74% $2^3\Pi$, 0.19% $2^1\Sigma^+$, 0.07% $X^1\Sigma^+$
$2^3\Pi_1$	94.6%, 5.26% $1^1\Pi$, 0.08%, 0.01%
$2^3\Pi_2$	99.28% $2^3\Pi$, 0.46% $1^1\Delta$, 0.14% $1^3\Phi$, 0.13% $1^3\Delta$
$1^3\Phi_2$	99.52% $1^3\Phi$, 0.34% $1^1\Delta$, 0.10% $1^3\Pi$, 0.05% $1^3\Delta$
$2^1\Pi_1$	94.56% $2^1\Pi$, 5.19% $2^3\Pi$, 0.22% $1^3\Delta$, 0.03% $1^3\Pi$
$1^3\Phi_3$	99.95% $1^3\Phi$, 0.06% $1^3\Delta$
$1^3\Phi_4$	100% $1^3\Phi$

The ground state of LuF ($X^1\Sigma^+$), $1^1\Delta$, $1^1\Pi$ and $2^1\Sigma^+$ remain almost unaffected due to zero first order spin-orbit effects. Without spin-orbit effects, the $1^3\Delta$ is the first excited state followed by $1^3\Pi$, which is $\sim 1200 \text{ cm}^{-1}$ higher in energy according to MRCI+Q. However, with spin-orbit correction, the ordering of Ω -states is more complex to assess due to the closeness of the energetics gaps. The $^3\Delta$ and $^3\Pi$ states, spin-orbit corrected at MRCI and C-MRCI level follow the same ascending order: $^3\Pi_{0-}$, $^3\Pi_{0+}$, $^3\Pi_1$, $^3\Pi_2$ and $^3\Delta_1$, $^3\Delta_2$, $^3\Delta_3$. According to C-MRCI, the $1^3\Pi_{0-}$ is the first excited state followed by $1^3\Pi_{0+}$, which is $\sim 400 \text{ cm}^{-1}$ above in energy. However, for MRCI, the $1^3\Delta_1$ is the second excited followed by $1^3\Pi_{0+}$. The effect of the core orbitals is also felt on the bond lengths of $1^3\Pi_{0-}$, $1^3\Pi_{0+}$, and $1^3\Delta_1$ which drop by $\sim 0.01 \text{ Å}$ when using C-MRCI. For C-MRCI, the third

excited state is $^3\Delta_1$, followed by $^3\Pi_1$, $^3\Delta_2$, $^3\Pi_2$, and $^3\Delta_3$. The Ω -states of $1^3\Delta$ and $1^3\Pi$ span over a range of more than 3000 cm^{-1} , which shows a large spin-orbit contribution and the importance of including inner core correlation.

When comparing this work with Assaf *et al.*, their state ordering is different, and the $1^3\Delta_1$ is their first excited state followed by $1^3\Delta_2$ and then $1^3\Pi_{0-}$. These differences can be attributed to the use of a more state specific approach in the CASSCF and MRCI calculations, a higher level basis set in the present study. The inner orbitals of lutetium and fluorine were not considered in their calculations, but only the lutetium sub-valence $4f^{14}$ was included along with the $2p^5$ of fluorine. The use of inner core orbitals results in significant differences in bond lengths and spectroscopic constants.

In terms of composition (Table 4.4), Ω -states=1, 2 for $1^3\Delta$ and $1^3\Pi$ are heavily mixed, but $1^3\Delta_3$ can only mix with $1^3\Phi_3$. The next excited is $^3\Sigma^+$, which splits into $^3\Sigma_{0+}^+$ and $^3\Sigma_1^+$. The bond length dropped $\sim 0.07\text{ \AA}$ when using C-MRCI, and the T_e is $\sim 500\text{ cm}^{-1}$ for both Ω states above MRCI. The next three states have minimal spin-orbit effects, but the inclusion of the core orbitals for C-MRCI changed their bond lengths by almost $0.1\text{ \AA}$, and the T_e is $\sim 1000\text{ cm}^{-1}$ above MRCI. The last three states considered in Figure 4.3 are $2^3\Pi$, $2^1\Pi$ and $1^3\Phi$. The $2^3\Pi$ follows the same ordering for its Ω states as the $1^3\Pi$. When comparing MRCI and C-MRCI, the bond length for this state only varies $0.02\text{ \AA}$ on average. C-MRCI still sits $\sim 1000\text{ cm}^{-1}$ above MRCI. $2^1\Pi_1$ is in between the $1^3\Phi$ Ω states. $1^3\Phi_2$ is a heavily mixed state as reported in Table 4.4. $1^3\Phi_3$ can only mix with $1^3\Delta_3$, but $1^3\Phi_4$ is a pure state. For the $1^3\Phi$ splitting, C-MRCI also drops the bond length by almost $\sim 0.1\text{ \AA}$, for the three Ω states. The T_e for C-MRCI is also on average 1000 cm^{-1} above MRCI.

Dissociation energy

Dissociation energies calculated in this work as well as those reported previously from both theoretical and experimental studies are included in Table 4.5.

Table 4.5: Dissociation energy of LuF in kcal·mol⁻¹ with different levels of theory and a range of basis sets.

Method	Frozen-core	Relativistic treatment	D ₀ (dz)	D ₀ (tz)	D ₀ (qz)	D ₀ CBS
CR-CCSD(T)	FC-val	DKH3	177.93	180.35	180.38	180.3
CR-CCSD(T)	FC-subval	DKH3	163.01	167.20	167.66	167.9
CCSD(T)	FC-val	DKH3	178.99	182.16	182.4	182.3
CCSD(T)	FC-subval	DKH3	164.16	169.25	169.96	170.4
CCSD(T)	FC-val	ECP28-Def2	-	172.35	169.47	167.9
CCSD(T)	FC-subval	DC	158.33	165.63	-	
MP2		DC	158.78	167.19	-	
HF		DC	134.15	137.49	137.74	137.8
CASSCF(8,15)		ECP28-ANO	-	159.74	-	-
PBE		DKH3	174.4	172.29	176.04	-
TPSS		DKH3	170.64	167.87	171.73	-
M06-L		DKH3	169.02	169.77	171.93	-
B3LYP		DKH3	167.72	166.09	169.97	-
Other theoretical values						
Composite ¹⁴³						169.7
Composite ¹⁴⁴						173.32
PP-CCSD(T) ¹³⁸	Valence	ECP60				173
PP-MRACPF ¹³⁸	Valence	ECP60				175
DFT: SOAP ¹³⁹					96.6	-
PBE ¹⁴⁰		ZORA				174
DFT ^{141,a}		DKH3				195.3-161.6
EOM-CR-CCSD(T) ¹⁴²	Valence	DKH3				171.3
EOM-CR-CCSD(T) ¹⁴²	Full	DKH3				139.6
Experimental value						
Mass Spectroscopy ¹²¹			136 ± 12			
Ligand Field theory ¹²²			124			
Fitting PES ¹²⁸			79			
Fitting PES ¹²³			105			
Fitting PES ¹²⁷			96.0 ± 2.4			

^a DFT functionals used are SVWN, BP86, BLYP, PW91, PBE, B97-D, SSB-D, M06-L, TPSS, PBE0, B3LYP, B3LYP, B3P86, MPW1K, B97-1, X3LYP, M06, M06-2X, TPSSh, M11, CAM-B3LYP, and B2PLYP.

For the correlation, two approaches to the valence space were considered: FC-val, which includes only valence electrons ($6s^2$, $5d^1$ of Lu and $2s^2$, $2p^5$ of F) and FC-subval includes sub-valence orbitals ($5s^2$, $5p^6$ of Lu). In addition, the effects of using a full relativistic Hamiltonian, and ECPs (28 electrons) were probed. For *ab initio* calculations, CCSD(T), CR-CCSD(T), and MP2 were

utilized. For DFT, a variety of functionals were considered: PBE, TPSS, M06-L and B3LYP.

The dissociation energy difference between the Sapporo-DZ and Sapporo-TZ for CR-CCSD(T) is 2 kcal·mol⁻¹ with FC-val, while between Sapporo-TZ and Sapporo-QZ basis sets, the energy difference drops to 0.03 kcal·mol⁻¹ which implies that the energy is almost converged at the triple- ζ level. The same trend is observed for CCSD(T), where the energy at the triple- ζ level is almost converged. When the sub-valence electrons from Lu are added (FC-subval results), the dissociation energy with CR-CCSD(T)/Sapporo-DZ dropped by 14 kcal·mol⁻¹ and by $\sim$ 13 kcal·mol⁻¹ at the CBS limit. At the CCSD(T) level of theory, the difference between FC-val and FC-subval dissociation energies are $\sim$ 14 and $\sim$ 12 kcal·mol⁻¹ with the Sapporo-DZ and at the CBS limit, respectively. Such a large difference arising from the choice of valence indicates that the electron correlation arising from the sub-valence electrons is important in the overall energy.

The basis set superposition error has been investigated by using the counterpoise method suggested by Boys and Bernardi for CCSD(T) and CR-CCSD(T) at the CBS limit for FC-val and FC-subval.¹⁶² For both FC-val calculations, considering CCSD(T) and CR-CCSD(T), the BSSEs extrapolated to the CBS limit using a mixed exponential/gaussian by Peterson is 0.87 kcal·mol⁻¹.⁶⁰ For CCSD(T) and CR-CCSD(T) using sub-valence electrons, 0.81 and 0.59 kcal·mol⁻¹ were obtained, respectively for BSSE corrections at CBS. As an example, for CCSD(T)/FC-subval at a double-, triple- and quadruple- ζ basis set levels, the BSSE is 6.82, 3.52 and 1.21 kcal·mol⁻¹ respectively, which at CBS yields 0.81 kcal·mol⁻¹.

In addition, the dissociation energy of LuF was evaluated using the ECP28MWB pseudopotential and Def2-TZVPP, Def2-QZVPP (Lu) and aug-cc-pVTZ, aug-cc-PVQZ (F) basis sets. The value obtained at the quadruple- ζ level is very close to DKH3 predictions mentioned earlier, while the triple- ζ result is slightly higher than the DKH3 dissociation reported. The pseudopotential used for lutetium accounts for relativistic effects arising from the inner-core electrons. To evaluate the spin-orbit contribution to the ground state, the Dirac-Coulomb (DC) four component Hamiltonian was utilized. CCSD(T), MP2 and HF were probed for this step. The utility of the double- and triple- ζ CBS extrapolation by Martin¹⁶⁰ has been considered for CCSD(T)/FC-subval/DKH3. This

double-, triple- ζ CBS extrapolation scheme results in a dissociation energy of 171.1 kcal·mol⁻¹, while considering a two-point scheme extrapolation with triple- ζ and quadruple- ζ basis sets, 170.4 kcal·mol⁻¹ is obtained. Considering the unextrapolated triple- ζ basis set, the value obtained is 169.3 kcal·mol⁻¹. For CCSD(T)/FC-subval/DKH3, the double-, triple- ζ CBS extrapolated energy is closer to the triple-,quadruple- ζ extrapolated energy than the unextrapolated triple- ζ energy.

This shows that the spin-orbit contribution is small to the ground state, which is expected for a $^1\Sigma^+$. In terms of calculations at the Hartee-Fock level, the necessary electron correlation is not present, so its dissociation energy prediction is very far from the best estimate. Finally, CASSCF was also used to calculate the dissociation energy by using the state-averaged wavefunction utilized to construct Figure 4.1 The prediction is 159.74 kcal·mol⁻¹ at a triple- ζ level, which is ~ 9 kcal·mol⁻¹ from the CCSD(T)/DKH3/FC-subval dissociation energy.

CR-CCSD(T) and CCSD(T) results obtained in this study are in good agreement with other theoretical dissociation energies from literature. When comparing the current results with Solomonik and Smirnov, a difference of 2 kcal·mol⁻¹ is obtained when using a sub-valence space correlation.¹⁴³ Solomoniks dissociation energy was obtained with a composite scheme based on CCSD(T)/CBS with core-valence correlation energy, spin-orbit and scalar relativistic effects. The CCSD(T)/CBS results herein, are in very good agreement with previous work from Lu.¹⁴⁴ A composite scheme utilizing the Feller-Peterson-Dixon scheme, renders a value of 173.32 kcal·mol⁻¹, which is only ~ 3 kcal·mol⁻¹ from our best CCSD(T)/CBS results and 5 kcal·mol⁻¹ from CR-CCSD(T). K $\ddot{u}$ chle *et al.*,¹³⁸ used the multireference averaged coupled-pair functional (MRACPF), and their dissociation energy is 4 and 7 kcal·mol⁻¹ higher than the results obtained in the CCSD(T)/CBS and CR-CCSD(T)/CBS predictions herein, respectively. However, both CCSD(T)/CBS and CR-CCSD(T)/CBS dissociation energies, are quite distant from reported experimental values. In Table 4.4, the smallest difference in dissociation energy between experiment and our predictions was obtained by mass-spectroscopy (Zmbov and Margrave,¹²¹ 136 kcal·mol⁻¹). The other experimental values presented in Table 4.4 have large energetic differences from our calculated values, with a maximum ΔE of ~ 90 kcal·mol⁻¹. This shows the large discrepancy between experiment and theory.

Additionally, the potential utility of several DFT functionals in the determination of the dissociation energy of LuF has been considered. The PBE, TPSS, M06-L and B3LYP functionals have been used, along with a DKH3 Hamiltonian, and the Sap-*nz* basis set. The PBE dissociation energy obtained in this study is in agreement with the one predicted by Hong *et al.*¹⁴⁰ using PBE and the ZORA Hamiltonian. The dissociation energies obtained with the functionals are in a range between 166 and 176 kcal·mol⁻¹. B3LYP at the triple- ζ level results in the lowest dissociation energy (166.09 kcal·mol⁻¹), while PBE with the quadruple- ζ basis set leads to the largest dissociation (176.04 kcal·mol⁻¹). These results largely compare with the DFT dissociation energies reported by Grimmel and co-workers.¹⁴¹ However, in their study, a larger range of functionals were used, with SVWN leading to the largest dissociation energy at 195.3 kcal·mol⁻¹, and BHLYP resulting in the lowest energy at 161.6 kcal·mol⁻¹. Moreover, from the prior effort, B97-1 predicted a dissociation energy that is the closest to our CCSD(T)/CBS with DKH3/FC-subval dissociation energy. Finally, when comparing PBE, TPSS, M06-L and B3LYP dissociation energies from our work and Grimmel *et al.*, PBE has the largest dissociation energy among the four functionals and B3LYP the lowest. (To note, the differences between the Grimmel study and the present one are the use of a larger basis set (quadruple- ζ) in this study as well as a different type of basis set for the ligand.)

4.4 Conclusion

The bond lengths, spectroscopic constants, energetics, and potential energy curves are reported, which include four dissociation channels, and detailed information concerning intersystem and avoided crossings. In addition, spin-orbit effects are calculated at a level of correlation that, can aid experimentalists in further pursuits of the description of the ground and excited states and their spectroscopic data. The use of sub-valence orbitals at spin-orbit demonstrated that they are necessary to recover the necessary correlation to obtain results that are in agreement with experiment, especially for the low-lying excited states. The first excited state of LuF at spin-orbit C-MRCI is $1^3\Pi_{0-}$, followed by $1^3\Pi_{0+}$ and $1^3\Pi_1$, which shows the importance of considering sub-valence and inner core orbitals to calculate spectroscopic constants and bond lengths.

In the second part of this work, the sub-valence orbitals are of paramount importance for

predicting dissociation energies and can shift the dissociation energy by up to $\sim 13 \text{ kcal}\cdot\text{mol}^{-1}$. CR-CCSD(T) and CCSD(T) at the CBS limit estimate the dissociation energy as 167.9 and 170.4 $\text{kcal}\cdot\text{mol}^{-1}$, respectively. Utilizing a four-component Hamiltonian (Dirac-Coulomb) resulted in a dissociation energy $\sim 2 \text{ kcal}\cdot\text{mol}^{-1}$ lower than the DKH3 calculations. The DFT calculations are overall in good agreement with our best estimate (from ~ 1 to $\sim 6 \text{ kcal}\cdot\text{mol}^{-1}$ to 170.40 $\text{kcal}\cdot\text{mol}^{-1}$). Due to the large discrepancies between the results in this study as well as other theoretical data and the experiment, the experimental dissociation energy might need to be revisited. Finally, while in this case DFT gave similar dissociation than *ab initio* methods, a study of an open-shell molecule with a multireference character at the ground state might need more robust methods such as *ab initio* method.

Overall, lanthanide species are difficult to investigate from both theoretical and experimental perspectives. The high density of states, which can be very close in energy (herein, 132 states, most of which are bound and in a $\sim 55000 \text{ cm}^{-1}$ range, just below the dissociation energy), the effect of spin-orbit on the ground and excited states, as well as the influence of the sub-valence electrons are effects that should be included in a detailed analysis. *Ab initio* methods, as utilized herein, are vital to the description of the complex electronic manifold. Already for diatomics, such analysis is significantly demanding, and requires judicious selection of the active space, the electron correlation space and the method.

CHAPTER 5

***AB INITIO* COMPOSITE STRATEGIES AND MULTIREFERENCE APPROACHES FOR LANTHANIDE SULFIDES AND SELENIDES**

Reproduced from N. M. S. Almeida, T. R. L. Melin, S. C. North, B. K. Welch and A. K. Wilson, *J. Chem. Phys.* **157**, 024105 (2022),⁶³ with the permission of AIP publishing. Timoth  Melin, Dr. Nuno Almeida, Dr. Sasha North and Dr. Bradley Welch each focused on four molecules. The DFT calculations on ErSe and the CCSD(T)/CBS calculation were done by Timoth  Melin. The data presentation in the interest of reporting the entire scientific study was done by Timoth  Melin, and he came up with the categorization scheme. The initial draft was written by Dr. Nuno Almeida and each author contributed equally to develop and finalize the paper.

5.1 Introduction

Lanthanide compounds have a number of important uses, such as in catalysis, phosphors, and magnets, with applications in superconductors, hybrid cars, display devices, stealth technology, and potential anti-cancer agents.¹⁶⁴ Lanthanide chalcogenide (selenium and sulfur) clusters have been synthesized^{165–167} for use in optical fibers,^{168–172} and lanthanide sulfides have applications in thermoelectric materials,¹⁷³ infrared window materials,¹⁷⁴ and photovoltaic energy materials.¹⁷⁵ In all of these applications, thermodynamic properties, such as enthalpies of formation, bond dissociation energies, and reaction enthalpies, are of paramount importance in understanding reactivity. However, fundamental properties such as bond dissociation energies (BDEs) can be difficult to describe for heavy element species.

Gaining structural and energetic insight into lanthanides requires both experiment and theory. Experimentally, there is a severe lack of small molecule experimental data. One of the reasons for the lack of experimental studies for some heavy element species is attributed to their radioactivity (for example, promethium in the lanthanide series, and all the actinide series), limiting studies due to special facilities, and/or requirements in their handling. Some elements are not naturally occurring, at least in appreciable amounts (promethium in the lanthanide series and all but uranium and thorium in the actinide series) and, therefore, must be synthesized, which can be a slow (sometimes an

atom-at-a-time!) and costly process.

For lanthanide species, historically, the Knusden cell massspectrometry has been used to study equilibrium vapor properties at high temperatures. Prior work, largely done from the late 1960s to the 1980s, has provided rare-earth gas phase thermodynamic data, including bond dissociation energies.^{176–178} However, the bond dissociation energies determined from such experiments were obtained through the use of the second and third laws of thermodynamics, requiring high precision measurements, which were often not achieved with the Knusden cell approach. As well, the use of these laws required statistical thermodynamic assumptions, which led to significant errors for a number of species.¹⁷⁹

Theory provides another route to address lanthanide species. However, due to the energetic complexity of lanthanide-containing molecules, *ab initio* calculations on these species can be computationally prohibitive [in terms of central processing unit (CPU) times to obtain results, required memory, and disk space], so predictions for these species have traditionally been dominated by density functional theory (DFT). For example, Dolg *et. al.* performed DFT calculations using the B88 and P86 functionals combined with small core pseudopotentials (ECP) on lanthanide high spin complexes, specifically on GdX diatomics (X = H, N, O, F, P, S, Cl, and Gd), to predict chemical binding.¹⁸⁰ Luo and co-workers used DFT approaches to predict the first ionization potentials for lanthanide monosulfides (LnS where Ln = La, Ce, Eu, Gd, Yb, and Lu).¹⁷⁰ Three density functionals were used: the Becke functional for exchange-correlation paired with Perdew (BP),¹⁵ Perdew and Wang (BPW),^{181,182} and Lee-Yang-Par (BLYP).¹⁸³ The Perdew functional yielded the smallest mean absolute deviations from the experimental first ionization potentials and, thus, was employed to calculate bond lengths, vibrational frequencies, populations, and dissociation energies for the diatomic lanthanide sulfides. Xu *et. al.*¹⁸⁴ examined the electronic configurations of LnX with Ln = La-Eu and X = O, S, Se, and Te using relativistic DFT [Vosko-Wilk-Nusair (VWN), Perdew-Burke-Ernzerhof (PBE), BP, PW91, and BLYP functionals]. The authors used the fractional occupation number approach to determine the ground state electronic configurations for the diatomic, concluding that a σ molecular orbital is involved in the bonding when the lanthanide atom has less

than three electrons in its 4*f* orbital shell. The most intensive single reference-based computations to date were carried out on a set of lanthanum oxide and halides by Solomonik and Smirnov.¹⁴³ The authors utilized a coupled cluster based scheme that includes core-valence contributions and contributions from full triples, and perturbative quadruples, obtaining a mean absolute deviation (MAD) of 1.8 kcal·mol⁻¹, showing the utility of the higher order coupled cluster contributions for a subset of the molecules (EuO, YbF, and LuO).¹⁴³

Recent calculations utilizing the Ln54^{96,141} and An66 sets,¹⁸⁵ sets of 54 lanthanides and 66 actinide enthalpies of formation and dissociation energies, with experimental uncertainties of 5 kcal·mol⁻¹ or less, have assessed the utility of a number of density functionals for heavy element species. For lanthanide molecules, the typical errors for these thermodynamic properties are on the order of 23 kcal·mol⁻¹; the errors for the actinide molecules are on the order of 10 kcal·mol⁻¹. The smaller errors for actinides are attributed to the more covalent bonding nature of actinide species, as compared to lanthanides. However, the errors in either case are not ideal. A leading reason for large errors in lanthanide and actinide chemistry pertains to the fact that available functionals have not been parameterized for use with heavy elements. An additional drawback to the use of DFT is its difficulties in describing systems with degeneracies or near-degeneracies (multireference systems).¹⁸⁶ Even though approaches to multireference DFT have been developed, such as multiconfiguration pair-density functional theory (MC-PDFT), only recently have they begun to be considered for actinide complexes.¹⁸⁷ Most forms of DFT are less than ideal for the *f*-elements, as their partially filled valence shells often lead to multireference character. As well, as demonstrated in the Ln54 and An66 studies, the utility of the functionals deviates very substantially from molecule to molecule, depending upon lanthanide or actinide, ligand, and property, with differences that can be as large as 100 kcal·mol⁻¹ or more for the dissociation energy or enthalpy of formation, complicating the choice of functional. For example, the enthalpy of formation obtained for the UO₃ molecule in the An66 set is 243.8 kcal·mol⁻¹ when using the Vosko-Wilk-Nusair with Slater exchange (SVWN) functional and 141.3 kcal·mol⁻¹ when using the Becke-Half-and-Half Lee-Yang-Parr (BHLYP) functional.¹⁸⁵

High quality experimental data with small experimental error bars have been vital to gauging the utility of theoretical methods in earlier parts of the periodic table. For the heavy elements, not only are the numbers of studies limited but also some of the very best results have had large experimental uncertainties. For example, for diatomic lanthanide sulfides and selenides in particular, there are few experimental studies^{188–191} where the bond dissociation energies have been determined. The studies, which occurred in the 1960s and 1970s, resulted in errors of $\sim 15 \text{ kJ}\cdot\text{mol}^{-1}$ or 0.2 eV, which is near the value selected for lanthanide chemical accuracy ($5 \text{ kcal}\cdot\text{mol}^{-1}$).¹⁴¹ This term was proposed more recently, based upon the average experimental uncertainties in the determination of enthalpies of formation and bond energies for a set of 54 lanthanide energies, which were then utilized to gauge predictions by a number of density functional approaches.¹⁴¹ Such large experimental uncertainties, though somewhat useful (as some data are better than none), make it more difficult to assess computational methodologies.

Recently, new experiments have enabled unprecedented accuracy in the prediction of dissociation, with an estimated accuracy of $\sim 0.004 \text{ eV}$.¹⁹² The Morse group has utilized a predissociation-based two-photon ionization (R2PI) method⁷⁴ to obtain bond dissociation energies for these species for transition metal and inner transition metal (*f*-element) containing silicides,^{86,193,194} selenides,¹⁹⁵ and sulfides.¹⁹⁵ The R2PI method makes use of the spin-orbit and nonadiabatic couplings of the large density of electronic states near the ground separated atom limit that allow the molecules to predissociate rapidly when the bond dissociation energy is reached or exceeded. Predissociation in this case is molecular dissociation that occurs long before the separated atom limit is reached. It is the key phenomenon that allows bond dissociation energies of transition metal and inner-transition metal molecules to be measured with high precision. When a predissociation threshold is observed by a sharp drop in signal in the absorption spectrum of the molecule of interest, its value directly provides the BDE of the system. The predissociation technique is a significant breakthrough in the determination of accurate thermodynamic data.

When considering the often-prohibitive computational cost of *ab initio* electron correlation calculations on heavy element containing species, composite methods are uniquely suited to pro-

vide accurate results with respect to experimental energies and thermodynamic properties while lowering the computational cost compared to traditional *ab initio* methods. Among the most used composite methods are the Gaussian-n (Gn),^{48,49,77,196–198} Weizmann-n (Wn),^{51,52,78,199} Complete Basis Set (CBS-n),^{200–204} High accuracy extrapolated *ab initio* thermochemistry (HEAT),^{53,81,82,205} Feller-Peterson-Dixon (FPD),^{54,206,207} and our own correlation consistent Composite Approach (ccCA).^{55,208} ccCA was first constructed for use on main group species; for the main group and s-block metals, ccCA was shown to achieve chemical accuracy ($\pm 1 \text{ kcal}\cdot\text{mol}^{-1}$), on average.^{55,56,208,209} Later, the method was developed for use with 3d transition metals (ccCA-TM)^{91,210} achieving transition metal chemical accuracy ($\pm 3 \text{ kcal}\cdot\text{mol}^{-1}$), on average, for the prediction of over 200 transition metal enthalpies of formation. Recently, Welch *et. al.* introduced Super-ccCA (s-ccCA), a composite method capable of achieving transition metal dissociation energies of $< 1\text{-}2 \text{ kcal}\cdot\text{mol}^{-1}$ within those from R2PI experiments. s-ccCA utilizes contributions to the composite including higher level coupled cluster corrections (triple and quadruple excitations) and spin-orbit contribution/correction from a Breit-Pauli Hamiltonian.²¹¹ The theoretical predictions were compared to experimental R2PI results.^{86,193,194,212,213} Recently, the *f*-block *ab initio* correlation consistent composite approach (*f*-ccCA) has been applied to lanthanide complexes, namely, oxides and halides, which were also compared to experimental R2PI predissociation energy results.^{213,214}

A challenge, however, for single reference methodologies, is that they can fail for lanthanide chemistry, as a single reference determinant may not be able to describe low-lying excited states or may converge to the wrong ground state. As well, Hartree-Fock orbitals used as initial guess orbitals may not be able to properly describe the ground state and lead to erroneous convergence issues.^{215,216} Paired with composite methods, multiconfigurational approaches are typically necessary for lanthanide complexes, to obtain accurate thermochemical and spectroscopic properties. The detailed work of Ruedenberg *et. al.* on the nature of the chemical bond, localized orbitals, and on Full Optimized Reaction Spaces (FORS), led to the widely used complete active space self-consistent field (CASSCF) approach and demonstrated the importance of multireference approaches.]^{217–220} By using multireference approaches, correct energetic assessment of ground states and spin-orbit

contributions can be performed confidently and will be considered in this effort.

Herein, *f*-ccCA has been used for the determination of bond dissociation energies of lanthanide sulfides and selenides. The highly accurate bond dissociation energies from Morse provide an excellent gauge of the ability of ccCA to model this molecular property for these lanthanide species.

5.2 Computational details

Geometry optimizations were performed for each of the molecules Ln–S and Ln–Se (Ln = Pr, Nd, Sm, Eu, Gd, Tb, Er, and Lu) using two different methods: the nonlocal exchange-correlation Perdew and Wang (PW91) functional¹⁰² and coupled cluster with single, double, and perturbative triple excitations [CCSD(T)]. PW91 was selected for its demonstrated efficacy for transition metals in bond dissociation predictions and as guess orbitals, which are important to describe the ground state correctly and for obtaining accurate thermodynamic properties.²²¹ However, Hartree-Fock orbitals are used throughout *f*-ccCA; if there are significant differences between these orbitals and natural orbitals, the calculation of accurate thermodynamic properties can be difficult.^{215,216} For example, for the NdS complex studied herein, the ground state is composed of a linear combination of four determinants. CCSD(T) was utilized due to its overall utility in predicting bond dissociation. Two routes were utilized to incorporate relativistic effects. The first was the use of a third-order Douglas-Kroll-Hess Hamiltonian (DKH3) that was considered for PW91 and CCSD(T) and will be referred to PW91-DKH3 and CCSD(T)-DKH3, respectively, throughout the discussion.²²² For these calculations, the cc-pVTZ-DK3²²³ basis set for lanthanides, aug-cc-pV(T+d)Z-DK^{224,225} for sulfur, and aug-cc-pVTZ-DK²²⁶ for selenium were utilized. The other route was to use effective core potentials (ECP) that will be referred to CCSD(T)/ECP. For each lanthanide, an atomic natural orbital (ANO) basis set was used with the corresponding ECP28MWB^{154,227,228} pseudopotential, accounting for scalar relativistic effects. For sulfur and selenium, aug-cc-PV(T+d)Z and aug-cc-pVTZ-PP with a ten-electron ECP (ECP10MDF) were employed, respectively.²²⁹

A triple- ζ level basis set was utilized for the geometry optimization steps due to its utility with respect to experiment and computational cost.¹¹⁴ In prior work on LuF, for example, despite the significant gain in computational cost incurred by the use of a quadruple- ζ level basis set, the bond

length did not improve, and even the triple- ζ level basis set led to a CCSD(T) optimized structure that was within 0.002 Å of experiment.¹¹⁴ In this work, frequencies were calculated to correct for zero-point energies (ZPEs), which were also determined at the triple- ζ level. For the CCSD(T) geometry optimizations, restricted Hartree-Fock (RHF) orbitals were generated and then used as guess orbitals for unrestricted CCSD(T) calculation (for the open-shell systems) and restricted CCSD(T) (closed-shell systems). The calculations were performed with MOLPRO 2020.¹⁴⁷ In MOLPRO 2020, the abelian point group, C_{2v} was utilized, as the full point group symmetry (C_∞) is not available for diatomics.

In the second part of this work, the f -ccCA ab initio composite scheme, recently introduced by Welch *et. al.*, was used to obtain the dissociation energy (D_0) for each of the lanthanide compounds.⁶² The methodology has been introduced in section 2.9. For the multireference wavefunction calculations, the $4f(f_{yz^2}, f_{xyz}, f_{y(3x^2y^2)}, f_{xz^2}, f_{z(xy^2)}, f_{x(x^2zy^2)}$ and $f_{z^3})$ and $6s$ hybrid orbitals of the metal were included in the CASSCF active space, except for gadolinium and lutetium. For the former, the $5d(d_{x^2y^2}, d_{z^2}, d_{xy}, d_{xz}, d_{yz})$ hybrid orbitals were included in conjunction with the $4f$ and $6s$ orbitals. For lutetium, only the $5d$ and $6s$ hybrid orbitals were included in the active space because the $4f$ set of hybrid orbitals are occupied by 14 electrons and, thus, do not have a large contribution toward spin-orbit coupling. The $3p_{x,y,z}$ and $4p_{x,y,z}$ sets of orbitals from sulfur and selenium are doubly occupied at the equilibrium bond length and were not included in the CASSCF active space. For Multireference Configuration Interaction (MRCI), the same active space that was utilized for the CASSCF calculations was chosen. In addition, the $3p$ and $4p$ orbitals for sulfur and selenium, respectively, were also considered. The inclusion of these p orbitals allows for single and double excitations from these orbitals to the active space in the MRCI calculations (orbitals were included at the core level in the MOLPRO implementation).¹⁴⁷ The spin-orbit coupling energies were calculated using the Breit-Pauli Hamiltonian and the MRCI wavefunction. The ground and first excited states were included in the state-averaged CASSCF calculations. Then, the CI vectors and energetics for MRCI were utilized to gauge potential multireference character in the diatomic by comparing the weight of configurations. As well, for all of the molecules, the multireference

character was assessed by determining the T_1 and D_1 diagnostics,^{230,231} coupled cluster singles and doubles amplitudes ($|T_{1\max}|$ and $|T_{2\max}|$, respectively),²³² and spin contamination at the triple- ζ level.

For several species (TbS, TbSe, ErS, and ErSe), CCSD(T)-DKH3 energies were determined at the double-, triple-, and quadruple- ζ basis set levels and were extrapolated to the CBS limit using eq. 2.40. Two different guess orbitals (RHF and RPW91) were utilized for CCSD(T).

For ErSe, a functional from each of four density functional families – local-density approximation (LDA), generalized gradient approximation (GGA), meta-GGA (M-GGA), and hybrid M-GGA – was employed for bond dissociation energy predictions. The corresponding functionals were PW91 (LDA), Tao, Perdew, Staroverov, Scuseria (TPSS, GGA),⁹⁵ Minnesota 2006 local functional (M06-L, M-GGA),¹⁰⁴ and the Becke, 3-parameter, Lee-Yang-Parr (B3LYP, hybrid M-GGA)^{15,16} that were utilized at a restricted level (RKS). These DFT calculations were performed at a double-, triple- and quadruple- ζ level.

5.3 Results and discussion of results

Geometry optimization of sulfides and selenides

In Table 5.1, the optimized bond lengths and vibrational frequencies determined for the PW91 and CCSD(T) calculations for the sulfide complexes (Pr, Nd, Sm, Eu, Gd, Tb, Er, and Lu) are provided. In terms of bond lengths, though a contraction in bond length is expected while the f shell is being filled, due to the complexity of the electronic structure of lanthanide chemistry, this was not observed for the sulfide diatomics.¹⁸⁴ In Table 5.1, CCSD(T) bond lengths for both one- and two-component Hamiltonians are longer for PrS to GdS, and LuS than for the corresponding PW91 optimized structures. The difference in the bond length for lanthanide sulfides comparing PW91 and CCSD(T) ranges from ~ 0.01 Å (GdS and LuS) to ~ 0.05 Å for EuS. NdS has a 0.02-0.03 Å shorter bond length than the DFT using the zeroth-order regular approximation (ZORA) at triple- ζ bond length given by Xu *et. al.*¹⁸⁴ This difference is not surprising, as NdS has a significant multireference character, requiring four CI vectors to describe the ground state Table 5.7). For RHF and CCSD(T),

the lowest electronic configuration corresponds to having unpaired electrons in the $4f_{z^3}$, $4f_{z(x^2y^2)}$, and $6s$ orbitals and another unpaired electron in the $4f_{x(x^2-3y^2)}$ or $4f_{y(3x^2-y^2)}$ orbital (Table 5.7). Even though this configuration is an excited state at the CAS/MRCI level, it is the ground state electronic structure for RHF and UCCSD(T).

For ErS and TbS complexes, their bond lengths are larger with PW91 than with CCSD(T). The vibrational frequencies for CCSD(T) with a two-component Hamiltonian are 60.44 and 83.31 cm^{-1} larger than the PW91-DKH3 frequencies for TbS and ErS, respectively. For the CCSD(T) calculations with a one-component Hamiltonian and an ECP basis set, there was no convergence for the SCF procedure for TbS and ErS (which is discussed in Sec. 5.3).

Overall, the difference in the bond length from one- and two-component Hamiltonian predictions was not very significant; on average, CCSD(T)-DKH3 bond lengths are ~ 0.01 Å longer than for CCSD(T)/ECP for all complexes. The results herein are on par with DFT, SCF, and configuration interaction with singles and doubles (CISD) bond lengths predicted in prior studies and given in Table 5.1. However, in comparing theoretical and experimental bond lengths, there are some substantial differences. For example, for EuS and GdS, the bond length variance between CCSD(T)-DKH3 and experiment is large, ~ 0.08 Å for EuS and ~ 0.16 or 0.08 Å for GdS, depending upon experiment used for comparison. The substantial difference in the bond length can be attributed to the estimated experimental bond lengths, rather than having direct measurements. For lanthanide sulfides, a one component Hamiltonian CCSD(T) with a robust ECP basis set can recover most of the electron correlation needed.

Table 5.1: DFT (PW91-DKH3), CCSD(T)/ECP, and CCSD(T)-DKH3 optimized bond lengths (Å) and harmonic vibrational frequencies (cm⁻¹) (in parentheses) of sulfide complexes using a triple- ζ level basis set.

Molecules	PW91-DKH3	CCSD(T)/ECP	CCSD(T)-DKH3	Previous theoretical results	Expt.
PrS	2.3154 (450.81)	2.3291 (459.21)	2.3310 (455.43)	2.347 ^{184,a} (432.7) ^{184,a} 2.338 ^{184,b}	—
NdS	2.3258 (428.93)	2.3251 (459.44)	2.3289 (-56.73)	2.345 ^{184,a} (447.5) 2.340 ^{184,b}	—
SmS	2.4186 (393.14)	2.4388 (375.16)	2.4378 (380.34)	2.414 ^{184,a} -390.6 2.334 ^{184,b}	—
EuS	2.3879 (384.25)	2.4315 (376.5)	2.4373 (374.56)	2.41 ^{170,c} (362 ^{170,c}) 2.396 ^{184,a} (383.5) ^{184,a} 2.343 ^{184,b}	2.51 ^{233,d} (400) ^{233,e}
GdS	2.2945 (436.62)	2.3019 (457.65)	2.3119 (453.89)	2.31, 2.29 ^{170,c} (431 ^{170,c}) 2.272-2.342 ^{180,f} (412-493) ^{180,f}	2.15 ^{233,d} 2.23 ^{233,d} (479) ^{233,d}
TbS	2.3309 (398.22)		2.3033 (458.66)		
ErS	2.3329 (372.70)		2.2834 (456.01)		
LuS	2.2572 (448.17)	2.2611 (452.87)	2.2762 (447.03)	2.27 ^{170,c} (456 ^{170,c})	2.17 ^{233,d} 2.10 ^{233,d} (500) ^{233,d}

^a FON-DFT calculation using ZORA Hamiltonian at a triple- ζ level.

^b FON-DFT calculation using spin-orbit coupled calculations.

^c DFT calculations using the BP, BPW and BLYP functional at the triple- ζ level.

^d Estimated from empirical relations.

^e Estimated based on experiments.

^f All electron DFT (LDASIC, B88, and B88P86) calculations and ECP calculation with SCF, CISD, and ACPF at the ANO triple- ζ level.

In Table 5.2, the optimized geometry for the selenide complexes (Pr, Nd, Sm, Eu, Gd, Tb, Er, and Lu) are provided. For the one-component Hamiltonian predictions, selenide complexes have ten more electrons than the sulfides, which were treated with an ECP. The shorter bond distances obtained in the calculations relative to experiment for sulfide complexes also occur for selenides, but it is not as pronounced. The difference in bond lengths between PW91 and CCSD(T) is quite large for SmSe, the difference is ~ 0.14 and ~ 0.15 Å, for one- and two-component Hamiltonian predictions, respectively. For the other seven selenium complexes, the bond lengths differ on average by ~ 0.01 Å between PW91 to CCSD(T), except for EuSe, which is 0.04 Å. NdSe has a shorter bond length than the fractional occupation number DFT (FON-DFT) results using the ZORA Hamiltonian (0.04-0.05 Å). The four different CI vectors needed to describe the ground state determinant make NdSe a multireference system (see Table 5.7), akin to NdS. For TbSe and ErSe, as for TbS and ErS, the bond length is larger for PW91 than for CCSD(T), and also, the ground state vibrational frequencies are smaller for PW91. In addition, the CCSD(T)/ECP optimization was also not possible due to the non-convergence of the SCF procedures. Overall, the equilibrium geometries are quite similar when comparing DFT (PW91) to CCSD(T). Comparing the CCSD(T) one- and two-component Hamiltonian bond length predictions, the difference is quite small, less than 0.01 Å for most complexes, and their vibrational frequencies differ by a maximum of ~ 5 cm⁻¹. The impact on the BDE predictions among the different geometries for all complexes is always less than 0.8 kcal·mol⁻¹.

Table 5.2: DFT (PW91-DKH3), CCSD(T)/ECP, and CCSD(T)-DKH3 optimized bond lengths (Å) and harmonic vibrational frequencies (cm⁻¹) (in parentheses) of selenide complexes using a triple- ζ level basis set.

Molecules	PW91-DKH3	CCSD(T)/ECP	CCSD(T)-DKH3	Previous theoretical predictions ^{184,a}
PrSe	2.4610 (-300.15)	2.473 (-308.4)	2.4781 (-305.67)	2.504 (-284.8) 2.497
NdSe	2.4745 (-282.50)	2.4669 (-307.38)	2.4738 (-306.60)	2.519 (-255.6) 2.523
SmSe	2.4347 (-259.35)	2.5761 (-252.27)	2.5831 (-253.10)	2.508
EuSe	2.5325 (-252.01)	2.5457 (-248.01)	2.5804 (-248.27)	2.552 (-250.9) 2.502
GdSe	2.4236 (-287.53)	2.439 (-302.68)	2.4514 (-300.44)	—
TbSe	2.4526 (-263.78)	—	2.4426 (-301.41)	—
ErSe	2.4852 (-238.22)	—	2.4182 (-295.43)	—
LuSe	2.3913 (-288.66)	2.3901 (-292.76)	2.4059 (-290.31)	—

^a FON-DFT calculation using ZORA Hamiltonian at a triple- ζ level.

Bond dissociation energies for lanthanide sulfides and selenides

In this section, using the three different geometries described above, *f*-ccCA is employed to calculate BDEs for eight sulfide and selenide complexes, and the BDEs are compared to experiment. The molecules are separated into three categories, depending upon the error in the *f*-ccCA dissociation energy predictions relative to Morses experimental data (see Ref. ¹⁹⁵): Cat I (± 2 kcal·mol⁻¹), Cat II (± 6 kcal·mol⁻¹), and Cat III (> 6 kcal·mol⁻¹). These differences are presented in Figure 5.1. In Table 5.3, a summary of the *f*-ccCA dissociation energies determined at each different optimized geometry is provided and compared with experiment. Moreover, in Tables 5.4, C.1, and C.2, the total atomization terms and dissociation energy predictions for CCSD(T)-DKH3, PW91-DKH3, and CCSD(T)/ECP geometries are shown.

The calculated dissociation energies (D_0) for sulfide and selenide complexes are decomposed into different contributions that comprise the *f*-ccCA composite, as described in Sec. 2.9, such as the reference energy ($\Delta MP2_{CBS}$ and ΔHF_{CBS}), core-valence (ΔCV), correlation contribution (ΔCC), and spin-orbit (ΔSO) contributions. Table 5.5 provides TbS, ErS, TbSe, and ErSe bond dissociation energies calculated with two different approaches: RHF and DFT (PW91) orbitals. In Table 5.6, dissociation energies considering different families of DFT are used to predict the most

problematic lanthanide complex: ErSe. Table 5.7 contains the corresponding CI vectors for the spin-orbit correction for each of the sulfide and selenide complexes. In Table C.4, multireference diagnostics are given for all sulfide and selenide complexes at CCSD(T)-DKH3 geometries. For PW91-DKH3 and CCSD(T)/ECP geometries, see Tables C.3 and 5.4 in the appendix C.

For the sulfide complexes, *f*-ccCA was quite successful for PrS, SmS, GdS, and LuS, resulting in differences of 1-2 kcal·mol⁻¹ from experiment for each of the three investigated geometries [PW91-DKH3, CCSD(T)/ECP, and CCSD(T)-DKH3]. These four complexes are part of Cat I. The ground state of these four complexes has been identified as PrS (⁴H), SmS (⁷Δ), GdS (⁹Σ), and LuS(²Σ⁺). The ΔMP2_{CBS} term for these four complexes ranges between 37 and 51 kcal·mol⁻¹ and the core-valence (CV) term contributes 1-2 kcal·mol⁻¹ to the total energy. The correlation term (CC) ranges from -4.75 (SmS) to 6.48 kcal·mol⁻¹ (PrS) considering CCSD(T)-DKH3 geometries (Table 5.4). The spin-orbit correction obtained with the Breit-Pauli Hamiltonian can render a large difference in the final BDEs, with contributions that can be as large as -3.91 kcal·mol⁻¹, as for LuS. Even though the molecule has small spin-orbit coupling, the individual atoms account for this large contribution. The choice of the method for geometry optimization [PW91-DKH3, CCSD(T)-DKH3, and CCSD(T)/ECP] has very little impact on the energy, each resulting in a BDE for LuS within ~0.01 kcal·mol⁻¹ from experiment (Table 5.3).

Table 5.3: f -ccCA dissociation energy for each geometry optimization method and corresponding experimental data in kcal·mol⁻¹.

Molecules	PW91-DKH3	CCSD(T)/ECP	CCSD(T)-DKH3	Exp. D ₀ ¹⁹⁵	Exp. D ₀
Category I					
PrS	118.69	118.59	118.57	120.61	112 ± 6 ^a
PrSe	102.35	102.29	102.24	103.68	—
SmS	92.70	92.63	92.65	92.52	—
SmSe	78.66	81.41	81.38	80.62	—
GdS	121.06	121.04	120.95	121.80	124 ± 6 ^a 125 ± 4 ^b
GdSe	105.77	105.74	105.63	106.20	102 ± 5 ^a 103 ± 4 ^c
LuS	119.69	119.62	119.62	119.70	120 ± 6 ^a 120 ± 4 ^b
LuSe	106.51	106.51	106.39	106.10	99 ± 6 ^a 100 ± 4 ^c
Category II					
NdS	116.73	116.53	116.50	111.15	120 ± 6 ^a 112 ± 4 ^b
NdSe	99.21	99.28	99.35	94.53	91 ± 6 ^a 91 ± 4 ^c
EuS	92.35	92.43	92.42	87.90	86 ± 6 ^a 86 ± 4 ^b
EuSe	82.34	82.4	82.32	76.50	66 ± 6 ^a 72 ± 4 ^c
Category III					
TbS	132.13	—	132.87	122.00	—
TbSe	92.02	—	92.47	106.10	—
ErS	89.01	—	89.94	98.20	99 ± 4 ^b
ErSe	-100.29	—	-100.92	82.70	—

^a Reference.¹⁷⁶

^b Reference.¹⁸⁹

^c Reference.¹⁹⁰

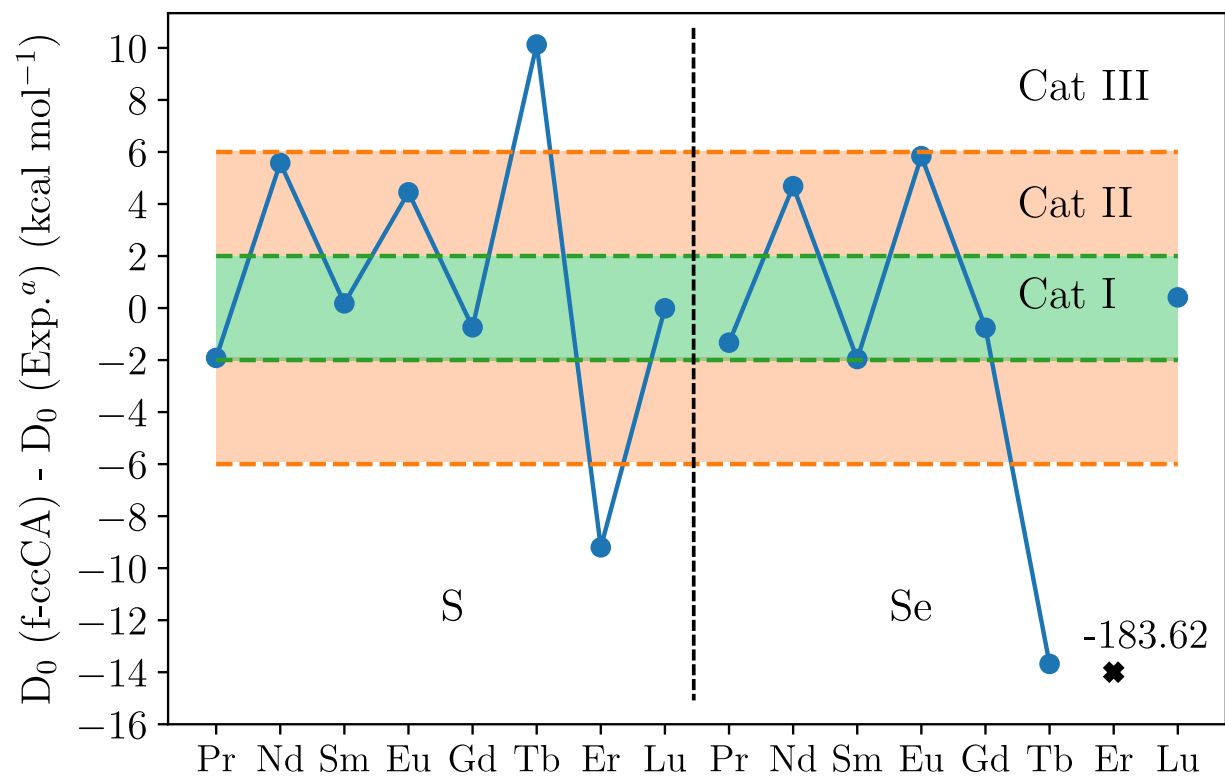


Figure 5.1: *f*-ccCA BDE divided in three categories based on their difference from experiment^{a, 195}: Cat I ($\pm 2 \text{ kcal mol}^{-1}$), Cat II ($\pm 6 \text{ kcal mol}^{-1}$), and Cat III (more than 6 kcal mol^{-1}).

Table 5.4: Total atomization terms and dissociation energy obtained with f -ccCA using CCSD(T)-DKH3 geometries, along with all electron basis sets. Theoretical and experimental dissociation energies are in kcal·mol⁻¹.

	Δ MP2 _{CBS}	Δ HF _{CBS}	Δ CV	Δ CC	Δ SO	f -ccCA D ₀	Exp. D ₀ ¹⁹⁵	Exp. D ₀
Category I								
PrS	44.69	69.70	1.65	6.48	-2.66	118.57	120.61	112 ± 6 ^a
PrSe	39.44	56.99	2.66	8.83	-4.82	102.24	103.68	-
SmS	50.36	47.06	-0.64	-4.75	1.17	92.65	92.52	-
SmSe	43.52	41.31	0.62	-2.39	-1.32	81.38	80.62	-
GdS	48.71	79.69	0.61	-4.33	-3.08	120.95	121.80	124 ± 6 ^a 125 ± 4 ^b
GdSe	43.53	68.33	1.49	-2.06	-5.22	105.63	106.20	102 ± 5 ^a 103 ± 4 ^c
LuS	37.74	88.45	1.22	-3.22	-3.91	119.62	119.70	120 ± 6 ^a 120 ± 4 ^b
LuSe	36.45	77.54	0.63	-1.69	-6.12	106.39	106.10	99 ± 6 ^a 100 ± 4 ^c
Category II								
NdS	29.26	88.39	1.77	-1.47	-0.22	116.50	111.15	120 ± 6 ^a 112 ± 4 ^b
NdSe	24.10	76.28	2.83	0.72	-3.84	99.35	94.53	91 ± 6 ^a 91 ± 4 ^c
EuS	49.58	43.52	0.36	-4.95	4.44	92.42	87.90	86 ± 6 ^a 86 ± 4 ^b
EuSe	42.89	38.19	1.38	-2.54	2.77	82.32	76.50	66 ± 6 ^a 72 ± 4 ^c
Category III								
TbS	-29.85	155.87	-0.63	8.78	-0.71	132.87	122.00	-
TbSe	-35.14	144.32	-24.54	11.07	-2.86	92.47	106.10	-
ErS	-67.29	130.88	-0.92	27.66	0.17	89.94	98.20	99 ± 4 ^b
ErSe	-183.70	9.72	32.65	41.80	-1.06	-100.92	82.70	-

^a Reference.¹⁷⁶

^b Reference.¹⁸⁹

^c Reference.¹⁹⁰

For NdS (⁵I) (Cat II), the BDE is ~5 kcal·mol⁻¹ from experiment when different optimized

geometries are considered (shown in Tables 5.4, C.1, and C.2). When investigating the ground state of the neodymium atom (Nd), the electronic structure is more complex, which is a hurdle while treating these molecules with single reference methods. The neodymium ground state is a ^5I , resulting in 13 ways of constructing the ground state.²³⁴ However, at the Hartree-Fock level, all electronic configurations are not all balanced in the same way; they are not all degenerate. For neodymium, the lowest energy at Hartree-Fock was obtained by placing two unpaired electrons in the same symmetry, and the other two in two different symmetries. In addition, in Table 5.7, for NdS, there are four equally important CI coefficients (0.50), resulting in a complex ground state. In Table C.4, the D_1 value is 0.10 and the $|T_{1\text{max}}|$ for this complex is 0.13, which is one of the highest among the studied complexes. The reference value for D_1 and $|T_{1\text{max}}|$ for transition metals is 0.15 and 0.05, respectively.²³⁵ The multireference ground state of NdS, i.e., for large CI contributions (see Table 5.7), along with a bad set of guess orbitals for the complex led to predictions that are ~ 5 kcal·mol⁻¹ from experiment.

For EuS ($^8\Sigma$) (Cat II), the spin-orbit contribution from the ground state of the complex is large [4.44 kcal·mol⁻¹ at the CCSD(T)-DKH3 geometry], although, for europium and sulfur atoms, the spin-orbit splitting is zero and close to zero, respectively. The largest CI coefficient for EuS is 0.99 (Table 5.7), and its ground state is a $^8\Sigma$; here, a difference of 5 kcal·mol⁻¹ from experiment in the BDE was obtained. At the CAS/MRCI level, a $4f^6 6s^1$ electronic configuration is predicted (see Table 5.7); while at RHF/UCCSD(T), a $4f^7 6s^0$ configuration is obtained. Single reference methods cannot predict the correct configuration for the ground state, which leads to a larger difference when compared to experiment. Similarly, for SmS, CASSCF predicts a ground state that has one electron in the 6s orbital ($4f^5 6s^1$), rather than in a $4f^6$ configuration. In addition, at HF/CCSD(T), the ground state converges to a $4f^6$ electronic configuration. However, for SmS, the impact of the electron configuration in the final predicted BDE was small when compared to experiment.

The last two molecules TbS ($^8\Phi$) and ErS ($^5\Delta$), which are part of Cat III, are the most problematic complexes to treat with single reference methodologies. TbS primarily has one large contribution to the ground state, but the Tb (^6H) atom is quite multireference in nature. ErS has five main CI

contributions to the ground state (see Table 5.7), which make it quite complex to address using single reference methods. In addition, in Table C.4, these two complexes have two of the largest $|T_{1\max}|$ amplitudes, which can generate issues for Hartree-Fock initial orbital guesses resulting from multireference character. For TbS and ErS, the $\Delta\text{MP2}_{\text{CBS}}$ term is negative [30.65 and 68.77 kcal·mol⁻¹ for PW91-DKH3 and 29.85 and 67.29 kcal·mol⁻¹ for CCSD(T)-DKH3, respectively], which means that according to MP2, the molecular complexes do not form, i.e., the atoms are more stable than the complexes. The coupled cluster correction, along with HF/CBS extrapolation make up for the negative MP2 energetics, and the final *f*-ccCA values are ~10 and 9 kcal·mol⁻¹ from experimental energies for TbS and ErS, respectively. Since the *f*-ccCA procedure is not very accurate for these complexes likely due to their multireference nature, other approaches to predict bond dissociation energies were taken.

In Table 5.5, the CCSD(T)/CBS energies obtained at the CCSD(T)-DKH3 and PW91-DKH3 geometries are provided. For TbS, UCCSD(T)/CBS resulted in BDEs that are further from experiment than the *f*-ccCA BDEs. For example, UCCSD(T)/CBS using CCSD(T)-DK3 geometry renders a BDE of 137.0 kcal·mol⁻¹, *f*-ccCA with the same geometry renders a BDE of 132.87 kcal·mol⁻¹, and the experimental BDE is 122.0 kcal·mol⁻¹. However, for ErS, the BDE was ~6 kcal·mol⁻¹ from experiment. Since Hartree-Fock provides a poor initial guess for these complexes, DFT orbitals were considered for CCSD(T), probing their effect on dissociation energy predictions. For TbS, DFT orbitals are a better guess for the complex; however, they are not for ErS. Our best estimate for TbS is 4 kcal·mol⁻¹ from experiment using PW91 orbitals, but for ErS, CCSD(T) using RHF orbitals provides the most accurate result [92.5 kcal·mol⁻¹ obtained using UCCSD(T)/CBS vs 98.2 kcal·mol⁻¹ from experiment].

Table 5.5: Dissociation energies at CBS in kcal·mol⁻¹, with CCSD(T), paired with double-, triple-, and quadruple- ζ basis sets and two different geometries.

	Geometry	CBS ^a	CBS ^a	Exp. ¹⁹⁵	Exp. ¹⁹³
TbS	CCSD(T)-DKH3	137.0	126.8	122.0	—
	PW91-DKH3	136.6	126.4		
TbSe	CCSD(T)-DKH3	124.7	110.7	106.1	—
	PW91-DKH3	124.6	110.6		
ErS	CCSD(T)-DKH3	92.5	81.7	98.2	99 $\pm$ 4
	PW91-DKH3	91.5	80.7		
ErSe	CCSD(T)-DKH3	-125.8	-2.2	82.7	—
	PW91-DKH3	-129.8	no convergence		

^a Using restricted Hartree-Fock as an initial guess for the orbitals.

^b Using restricted PW91 as an initial guess for the orbitals.

The lanthanide selenides also proved to be challenging for single reference wavefunction-based methodologies. As seen for the sulfide complexes, the Cat I molecules PrSe (⁴H), SmSe (⁷ Δ), GdSe (⁹ Σ^+), and LuSe (² Σ^+) resulted in bond dissociation energies within 1-2 kcal·mol⁻¹ when compared to the experimental values determined by Sorensen *et. al.*¹⁹⁵ Furthermore, PrSe has the largest $|T_{2\max}|$ in Table C.4, which is within the threshold of multireference complexes ($|T_{2\max}| > 0.15$).²³⁵ However, the predicted dissociation energies are still accurate compared to experiment when calculated using single reference methodologies. In Tables 5.4 and C.1, the BDEs for PrSe obtained using the PW91-DKH3 and CCSD(T)/DKH3 geometries and the *f*-ccCA composite method are 102.35 and 102.24 kcal·mol⁻¹, respectively, compared to the experimentally determined BDE of 103.68 kcal·mol⁻¹. Regarding the other complexes of Cat I, GdSe and LuSe are indicated to be of a single reference character based upon the diagnostics in Table C.4. In addition, they only have one main CI coefficient (see Table 5.7), so are expected to perform better with single reference methodologies.

NdSe (Cat II), which has a ⁵I ground state and has the same four CI coefficients as NdS, has a ground state with significant multireference character. It also has a large $|T_{1\max}|$ value of 0.13 and D_1 of 0.10 (Table C.4). Due to the complexity of the electronic structure of this complex and

the mixing of states, the spin-orbit contribution of this complex is quite large [$3.84 \text{ kcal}\cdot\text{mol}^{-1}$ at the CCSD(T)-DKH3 geometry]. The dissociation energies determined at each of the optimized geometries, are $\sim 5 \text{ kcal}\cdot\text{mol}^{-1}$ from experiment.¹⁹⁵ For EuSe ($^8\Sigma$) (Cat II), similar observations to those made for EuS are made for the BDE and spin-orbit contribution: the BDE is $\sim 6 \text{ kcal}\cdot\text{mol}^{-1}$ from experiment, with a large spin-orbit contribution of $2.77 \text{ kcal}\cdot\text{mol}^{-1}$. Its ground state also converges to a $4f^7$ configuration as was seen for EuS, while at the CAS/MRCI level, a $4f^6 6s^1$ electronic configuration is observed. For SmSe, the $4f^5 6s^1$ configuration is also more stable with a multireference wavefunction (CASSCF) than a $4f^6$ configuration.

Table 5.6: Erbium selenide dissociation energies ($\text{kcal}\cdot\text{mol}^{-1}$) using CCSD(T)-DKH3 geometries, paired with double-, triple-, and quadruple- ζ basis sets, considering four families of DFT functionals, ZPVE (CCSD(T)-DKH3 geometry), and spin-orbit corrected.

DFT functional	Double- ζ	Triple- ζ	Quadruple- ζ	Exp. ¹⁹⁵
TPSS	245.3	252.5	193.3	82.7
B3LYP	86.3	92.0	70.8	
M06-L	134.5	126.4	81.0	
PW91	234.0	217.2	209.2	

Focusing on the Cat III compounds, TbSe and ErSe, the same observations made for TbS and ErS are again seen here. The $\Delta\text{MP2}_{\text{CBS}}$ term is negative for both complexes. For ErSe, values of 181.13 and $183.70 \text{ kcal}\cdot\text{mol}^{-1}$ are obtained for the PW91-DKH3 and CCSD(T)-DKH3 optimized structures, respectively. For TbSe, a less negative term is observed [$\sim 35 \text{ kcal}\cdot\text{mol}^{-1}$ for both PW91-DKH3 and CCSD(T)-DKH3], but according to MP2, the atoms are more stable than the complex. For ErSe, not only does MP2 fail for this complex, but in contrast to what happens for the sulfides, this large difference is not corrected by the CCSD(T) calculation (correlation term) with the composite, so large negative values for the dissociation energy are obtained at f -ccCA level. For TbSe, $\Delta\text{HF}_{\text{CBS}}$ and Δ_{CC} terms show that the complex forms, but it is due to error cancellations. On the contrary to their sulfide counterpart, for ErSe, the $\Delta\text{HF}_{\text{CBS}}$ and Δ_{CC} terms are not enough to compensate for the huge negative $\Delta\text{MP2}_{\text{CBS}}$ term, leading to the negative dissociation energies, which means at

f-ccCA level, ErSe does not form. The largest $|T_{1\max}|$ and D_1 for this complex are 0.25 and 0.18, which correspond to the largest values for these diagnostics found in the appendix (Table C.4). In Table 5.5, CCSD(T)/CBS energies are given. The dissociation energy predictions [obtained using the PW91-DKH3 and CCSD(T)-DHK3 geometries] for TbSe and ErSe are quite far away from experiment and, for ErSe, the complex does not form according to CCSD(T)/CBS. It is clear from this Table how poor the restricted Hartree-Fock guess orbitals are for these complexes. In Table 5.5, for TbSe and ErSe, DFT orbitals are also used as guesses for CCSD(T). For TbSe, these sets of orbitals prove to be of value and are better guesses for CCSD(T), which put the dissociation energy extrapolated at CCSD(T)/CBS at 4 kcal·mol⁻¹ from experiment. However, for ErSe, even with DFT orbitals, the dissociation energy is still negative. The use of a multireference wavefunction could help in determining the bond dissociation energy. The ground state of Er is a ³H, which leads to a large number of states generated and can complicate the determination of the complete dissociation channel [Er(³H) + S(³P)]. In addition, even MRCI+Q is not a size extensive method, which carries an additional layer of complexity to these calculations. Since ErSe is such a difficult complex from an electronic structure perspective, DFT was probed and different functionals were considered for calculating dissociation energies for ErSe. DFT offers a computationally less costly theoretical approach than ab initio correlated methods that can lead to an easy comparison to experimental values. In Table 5.6, four families of DFT were considered using one representing functional from each family and three different levels of basis sets. The meta-GGA functional, M06-L, obtained the closest results to experiment at a quadruple- ζ level. However, fortuitous error cancellation is the most likely cause for these predictions because, when comparing double- and triple- ζ results, a consistent trend is not present. TPSS and PW91 completely fail predicting the dissociation energies. Finally, for the hybrid functional, B3LYP, there is a sudden drop in dissociation energy from the triple- ζ to quadruple- ζ levels. As demonstrated in Ref.⁹⁶, the performance of different density functionals for lanthanide containing species can be erratic at best, and this can be magnified with increasing or decreasing basis set size.¹⁸⁹

Table 5.7: CI vectors from CASSCF calculations for the equilibrium bond lengths of sulfide and selenide complexes. 0, α , 2, and "—" correspond to orbital occupations: zero, one (α -spin), two (doubly occupied), and not included in the active space electrons, respectively.

Molecules	Coeff.	$4f_{yz^2}$	$4f_{xyz}$	$4f_{y(3x^2-y^2)}$	$4f_{xz^2}$	$4f_{z(x^2-y^2)}$	$4f_{x(x^2-3y^2)}$	$4f_{z^3}$	$6s$	$n^a p_x$	$n^a p_y$	$n^a p_z$	$5d_{zy}$	$5d_{xy}$	$5d_{yx}$	$5d_{x^2-y^2}$	$5d_z^2$
PrS (4H)	0.71	0	0	0	0	α	α	0	α	2	2	2	—	—	—	—	—
	-0.71	0	α	α	0	0	0	0	α	2	2	2	—	—	—	—	—
NdS (5I)	0.50	α	α	α	0	0	0	0	α	2	2	2	—	—	—	—	—
	-0.50	0	α	0	α	0	α	0	α	2	2	2	—	—	—	—	—
	0.50	0	0	0	α	α	α	0	α	2	2	2	—	—	—	—	—
	-0.50	0	0	α	α	α	0	0	α	2	2	2	—	—	—	—	—
SmS (7A)	0.87	0	α	α	0	α	α	α	α	2	2	2	—	—	—	—	—
	0.49	α	0	α	α	0	α	α	α	2	2	2	—	—	—	—	—
GdS ($^9\Sigma^-$)	0.99	α	α	α	α	α	α	α	α	2	2	2	0	0	0	0	0
EuS ($^8\Sigma^-$)	0.99	α	α	α	α	α	α	0	α	2	2	2	—	—	—	—	—
TbS ($^8\Phi$)	1.00	α	α	α	α	2	α	α	α	2	2	2	—	—	—	—	—
ErS (5A)	0.55	2	2	2	α	2	α	α	α	2	2	2	—	—	—	—	—
	-0.34	α	2	2	2	α	α	2	α	2	2	2	—	—	—	—	—
	-0.34	α	2	α	α	2	2	2	α	2	2	2	—	—	—	—	—
	-0.32	2	α	2	2	α	α	2	α	2	2	2	—	—	—	—	—
	0.32	2	α	α	α	2	2	2	α	2	2	2	—	—	—	—	—
	-0.32	2	2	2	α	α	2	α	α	2	2	2	—	—	—	—	—
	0.32	2	2	α	2	2	α	α	α	2	2	2	—	—	—	—	—
LuS ($^2\Sigma^+$)	-0.96	2	2	2	2	2	2	2	α	2	2	2	0	0	0	0	0

^a Correspond to $3p$ for sulfur and $4p$ for selenium.

Table 5.7: Table (continued)

Molecules	Coeff.	$4f_{yz^2}$	$4f_{xyz}$	$4f_{y(3x^2-y^2)}$	$4f_{xz^2}$	$4f_{z(x^2-y^2)}$	$4f_{x(x^2-3y^2)}$	$4f_{z^3}$	6s	$n^a p_x$	$n^a p_y$	$n^a p_z$	$5d_{zy}$	$5d_{xy}$	$5d_{yx}$	$5d_{x^2-y^2}$	$5d_z^2$
PrSe (4H)	0.71	0	0	α	0	α	0	0	α	2	2	2	—	—	—	—	—
	0.71	0	0	0	0	α	α	0	α	2	2	2	—	—	—	—	—
NdSe (5I)	0.50	α	α	α	0	0	0	0	α	2	2	2	—	—	—	—	—
	-0.50	0	α	0	α	0	α	0	α	2	2	2	—	—	—	—	—
	0.50	0	0	α	α	α	0	0	α	2	2	2	—	—	—	—	—
	-0.50	α	0	0	0	α	α	0	α	2	2	2	—	—	—	—	—
SmSe ($^7\Delta$)	0.86	0	α	α	0	α	α	α	α	2	2	2	—	—	—	—	—
	0.50	α	0	α	α	0	α	α	α	2	2	2	—	—	—	—	—
EuSe ($^8\Sigma^-$)	0.99	α	α	α	α	α	α	0	α	2	2	2	—	—	—	—	—
GdSe ($^9\Sigma^-$)	1.00	α	α	α	α	α	α	α	α	2	2	2	0	0	0	0	0
TbSe ($^8\Phi$)	1.00	α	α	α	α	2	α	α	α	2	2	2	—	—	—	—	—
ErSe ($^5\Delta$)	0.55	2	2	2	α	2	α	α	α	2	2	2	—	—	—	—	—
	-0.34	α	2	2	2	α	α	2	α	2	2	2	—	—	—	—	—
	-0.34	α	2	α	α	2	2	2	α	2	2	2	—	—	—	—	—
	-0.32	2	α	2	2	α	α	2	α	2	2	2	—	—	—	—	—
	0.32	2	α	α	α	2	2	2	α	2	2	2	—	—	—	—	—
	-0.32	2	2	2	α	α	2	α	α	2	2	2	—	—	—	—	—
	0.32	2	2	α	2	2	α	α	α	2	2	2	—	—	—	—	—
LuSe ($^2\Sigma^+$)	0.94	2	2	2	2	2	2	2	α	2	2	2	0	0	0	0	0

^a Correspond to 3p for sulfur and 4p for selenium.

Table 5.8: Multireference considerations for lanthanide sulfides and selenides using CCSD(T)-DKH3 geometries. T_1 and D_1 are common diagnostics, $|T_{1\max}|$ and $|T_{2\max}|$ are the absolute values of the largest amplitudes, and S^2 is the spin contamination.

	T_1	D_1	$ T_{1\max} $	$ T_{2\max} $	Spin contamination (S^2)
PrS	0.03	0.11	0.12	0.2	0.004
NdS	0.03	0.10	0.13	0.11	0.003
SmS	0.03	0.09	0.07	—	0.010
EuS	0.03	0.03	0.08	—	0.003
GdS	0.03	0.10	0.12	—	0.030
TbS	0.03	0.10	0.13	—	0.007
ErS	0.03	0.10	0.13	0.06	0.002
LuS	0.03	0.10	0.13	—	0.001
PrSe	0.04	0.11	0.13	0.2	0.005
NdSe	0.04	0.11	0.13	0.11	0.004
SmSe	0.03	0.10	0.08	—	0.010
EuSe	0.03	0.11	0.09	—	0.003
GdSe	0.03	0.10	0.13	—	0.050
TbSe	0.03	0.10	0.13	—	0.009
ErSe	0.05	0.18	0.25	0.1	0.009
LuSe	0.03	0.11	0.15	—	0.001

5.4 Conclusion

In this work, bond dissociation energy predictions of lanthanide sulfides and selenides were investigated. Three different methods were considered for geometry optimizations. While DFT offers a fast solution for geometry optimizations, CCSD(T) is a more reliable methodology for lanthanide sulfide and selenide diatomics. Between one- and two-component Hamiltonian CCSD(T) calculations, the difference in the optimized geometry was small. A one component Hamiltonian with the ECP basis set offers a balance between speed and reliability that can be used in the future.

f -ccCA is shown here to be a reliable composite scheme. For bond dissociation energy predictions, eight (Pr, Sm, Gd, and Lu complexed with S and Se) of the complexes had energies within 2 kcal·mol⁻¹ from the experimental BDEs. It does have some limitations for molecules with a significant multireference character. However, herein some different routes are offered, which can

be helpful in addressing these limitations. A Breit-Pauli Hamiltonian for spin-orbit calculations proved to be paramount to analyze ground state contributions, orbital occupations, identify state symmetries, and predict accurate spin-orbit contributions while having a mixture of ground and several excited states. Multireference diagnostics aid in identifying problematic molecules and help explain differences from experimental values. Four complexes (EuS, NdS, EuSe, and NdSe) had BDEs within 6 kcal·mol⁻¹ from experiment. Some of these complexes have a large multireference character or different RHF/CCSD(T) electronic configurations for the ground state when compared to those arising from CAS/MRCI, which led to deviations from experiment. For TbS and TbSe, DFT (PW91) provided better quality orbitals than CCSD(T) for the description of their ground state. The degree of multireference character observed for ErS and ErSe was not easily overcome with single reference methods. The use of DFT orbitals in CCSD(T) was not a suitable route for these complexes and even considering different functionals for energy dissociation predictions led to fortuitous results. However, M06-L at a quadruple- ζ level is only 1.7 kcal·mol⁻¹ from experiment; for ErSe, this result should be treated with caution based on the inconsistent behavior of DFT for the dissociation energies of transition metal and lanthanide containing molecules,⁹⁶ as well as the significant and inconsistent shifts in energy with respect to increasing basis set level.

The complexity of these calculations is tremendous, and the aim of this project was to address gaps in the literature in terms of the lanthanide chemistry of sulfides and selenides. For EuS, SmS, EuSe, and SmSe at CAS/MRCI, a new $4f^x 6s^1$ (x = number of electrons) electronic configuration of the lanthanide was postulated for the ground state of these complexes. Overall, this study offers routes that are important in calculating accurate bond dissociation energies for small lanthanide species without significant multireference wavefunction character, though great care is needed to properly describe the correct ground states.

CHAPTER 6

THERMOCHEMISTRY OF PER- AND POLYFLUOROALKYL SUBSTANCES (PFAS)

Timothé Melin lead the project with Betoul Ali and Preston Harell (two undergraduates) running part of the calculations. The article was written by Timothé Melin and Narasimhan Loganathan aided with discussions and revisions.

6.1 Introduction

Per- and polyfluoroalkyl substances (PFAS) represent a diverse family of >5,000 anthropogenic organic molecules which are now widely distributed in various natural settings. The unique hydrophobic and lipophobic properties of PFAS have resulted in their use in many industrial, commercial, and domestic products for more than five decades. Consequently, PFAS are broadly used in common applications including fire-fighting foams, non-stick cookware, paints, cosmetics, carpeting, food wrappers, and attire (leather boots and rain repellent jackets). The extreme persistence of PFAS to degradation (thermal, biological and chemical) originates from the strong carbon-fluorine (C–F) bonds. This persistence can lead to PFAS bioaccumulation, and has been linked to a broad range of detrimental effects on humans, animals, and environmental health.^{236–239} The adverse effects include but are not limited to carcinogenesis, developmental and reproductive disorders, and immunotoxicity.^{237,240} In 2020, the U.S. Environmental Protection Agency (EPA) placed limitations on manufacturing, processing, or importing products containing a number of long-chain PFAS due to the persistence of PFAS in the environment and health effects.

There are numerous ongoing studies across the globe to better understand the linkages and extent of toxicity of PFAS compounds on humans, animals, and the environment.²⁴¹ For example, in recent work, we have illustrated the binding characteristics of a series of prevalent short- and long-chain PFAS with the human pregnane X receptor (hPXR) and with the peroxisome proliferator receptor gamma (PPAR γ) receptor using molecular dynamics simulations.^{237,242}

In considering PFAS in the environment, the majority of studies have been focused upon PFAS in the aqueous/condensed phase. However, in 2011, Ahrens *et. al.*²⁴³ indicated that the atmosphere can play a critical role in the long-range transport of PFAS molecules. In 2012, air sampling and

chemical analysis studies by Del Vento *et. al.*²⁴⁴ identified the presence of PFAS molecules in the atmosphere of the western Antarctic. Such volatile PFAS could explain their presence in remote locations such as the Antarctic ice or lakes in remote parts of Canada. Importantly, recent studies have shown that the inhalation of volatile PFAS from indoor environments amounts to substantial intake of PFAS in humans.²⁴⁰ Furthermore, there is a high risk of direct exposure to volatile PFAS at high concentrations in places where aqueous film-forming foams (AFFF) are utilized.²⁴⁵ Thus, it is vital to have a comprehensive understanding of the stability of PFAS in the atmosphere.

The stability of PFAS and their precursor molecules strongly depends on their gas phase thermochemical properties such as dissociation energies or enthalpies of formation. However, to the best of our knowledge, only a small number of experimental and theoretical studies have examined the gas phase properties of PFAS molecules.^{245–250} Furthermore, the experimental gas phase measurements of PFAS are mostly limited to the identification of different PFAS molecules and their concentrations in the atmosphere.^{245,246} They do not provide quantitative energetic information nor insight about their stability in the gas phase. Similarly, most theoretical investigations on the gas phase properties of PFAS focus on reporting the relative Gibbs free energy of formation for PFOA, PFOS, and their isomers,^{247–249,251} or the enthalpy, entropy, and Gibbs free energies of perfluorinated sulfonic acid.²⁵⁰ Importantly, there are very little, if any, experimental and theoretical gas-phase enthalpies of formation for the list of PFAS molecules investigated in this study. Gaining such thermodynamic insight for a broader range of short and long-chain PFAS and their precursor molecules is important to understand the transport and stability of PFAS in the atmosphere.

In the current work, gas-phase enthalpies of formation for short- and long-chain PFAS molecules are reported. While the goal of this effort is to predict the thermochemistry of PFAS species, as a first step to this, determining which theoretical approaches are useful for PFAS is important. Here, in the absence of experimental gas phase enthalpies of formation for PFAS, selected methodologies will first be gauged for a series of linear alkanoic acids and linear perfluoroalkanes, with detailed analysis. The experimental enthalpies of formation for alkanoic acids are well described, while for perfluoroalkanes, experimental enthalpies are available for several species.^{68,252} Three different

types of linear PFAS will then be considered – perfluorocarboxylic acids, perfluoroalkyl methanol, and fluorotelomer carboxylic acids.

6.2 Methodology

To provide well-gauged predictions and an approach that will be viable for a broader range of PFAS species, the predictions have been made using a number of computational chemistry approaches. While there are multiple routes that can be utilized to predict thermochemical properties (i.e., enthalpy of formation), the two most widely used routes are the total atomization energy⁷⁵ (TAE) and the isogyric family of reactions (isogyric, homodesmotic, hyperhomodesmotic,) approaches.⁷⁶ Importantly, the TAE approach provides a way to obtain enthalpies void of experiment, aside from high precision atomic energies.⁶⁴ Consequently, this approach is highly sensitive to computational method, and requires an accurate description of the energies of both independent atoms and molecules, which is typically disadvantageous to methods such as density functional theory (DFT) and lower-level *ab initio* methods. In contrast, the isogyric reaction approaches are highly dependent on the availability and quality of experimental enthalpies and are relatively less influenced by computational method choice. A comparison of these thermochemical TAE and isogyric approaches has been demonstrated in prior studies for organic molecules.^{55,76,93,253} Importantly, due to the lack of experimental thermodynamic data for PFAS and their precursor molecules needed for an isogyric approach, there will be a significant reliance upon a TAE, where the quality of the methodology utilized is critical.

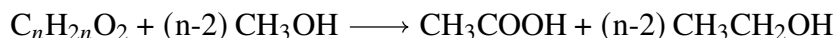
The computational methodologies considered in this effort include two density functionals, M06-2X¹⁰⁴ and B3LYP,^{15,16} and two *ab initio* methods, the domain-based local pair natural orbital coupled cluster single, double and perturbative triple (DLPNO-CCSD(T)),²⁵⁴ and the correlation consistent Composite Approach (ccCA).⁵⁵ The DFT approaches are selected due to their computational efficiency, which will be important in expanded studies of PFAS. B3LYP has been chosen because it is the most widely used functional for the calculation of thermochemical properties of organic molecule.¹⁷ The M06-2X functional developed with an empirical fitting is useful in the calculation of the atomization properties of organic molecules and has been effective in predicting the pK_as for

PFAS.^{255,256}

In terms of *ab initio* methods, DLPNO-CCSD(T) provides a more computationally cost-efficient form of CCSD(T) which is important as CCSD(T), when paired with a large, well-chosen basis set, can provide high-quality energetic data.^{254,257} The cost efficiency of DLPNO-CCSD(T) results from factors including localization schemes, cutoffs for the two electron integrals, and the reduction of the matrix element stored. The total energy and the computational cost of the molecules strongly depends on the selected cutoffs. Another alternative is ccCA, a well-demonstrated *ab initio* composite scheme which provides a route towards high-quality energetic data, akin to CCSD(T) results at or near the complete basis set (CBS) limit – the limit at which no further improvement to the basis set quality will improve the energetic predictions.^{258,259} The ccCA method is described in section 2.9. The enthalpy at 298K is calculated using a pressure of 1 bar and the ideal-gas, rigid-rotor and harmonic oscillator approximation.

Methodological details

The gas phase enthalpy of formation of alkanolic acids, perfluoroalkanes, and a series of PFAS molecules have been predicted at 298 K using the TAE and several isogyric reaction approaches. The isogyric reaction is defined as a reaction in which the same number of paired electrons are present on both sides of the equation. Moreover, the term homodesmotic refers to a reaction in which there are an equal number of carbon-carbon bond types and an equal number of *sp*, *sp*² and *sp*³ carbons bonded to zero, one, two or three hydrogens on both sides of the reaction. For the alkanolic acid, the homodesmotic scheme of the isogyric family of reaction has been used,⁶⁴



Concomitantly, two different isogyric reactions have been selected for use in this study where the number of unpaired electrons is equal on both sides and are as follows:



where n represents the number of carbons. While only one of the isogyric reactions is needed, CO_2 , $\text{HF}/\text{CF}_3\text{H}$ and F_2 are known byproducts in PFAS degradation. Thus, the two reactions are used to provide insight about the reliability of DFT, ccCA and DLPNO-CCSD(T) for the study of the thermochemical properties of PFAS and PFAS byproducts.^{260,261} These two reactions have been chosen for the isogyric approach as the experimental enthalpies of formation for F_2 , CF_4 , CO_2 , HF and CF_3H molecules are accurately known and are given in the appendix (see Table D.2).²⁶² To calculate the energy of the molecules and atoms, the B3LYP and M06-2X density functionals, ccCA and the DLPNO-CCSD(T) have been used. For the DLPNO-CCSD(T) calculations, the aug-cc-pVnZ ($n = D, T, Q$) basis set were used for each of the atoms, while for DFT only the aug-cc-pVTZ basis set were used, as convergence of structural and energetic properties of main group species is typically reached at the triple-zeta level in DFT calculations.

The TightPNO settings in ORCA were used for DLPNO-CCSD(T), with the following threshold: $T_{\text{CutPairs}} = 10^{-5}$ Hartree, $T_{\text{CutPNO}} = 10^{-7}$ and $T_{\text{CutMKN}} = 10^{-4}$, corresponding to the cutoffs of the occupation number (T_{CutPNO}), the orbital Mulliken populations (T_{CutMKN}) and to the differential overlap (T_{CutPairs}), respectively. The TightPNO settings were chosen based on a study done by Patel *et al.*²⁶³ where DLPNO was used to calculate the enthalpy of formation of alkane. For DLPNO-CCSD(T), energies were extrapolated to the CBS limit using a three-point mixed gaussian/exponential scheme.⁶⁰ (Further details about DLPNO methods can be found in Minenkov *et al.* and Patel *et al.*).^{253,263} Finally, for the ccCA and DLPNO-CBS determinations for the largest molecules, the enthalpies of formation were extrapolated based on the trends observed for each method. (An example of the procedure used is provided in the appendix D for alkanoic acids).

The optimized geometry for each molecule was determined using the B3LYP functional in combination with the 6-31G basis set to provide quick structure determination, followed by a frequency calculation using the same method and basis set to ensure that the geometry represents the minima on the potential energy curve. The optimized ground state geometries of all molecules examined correspond to a helical structure. The ground state structure of PFOA was consistent with earlier studies of PFOA by Liu *et al.*²⁶⁴ The 6-31G basis set has been used in numerous

studies to obtain structures for small organic molecules^{49,55,265} and larger molecules.²⁶⁶ Moreover, because of the low computational cost, B3LYP/6-31G provides a useful choice for large alkanolic, perfluoroalkanes and PFAS molecules.

The DFT and ccCA calculations were performed using NWChem 6.1,²⁶⁷ and MOLPRO 2020,¹⁴⁷ respectively, and the DLPNO-CCSD(T) were carried out using ORCA 4.2.1.²⁶⁸

Importantly, three groups of molecules were considered in this study as shown in Table 6.1: (1) alkanolic acids, (2) perfluoroalkanes, and (3) three varieties of linear PFAS, which are referred to herein as PFAS, PFAS-ol, and FTCA.

Table 6.1: List of molecules used in this study.

Name	Chemical formula	Carbon chain length
Alkanolic acid	$\text{H}_3\text{C}-(\text{H}_2\text{C})_n-\text{COOH}$	$n = 0 - 6$
Perfluoroalkanes	$\text{C}_2\text{F}_{2n+2}$	$n = 2 - 8$
Perfluoroalkanoic acid (PFAS)	$\text{F}_3\text{C}-(\text{F}_2\text{C})_n-\text{COOH}$	$n = 0 - 6$
Perfluoroalkyl methanol (PFAS-ol)	$\text{F}_3\text{C}-(\text{F}_2\text{C})_{n-1}-\text{CH}_2-\text{OH}$	$n = 1 - 7$
Fluorotelomer carboxylic acid (FTCA)	$\text{F}_3\text{C}-(\text{F}_2\text{C})_{n-1}-\text{CH}_2-\text{COOH}$	$n = 1 - 7$

While quality experiments are the ideal gauge for theoretical predictions, in the absence of these, an alternative approach is needed. Though more sophisticated theoretical approaches are preferred and are being examined here, an estimate – the empirical Joback approach,²⁶⁹ is considered here simply to provide a very quick qualitative comparison for other methods. The Joback method is based on a group contribution scheme – a scheme that presumes that the thermodynamic contributions from a specific structural component (i.e., $=\text{CH}_2$, $-\text{CO}_2\text{H}$, $-\text{F}$, $-\text{N}=\text{}$) of a molecule is the same from molecule to molecule. In the Joback approach, the contributions to thermodynamic property (in this case, enthalpy of formation) from each structural component, are combined additively and are empirically parameterized to predict an overall enthalpy for formation for a molecule. (The parameters for each chemical entity are reported in Joback and Reid). Consequently, this method provides a quick very approximate estimate for the enthalpies of formation for PFAS as experimental data is limited. To demonstrate its possible utility for comparable qualitative information, the Joback method is utilized on the alkanolic acids first, and subsequently applied to perfluoroalkanes. Finally,

in the appendix (see Table D.3), the relative timings of the B3LYP, M06-2X, ccCA and DLPNO-CCSD(T) calculations are given for $C_3H_6O_2$, C_3F_8 , and $C_3F_5HO_2$, providing a representative, molecule of similar size from each class of compounds for comparison.

6.3 Results and discussion

Alkanoic acids

The $\Delta H_f^{298\text{ K}}$ s of the alkanoic acids calculated using the total atomization approach are presented along with their corresponding experimental values in Table 6.2 and Figure D.1. The computed ccCA enthalpies for the acids are in excellent agreement with experimental values from NIST database.⁶⁸ The differences between experimental and computed $\Delta H_f^{298\text{ K}}$ are less than 1 kcal·mol⁻¹ until pentanoic acids, and ~3 kcal·mol⁻¹ for the hexanoic acid. For the heptanoic and octanoic acid, the estimation is also in good agreement with experiment with a difference error of ~3 kcal·mol⁻¹. Overall, this level of agreement (within 1 kcal·mol⁻¹ from those of experiment, on average) between ccCA and experiment for the prediction of enthalpies of formation has been previously demonstrated for organic molecules (from diatomics to n-octane).^{55,57}

Table 6.2: Enthalpies of formation ($\text{kcal}\cdot\text{mol}^{-1}$) for alkanolic acids determined with an atomization approach.

	B3LYP	M06-2X	ccCA	DLPNO-CCSD(T)				Joback	Exp ^b
				adz ^a	atz ^a	aqz ^a	CBS ^a		
C ₂ H ₄ O ₂	-92.49	-97.13	-102.04	-43.19	-78.68	-89.92	-98.9	-103.94	-103.49 ± 0.72
C ₃ H ₆ O ₂	-95.13	-103.16	-108.17	-31.9	-79.33	-93.19	-104	-108.87	-108.58 ± 0.47
C ₄ H ₈ O ₂	-98.00	-106.73	-113.98	-22.52	-80.38	-96.5	-109.2	-113.89	-112.80 ± 0.96
C ₅ H ₁₀ O ₂	-98.47	-110.92	-118.40	-7.21	-78.55	-97.69	-112.4	-118.74	-117.20 ± 4.78
C ₆ H ₁₂ O ₂	-101.94	-117.28	-125.40	3.32	-79.96	-101.75	-118.4	-123.67	-122.70 ± 0.96
C ₇ H ₁₄ O ₂	-103.00	-120.02	-131.24 ± 0.92 ^c	16.17	-79.03	-103.45	-122.1	-128.60	-127.70 ± 0.41
C ₈ H ₁₆ O ₂	-110.11	-130.36	-137.08 ± 0.92 ^c	23.17	-83.96	-111.1	-131.7	-133.38	-132.70 ± 0.31

^a DLPNO-CCSD(T).anz represent a DLPNO-CCSD(T) calculation with the corresponding aug-cc-pVnZ basis set, while DLPNO-CBS is the enthalpy obtained at the complete basis set limit.

^b All experimental values are taken from the NIST database.⁶⁸

^c Extrapolated. The uncertainties were based on the root mean square deviation of the extrapolated enthalpies of formation. The methodology used for extrapolation and uncertainties are provided in appendix D.

The DLPNO-CCSD(T).anz enthalpies for all of the molecules are far from experiment (Table 6.2) and exhibit very slow convergence toward the complete basis set limit. At the CBS limit, the $\Delta H_f^{298\text{ K}}$ s are overestimated (less negative) by ~ 2 to $\sim 5\text{ kcal}\cdot\text{mol}^{-1}$ as compared to experiment for all of the carboxylic acids with the exception of octanoic acid. Similarly, both DFT functionals overestimate $\Delta H_f^{298\text{ K}}$, with B3LYP resulting in larger errors (~ 11 - $22\text{ kcal}\cdot\text{mol}^{-1}$) relative to experiment and M06-2X being closer to experiment with an error of ~ 2 - $7\text{ kcal}\cdot\text{mol}^{-1}$. The large differences in the enthalpies for B3LYP and M06-2X were also observed in earlier DFT studies where the thermochemical properties of hydrocarbons are better predicted with M06-2X than B3LYP due to its higher exact exchange energy (54% for M06-2X and 20% for B3LYP).^{255,270}

Interestingly, the $\Delta H_f^{298\text{ K}}$ s calculated with the empirical Joback method are in very good agreement with the experimental data obtained from the NIST database⁶⁸ with an error of less than 1

kcal·mol⁻¹ for all alkanolic molecules. Such low errors could be attributed to the very low interactions between different –CH₂– groups (the electron of hydrogen is mostly involved in the bonding) for the alkanolic molecules.

Importantly, the consistency in the prediction of $\Delta H_f^{298\text{ K}}$ of a method is determined by comparing the variation in the enthalpies as the alkanolic molecules are increased in size (number of CH₂). It should be noted that the difference in experimental $\Delta H_f^{298\text{ K}}$ values between any two consecutive alkanolic molecules (number of CH₂ increasing by one) (Table 6.2), is on average ~4-5 kcal·mol⁻¹. Thus, if a method is capable of predicting similar enthalpy variation, then the results are likely to be consistent over the molecule set. For example, the ccCA enthalpy variation between ethanoic and propanoic acid is -6.1 kcal·mol⁻¹ which is 1 kcal·mol⁻¹ lower than the experimental variation (-5.1 kcal·mol⁻¹). Therefore, the ccCA method is consistent in predicting $\Delta H_f^{298\text{ K}}$ when the reported variation in enthalpies is similar between any two consecutive molecules.

Table 6.2 clearly shows that the enthalpy variation between any two consecutive alkanolic molecules determined with ccCA is ~4.5-6 kcal·mol⁻¹ until pentanoic acid, which increases by another ~2 kcal·mol⁻¹ between the pentanoic and hexanoic acid which could be attributed to the underestimation of the enthalpy of formation for hexanoic acid by ~3 kcal·mol⁻¹ in comparison to experiment. Nevertheless, this enthalpy variation between pentanoic and hexanoic acid with ccCA is ~1.5 kcal·mol⁻¹ when compared to their corresponding experimental enthalpy variation. For DLPNO-CBS, the enthalpy variation is in agreement with experiment for each molecule (difference of less than 1 kcal·mol⁻¹) with the exception of butanoic and octanoic acid for which the $\Delta H_f^{298\text{ K}}$ values are closer to experiment than for the other acids.

The $\Delta H_f^{298\text{ K}}$ values for alkanolic acid using a homodesmotic scheme are depicted in Table 6.3 and Figure D.2. It is evident that $\Delta H_f^{298\text{ K}}$ s computed using B3LYP and ccCA have the lowest error compared to experiment, with an error in the range of 0.1-3.6 kcal·mol⁻¹ and 0.6-2.4 kcal·mol⁻¹, respectively. Table 6.3 clearly illustrates that the computed $\Delta H_f^{298\text{ K}}$ values converge toward the experimental enthalpies of formation. Both ccCA and DLPNO-CBS have similar errors as compared to experiment: less than ~3 kcal·mol⁻¹ for every compound. In contrast to the TAE approach, the

$\Delta H_f^{298\text{ K}}$ values from B3LYP are better correlated with experiments than M06-2X.

For instance, the $\Delta H_f^{298\text{ K}}$ for propanoic acid using B3LYP with TAE is 13 kcal·mol⁻¹ larger than experiment, whereas with the homodesmotic approach, this difference is substantially reduced to less than 1 kcal·mol⁻¹ from experiment. This shows the impact of the method error-cancellation possible from the isogyric family of approaches, provided sufficient experimental data is available.

Finally, the enthalpy variation between any two consecutive molecules (the number of CH₂ increased by one) for ccCA is ~4-6 kcal·mol⁻¹ until propanoic acid and is consistent with the experimental range of ~5 kcal·mol⁻¹. Similarly, the DLPNO-CBS variation is also in good agreement with experiment (~5 kcal·mol⁻¹) except for C₈H₁₆O₂ for which the error in $\Delta H_f^{298\text{ K}}$ with experiment was much larger (~9 kcal·mol⁻¹) than for the smaller alkanoic acids. Overall, ccCA gave the best results compared to all methods with a mean unsigned error of 1.33 and 0.94 kcal·mol⁻¹ for the TAE and homodesmotic approach, respectively. Note that the $\Delta H_f^{298\text{ K}}$ values reported in Tables 6.2 and 6.3 are the same for the Joback method, as the method is not impacted by thermochemical reaction scheme.

Table 6.3: Enthalpies of formation ($\text{kcal}\cdot\text{mol}^{-1}$) for alkanoic acids determined with a homodesmotic approach.

	B3LYP	M06-2X	ccCA	DLPNO-CCSD(T)				Joback	Exp ^b
				adz ^a	atz ^a	aqz ^a	CBS ^a		
C ₂ H ₄ O ₂	-103.49	-103.49	-103.49	-103.49	-103.49	-103.49	-103.5	-103.94	-103.49 ± 0.72
C ₃ H ₆ O ₂	-108.64	-110.24	-109.24	-109.59	-109.33	-109.23	-109.1	-108.87	-108.58 ± 0.47
C ₄ H ₈ O ₂	-114.31	-114.71	-114.86	-117.8	-115.77	-115.21	-114.8	-113.89	-112.80 ± 0.96
C ₅ H ₁₀ O ₂	-117.03	-119.45	-118.72	-119.71	-118.96	-118.68	-118.5	-118.74	-117.20 ± 4.78
C ₆ H ₁₂ O ₂	-122.80	-126.30	-125.12	-126.35	-125.35	-125	-124.7	-123.67	-122.70 ± 0.96
C ₇ H ₁₄ O ₂	-126.47	-129.86	-130.19 ± 2.42 ^c	-131.00	-129.72	-129.27	-128.9	-128.60	-127.70 ± 0.41
C ₈ H ₁₆ O ₂	-136.31	-141.11	-135.26 ± 2.42 ^c	-141.59	-140.05	-139.59	-139.2	-133.38	-132.70 ± 0.31

^a DLPNO-CCSD(T).anz represent a DLPNO-CCSD(T) calculation with the corresponding aug-cc-pVnZ basis set, while DLPNO-CBS is the enthalpy obtained at the complete basis set limit.

^b All experimental values are taken from the NIST database.⁶⁸

^c Extrapolated. The uncertainties were based on the root mean square deviation of the extrapolated enthalpies of formation. The methodology used for extrapolation and uncertainties are provided in appendix D.

Perfluoroalkanes

The performance of the considered methods in computing the $\Delta H_f^{298\text{ K}}$ of perfluoroalkanes using TAE approach with respect to experimental values is reported in Table 6.4 and Figure D.3.

As shown in Table 6.4 and Figure D.3, the enthalpies computed with ccCA, M06-2X and DLPNO-CBS methods are closer to experiment than those determined with the other methods. Notably, the difference between experimental and computed $\Delta H_f^{298\text{ K}}$ values increase as the length of the molecule increases for all methods. For instance, the $\Delta H_f^{298\text{ K}}$ values for C₂F₆ have an error of 1 $\text{kcal}\cdot\text{mol}^{-1}$ with M06-2X and ccCA while for C₄F₁₀ the errors are ~14 and ~16 $\text{kcal}\cdot\text{mol}^{-1}$, respectively. In addition, the $\Delta H_f^{298\text{ K}}$ values from M06-2X and ccCA methods are very close to one another with a maximum difference of 4 $\text{kcal}\cdot\text{mol}^{-1}$ for C₅F₁₂, while the DLPNO-CBS enthalpies

are slightly higher by an additional 2 to 3 kcal·mol⁻¹.

The ccCA results for C₂F₆ are within 1 kcal·mol⁻¹ of the experimental values reported by Kolesov *et. al.*²⁷¹ and the computational enthalpies of formation by Paulechka *et. al.*²⁷² Similarly, DLPNO-CBS result in an error of ~0.5 kcal·mol⁻¹ with experiment for C₂F₆.

For C₃F₈, the ccCA and M06-2X errors compared to experiment are ~7 and ~6 kcal·mol⁻¹, respectively.²⁵² The $\Delta H_f^{298\text{ K}}$ values from these methods are in good agreement with those of Burcat *et. al.*²⁷³ (calculated using experimental data and an empirical equation) with an error of 1 kcal·mol⁻¹. In addition, the $\Delta H_f^{298\text{ K}}$ values reported using ccCA are closer to ones reported by Ventura *et. al.* (isogyric approach) using B3PW91 and CBS-Q with an error of ~5 and ~2 kcal·mol⁻¹, respectively.²⁷⁴

Table 6.4: Enthalpies of formation (kcal·mol⁻¹) for perfluoroalkanes determined with an atomization approach.

	B3LYP	M06-2X	ccCA	DLPNO-CCSD(T)				Joback	Exp ^b
				adz ^a	atz ^a	aqz ^a	CBS ^a		
C ₂ F ₆	-302.53	-321.57	-321.1	-255.03	-292.82	-308.1	-320.7	-305.63	-321.22 ± 0.96 ^b
C ₃ F ₈	-392.15	-420.79	-419.35	-329.73	-380.84	-401.36	-418.3	-406.40	-426.55 ± 1.40 ^c
C ₄ F ₁₀	-481.33	-519.94	-517.35	-404.68	-468.78	-494.4	-515.6	-507.17	-533.90 ^d
C ₅ F ₁₂	-570.58	-619.22	-615.6	-479.81	-557.08	-587.61	-612.8	-607.93	—
C ₆ F ₁₄	-659.83	-718.49	-713.77 ± 0.11 ^e	-555.28	-645.33	-680.92	-710.3	-708.70	—
C ₇ F ₁₆	-749.12	-817.78	-811.93 ± 0.11 ^e	-630.56	-733.53	—	-807.7 ± 0.1 ^e	-809.47	-809.13 ± 0.86 ^b
C ₈ F ₁₈	-838.30	-917.02	-910.10 ± 0.11 ^e	-706.00	-821.76	—	-905.0 ± 0.1 ^e	-910.24	—

^a DLPNO-CCSD(T).anz represent a DLPNO-CCSD(T) calculation with the corresponding aug-cc-pVnZ basis set, while DLPNO-CBS is the enthalpy obtained at the complete basis set limit.

^b Ref. ²⁷¹

^c Ref. ²⁵²

^d Ref. ²⁷⁵

^e Extrapolated. The uncertainties were based on the root mean square deviation of the extrapolated enthalpies of formation. The methodology used for extrapolation and uncertainties are provided in appendix D.

The $\Delta H_f^{298\text{ K}}$ values for C_4F_{10} using ccCA and DLPNO-CBS are in reasonable agreement (error of $\sim 2\text{ kcal}\cdot\text{mol}^{-1}$) with the NIST recommended enthalpy of $-515.2\text{ kcal}\cdot\text{mol}^{-1}$ computed with B88LYP methods as reported by Stewart *et. al.*²⁷⁵ In addition, the values of C_4F_{10} with ccCA and DLPNO-CBS are closer to the energy from Burcat *et. al.*²⁷³ ($-510.85\text{ kcal}\cdot\text{mol}^{-1}$) which is calculated using experimental data and an empirical equation, with a difference of $6\text{ kcal}\cdot\text{mol}^{-1}$. It is important to note that although all predicted enthalpies of formation tend toward $-515 \pm 5\text{ kcal}\cdot\text{mol}^{-1}$, the $\Delta H_f^{298\text{ K}}$ values do not agree with the experimental value of $\sim -533\text{ kcal}\cdot\text{mol}^{-1}$. Based upon the theoretical studies and NIST database utilizing the semiempirical computational data rather than reporting the existing experimental values, it is important to reevaluate the experiment.

For C_7F_{16} , the extrapolated ccCA and DLPNO-CBS enthalpy of formation are in better agreement with the experimental $\Delta H_f^{298\text{ K}}$ values with an error of $\sim 2\text{ kcal}\cdot\text{mol}^{-1}$. On the other hand, the $\Delta H_f^{298\text{ K}}$ errors of M06-2X compared to experiment is $\sim 8\text{ kcal}\cdot\text{mol}^{-1}$.

The empirical Joback approach results in large differences in $\Delta H_f^{298\text{ K}}$ values compared to experiment (i.e., $\sim 13\text{ kcal}\cdot\text{mol}^{-1}$ for C_2F_6 , $\sim 20\text{ kcal}\cdot\text{mol}^{-1}$ for C_3F_8 , and $\sim 26\text{ kcal}\cdot\text{mol}^{-1}$ for C_4F_{10}) with the exception of C_7F_{16} ($\sim 1\text{ kcal}\cdot\text{mol}^{-1}$). Such large errors from experiment demonstrate the limitations of the Joback approach, likely attributed, in part, to the lack of interactions incorporated in the model.

Importantly, the variation in $\Delta H_f^{298\text{ K}}$ values between any two consecutive molecules (with each addition of CF_2) using ccCA, M06-2X, DLPNO-CBS and Joback is $\sim 100\text{ kcal}\cdot\text{mol}^{-1}$ with a difference of $7\text{ kcal}\cdot\text{mol}^{-1}$ with respect to experiments (see Table 6.3).

Based on the performance of ccCA, M06-2X and DLPNO-CBS methods in reproducing the experimental enthalpies of formation for perfluoroalkanes and showing similar $\Delta H_f^{298\text{ K}}$ variation, these methods will be used as the routes to predict enthalpies of formation in the following sections on PFAS as there are no known experimental data for the PFAS molecules studied here. Furthermore, the good agreement between experimental and computed $\Delta H_f^{298\text{ K}}$ of alkanolic acids using ccCA (both TAE and isogyric approaches) clearly advocates the use of ccCA to gauge the accuracy of other methods for PFAS in the following sections. Finally, although the B3LYP results for the alkanolic

acids using the atomization approach are the furthest from experiment, B3LYP is in much better agreement with experiment when an isogyric approach is used. Thus, the performance of B3LYP will also be evaluated for PFAS molecules in order to assess its usefulness.

PFAS

The $\Delta H_f^{298\text{ K}}$ values for both short and long chain PFAS molecules using TAE are represented in Table 6.5 and Figure D.4.

Table 6.5: Enthalpies of formation ($\text{kcal}\cdot\text{mol}^{-1}$) for PFAS determined with an atomization approach.

	B3LYP	M06-2X	ccCA	DLPNO-CCSD(T)				Joback
				adz ^a	atz ^a	aqz ^a	CBS ^a	
C ₂ F ₃ HO ₂	-230.59	-244.80	-247.51	-177.01	-216.58	-232.49	-245.6	-246.64
C ₃ F ₅ HO ₂	-322.30	-345.69	-347.84	-253.68	-306.68	-327.79	-345.2	-347.41
C ₄ F ₇ HO ₂	-410.63	-443.33	-444.94	-327.52	-393.62	-419.81	-441.4	-448.18
C ₅ F ₉ HO ₂	-500.44	-544.97	-543.78	-403.25	-482.51	-513.82	-539.6	-548.94
C ₆ F ₁₁ HO ₂	-589.71	-642.99	-642.54	-478.66	-570.82	-607.11	-637	-649.71
			$\pm 1.32^b$					
C ₇ F ₁₃ HO ₂	-679.39	-743.27	-741.29	-554.34	-659.44	-700.64	-734.6	-750.48
			$\pm 1.32^b$					
C ₈ F ₁₅ HO ₂	-768.40	-837.77	-840.05	-629.43	-747.25	—	-832.4	-851.25
			$\pm 1.32^b$				$\pm 1.11^b$	

^a DLPNO-CCSD(T).anz represent a DLPNO-CCSD(T) calculation with the corresponding aug-cc-pVnZ basis set, while DLPNO-CBS is the enthalpy obtained at the complete basis set limit.

^b Extrapolated. The uncertainties were based on the root mean square deviation of the extrapolated enthalpies of formation. The methodology used for extrapolation and uncertainties are provided in appendix D.

For short chain PFAS molecules with $C < 6$, the $\Delta H_f^{298\text{ K}}$ values of M06-2X and DLPNO-CBS methods computed are very similar to those of ccCA, with a maximum difference of $\sim 4\text{ kcal}\cdot\text{mol}^{-1}$. In contrast, the $\Delta H_f^{298\text{ K}}$ values computed from DLPNO-CCSD(T) for all PFAS molecules varies substantially with their corresponding ccCA values and this difference increases as the length of the PFAS molecule increases. However, it is evident from Table 6.5 that the difference in $\Delta H_f^{298\text{ K}}$ values with ccCA decreases significantly with increase in the size of the basis set for DLPNO-CCSD(T). Nevertheless, despite the improvement in computing the $\Delta H_f^{298\text{ K}}$ values, there is a $\sim 15\text{ kcal}\cdot\text{mol}^{-1}$

difference in enthalpy even for the shortest molecule ($\text{C}_2\text{F}_3\text{HO}_2$) with the quadruple- ζ basis set in comparison to ccCA. It is only at the CBS limit that the enthalpy differences with ccCA are only a few $\text{kcal}\cdot\text{mol}^{-1}$ for each molecule.

For PFAS with $C>6$, the difference in $\Delta H_f^{298\text{ K}}$ values between M06-2X and DLPNO-CBS increases with a difference of up to $\sim 9\text{ kcal}\cdot\text{mol}^{-1}$ for PFHpA ($\text{C}_7\text{F}_{13}\text{HO}_2$). When comparing DFT methods, the $\Delta H_f^{298\text{ K}}$ s using B3LYP are significantly higher than M06-2X with a minimum difference of $\sim 15\text{ kcal}\cdot\text{mol}^{-1}$ and that increases with increase in the length of the PFAS molecules. Surprisingly, the $\Delta H_f^{298\text{ K}}$ values of PFAS using Joback method are closer to ccCA, M06-2X and DLPNO-CBS than for perfluoroalkanes despite the lack of interactions in the model. Importantly, the enthalpy variation between any two consecutive short chain PFAS molecules is in the range of ~ 98 to $101\text{ kcal}\cdot\text{mol}^{-1}$ for ccCA.

Therefore, based on the good correlation between experimental and ccCA enthalpies for perfluoroalkanes, the ccCA energies for PFAS are recommended. The $\Delta H_f^{298\text{ K}}$ values of PFAS with two isogyric reactions namely ISO1 (Table 6.6) and ISO2 (Table 6.7) are given in Figure D.5. Note, the enthalpies reported in this section for Joback method are similar to Table 6.4 for the same reason described earlier.

Table 6.6: Enthalpies of formation ($\text{kcal}\cdot\text{mol}^{-1}$) for PFAS determined with an isogyric approach (ISO1).

	B3LYP	M06-2X	ccCA	DLPNO-CCSD(T)				Joback
				adz ^a	atz ^a	aqz ^a	CBS ^a	
C ₂ F ₃ HO ₂	-247.87	-241.92	-246.6	-250.63	-248.42	-246.73	-245.2	-246.64
C ₃ F ₅ HO ₂	-348.51	-333.32	-347.64	-357.62	-350.94	-347.1	-343.8	-347.41
C ₄ F ₇ HO ₂	-447.04	-422.74	-446.73	-463.05	-451.57	-445.47	-440.3	-448.18
C ₅ F ₉ HO ₂	-546.41	-515.51	-546.92	-569.73	-553.52	-545.19	-538	-548.94
C ₆ F ₁₁ HO ₂	-645.18	-604.61	-647.03 $\pm 0.80^b$	-676.03	-654.81	-644.12	-635	-649.71
C ₇ F ₁₃ HO ₂	-744.39	-696.01	-747.14 $\pm 0.80^b$	-782.63	-756.46	-743.33	-732.2	-750.48
C ₈ F ₁₅ HO ₂	-842.91	-781.59	-847.25 $\pm 0.80^b$	-888.62	-857.27	—	-829.6 $\pm 0.73^b$	-851.25

^a DLPNO-CCSD(T).anz represent a DLPNO-CCSD(T) calculation with the corresponding aug-cc-pVnZ basis set, while DLPNO-CBS is the enthalpy obtained at the complete basis set limit.

^b Extrapolated. The uncertainties were based on the root mean square deviation of the extrapolated enthalpies of formation. The methodology used for extrapolation and uncertainties are provided in appendix D.

Table 6.7: Enthalpies of formation ($\text{kcal}\cdot\text{mol}^{-1}$) for PFAS determined with an isogyric approach (ISO2).

	B3LYP	M06-2X	ccCA	DLPNO-CCSD(T)				Joback
				adz ^a	atz ^a	aqz ^a	CBS ^a	
C ₂ F ₃ HO ₂	-241.45	-247.82	-244.75	-246.18	-245.14	-244.96	-244.8	-246.64
C ₃ F ₅ HO ₂	-342.08	-339.22	-345.79	-353.17	-347.66	-345.33	-343.4	-347.41
C ₄ F ₇ HO ₂	-440.61	-428.64	-444.89	-458.6	-448.29	-443.7	-439.9	-448.18
C ₅ F ₉ HO ₂	-539.99	-521.41	-545.08	-565.28	-550.24	-543.41	-537.7	-548.94
C ₆ F ₁₁ HO ₂	-638.75	-610.51	-643.83	-671.58	-651.53	-642.34	-634.6	-649.71
			$\pm 1.32^b$					
C ₇ F ₁₃ HO ₂	-737.96	-701.90	-742.59	-778.18	-753.18	-741.56	-731.8	-750.48
			$\pm 1.32^b$					
C ₈ F ₁₅ HO ₂	-836.48	-787.49	-841.35	-884.17	-853.99	—	-829.6	-851.25
			$\pm 1.32^b$				$\pm 1.11^b$	

^a DLPNO-CCSD(T).anz represent a DLPNO-CCSD(T) calculation with the corresponding aug-cc-pVnZ basis set, while DLPNO-CBS is the enthalpy obtained at the complete basis set limit.

^b Extrapolated. The uncertainties were based on the root mean square deviation of the extrapolated enthalpies of formation. The methodology used for extrapolation and uncertainties are provided in appendix D.

Table 6.6 and 6.7 clearly show that, the computed enthalpies are similar for both isogyric reactions for all methods used in this study. However, it is important to highlight that the $\Delta H_f^{298\text{ K}}$ s computed with these reactions are substantially different than their corresponding enthalpic values using TAE. For instance, it is evident from Table 6.6 and 6.7 that all DLPNO-CCSD(T) methods show a huge improvement in predicting enthalpies closer to ccCA with isogyric reactions than TAE which could be attributed to the efficiency of the isogyric type of approach with any methods to compute thermochemistry properties. It is clearly demonstrated by the $\Delta H_f^{298\text{ K}}$ difference between DLPNO-CCSD(T)/aug-cc-pVDZ and DLPNO-CCSD(T)/aug-cc-pVQZ for C₂F₃HO₂ amounts to $\sim 2\text{ kcal}\cdot\text{mol}^{-1}$ in the isogyric reactions in contrast to $\sim 55\text{ kcal}\cdot\text{mol}^{-1}$ with TAE.

For the small-chain PFAS (<6 carbons), DLPNO-CCSD(T)/aug-cc-pVQZ leads to the best results compared to ccCA for both isogyric schemes with an error varying between ~ 0.2 to $\sim 2\text{ kcal}\cdot\text{mol}^{-1}$. In contrast, at the CBS limit, the $\Delta H_f^{298\text{ K}}$ s obtained are slightly further from the ccCA reference. As observed for the alkanoic acids, $\Delta H_f^{298\text{ K}}$ determined with B3LYP are much closer to

ccCA, while the $\Delta H_f^{298\text{ K}}$ s of M06-2X is significantly different from ccCA in both isogyric reaction approaches. For instance, in the case ISO1, the B3LYP $\Delta H_f^{298\text{ K}}$ errors compared to the ccCA $\Delta H_f^{298\text{ K}}$ range between ~ 0.8 -2 kcal·mol⁻¹ in contrast to ~ 5 -32 kcal·mol⁻¹ between the M06-2X and ccCA $\Delta H_f^{298\text{ K}}$. Similarly, with ISO2, the error between the B3LYP and ccCA $\Delta H_f^{298\text{ K}}$ are in the range of ~ 3 -4 kcal·mol⁻¹, while the errors between the M06-2X and ccCA $\Delta H_f^{298\text{ K}}$ are 3-24 kcal·mol⁻¹.

For the longest PFAS (C>6), DLPNO-CCSD(T) is used for comparison due to increasing system size. (There is a version of ccCA that incorporates DLPNO-CCSD(T) to enable application to larger systems,⁶⁰ however, it was not used here, as it is first important to gauge the utility of DLPNO-CCSD(T).) For these molecules, the B3LYP $\Delta H_f^{298\text{ K}}$ differs by ~ 3.5 kcal·mol⁻¹ from the DLPNO-CCSD(T)/aug-cc-pVQZ $\Delta H_f^{298\text{ K}}$, while the M06-2X $\Delta H_f^{298\text{ K}}$ differs from the DLPNO-CCSD(T) /aug-cc-pVQZ $\Delta H_f^{298\text{ K}}$ by ~ 32 kcal·mol⁻¹. Finally, the extrapolated values of ccCA are ~ 10 kcal·mol⁻¹ away from DLPNO-CCSD(T)/aug-cc-pVQZ.

The influence of different terminal functional groups on formation enthalpies were evaluated using the atomization approach with PFAS-ol (Table 6.8) and FTCA (Table 6.9) are illustrated in Figure D.6.

Table 6.8: Enthalpies of formation ($\text{kcal}\cdot\text{mol}^{-1}$) for PFAS-ol determined with the atomization approach.

	B3LYP	M06-2X	ccCA	DLPNO-CCSD(T)				Joback
				adz ^a	atz ^a	aqz ^a	CBS ^a	
C ₂ F ₃ H ₃ O	-198.38	-211.53	-214.11	-148.5	-188.54	-202.42	-213.6	-199.31
C ₃ F ₅ H ₃ O	-289.04	-311.58	-313.44	-223.39	-277.78	-297.19	-312.9	-300.08
C ₄ F ₇ H ₃ O	-378.57	-411.79	-411.78	-296.15	-365.12	-389.89	-410	-400.85
C ₅ F ₉ H ₃ O	-468.57	-510.47	-510.73	-375.75	-453.57	-482.49	-506.1	-501.61
C ₆ F ₁₁ H ₃ O	-557.02	-609.69	-609.64 $\pm 0.41^b$	-450.18	-540.85	-574.88	-602.7	-602.38
C ₇ F ₁₃ H ₃ O	-646.34	-709.62	-708.47 $\pm 0.41^b$	-525.50	-629.11	-668.04	-699.9	-703.15
C ₈ F ₁₅ H ₃ O	-736.40	-807.56	-807.34 $\pm 0.41^b$	-601.75	-718.3	—	-797.2 $\pm 1.1^b$	-803.91

^a DLPNO-CCSD(T).anz represent a DLPNO-CCSD(T) calculation with the corresponding aug-cc-pVnZ basis set, while DLPNO-CBS is the enthalpy obtained at the complete basis set limit.

^b Extrapolated. The uncertainties were based on the root mean square deviation of the extrapolated enthalpies of formation. The methodology used for extrapolation and uncertainties are provided in appendix D.

As shown in Table 6.8, the enthalpies of formation of PFAS-ol determined using M06-2X and DLPNO-CBS are very close to the values obtained with ccCA methods. The maximum difference in $\Delta H_f^{298\text{ K}}$ values between M06-2X and ccCA is $\sim 2.5 \text{ kcal}\cdot\text{mol}^{-1}$ for C₂F₃H₃O that decreases with increase in the length of the hydrocarbon and reaches $\sim 0.02 \text{ kcal}\cdot\text{mol}^{-1}$ for C₄F₇H₃O. The $\Delta H_f^{298\text{ K}}$ using DLPNO-CBS are in good agreement with ccCA values with a largest difference of $\sim 1.8 \text{ kcal}\cdot\text{mol}^{-1}$ for C₄F₇H₃O. For PFAS-ol molecules with C>4, the difference in $\Delta H_f^{298\text{ K}}$ values between M06-2X and DLPNO-CBS varies substantially with increasing length of molecules, with a difference of $\sim 4 \text{ kcal}\cdot\text{mol}^{-1}$ for C₅F₉H₃O and $\sim 10 \text{ kcal}\cdot\text{mol}^{-1}$ for C₇F₁₃H₃O. However, the M06-2X $\Delta H_f^{298\text{ K}}$ are in very good agreement with the extrapolated ccCA $\Delta H_f^{298\text{ K}}$. Similar to the DLPNO-CCSD(T) values of PFAS molecules with TAE approach, the $\Delta H_f^{298\text{ K}}$ s varies significantly from one basis set to another using DLPNO-CCSD(T) but still are further from ccCA values. Among DFT methods, the difference in $\Delta H_f^{298\text{ K}}$ values computed with B3LYP and M06-2X are $\sim 13 \text{ kcal}\cdot\text{mol}^{-1}$ for C₂F₃H₃O that increases to $\sim 71 \text{ kcal}\cdot\text{mol}^{-1}$ higher for C₈F₁₅H₃O using TAE. On the other hand,

Joback gave enthalpies of formation that are consistently at least $10 \text{ kcal}\cdot\text{mol}^{-1}$ higher than ccCA $\Delta H_f^{298 \text{ K}}$ s even for molecules with $C < 5$. Finally, the variation in enthalpy between two consecutive PFAS-ol is $\sim 99 \text{ kcal}\cdot\text{mol}^{-1}$ for ccCA which is $\sim 1 \text{ kcal}\cdot\text{mol}^{-1}$ lower than the values obtained for the perfluoroalkanes and also agrees well with the one observed in PFAS molecules.

For the FTCA (Table 6.9), the $\Delta H_f^{298 \text{ K}}$ values computed with M06-2X and DLPNO-CBS are slightly higher than the ccCA values using the atomization approach with a maximum difference of ~ 3 and $\sim 5 \text{ kcal}\cdot\text{mol}^{-1}$, respectively. For $\text{C}_3\text{F}_3\text{H}_3\text{O}_2$, ccCA gives an enthalpy of formation of $-261.01 \text{ kcal}\cdot\text{mol}^{-1}$ while DLPNO-CBS and M06-2X enthalpies are $\sim 4 \text{ kcal}\cdot\text{mol}^{-1}$ higher than ccCA. On the other hand, for $\text{C}_5\text{F}_7\text{H}_3\text{O}_2$, the error between ccCA and M06-2X, and, ccCA and DLPNO-CBS are ~ 1 and $\sim 5 \text{ kcal}\cdot\text{mol}^{-1}$ respectively. For the longest FTCA ($C > 6$) the M06-2X enthalpies are taken as references in the absence of ccCA results and are within $\sim 2 \text{ kcal}\cdot\text{mol}^{-1}$ with the extrapolated value using ccCA. Finally, the enthalpy variation between any two consecutive molecules for ccCA and DLPNO-CBS is ~ 98.5 and $\sim 97 \text{ kcal}\cdot\text{mol}^{-1}$, respectively. While for M06-2X this variation corresponds to $\sim 99 \text{ kcal}\cdot\text{mol}^{-1}$ which is consistent with those obtained for PFAS and perfluoroalkanes.

Table 6.9: Enthalpies of formation ($\text{kcal}\cdot\text{mol}^{-1}$) for FTCA determined with the atomization approach.

	B3LYP	M06-2X	ccCA	DLPNO-CCSD(T)				Joback
				adz ^a	atz ^a	aqz ^a	CBS ^a	
C ₃ F ₃ H ₃ O ₂	-240.48	-257.69	-261.01	-173.17	-224.53	-242.93	-257.9	-251.58
C ₄ F ₅ H ₃ O ₂	-328.76	-354.95	-358.06	-246.56	-311.28	-334.9	-354.1	-352.34
C ₅ F ₇ H ₃ O ₂	-420.27	-457.60	-458.62	-323.88	-401.8	-430.41	-453.7	-453.11
C ₆ F ₉ H ₃ O ₂	-509.27	-554.82	-557.47	-399.01	-489.87	-523.52	-551.0	-553.88
			$\pm 1.75^b$					
C ₇ F ₁₁ H ₃ O ₂	-598.52	-656.12	-656.22	-474.37	-578.06	—	-648.7	-654.65
			$\pm 1.75^b$				$\pm 1.4^b$	
C ₈ F ₁₃ H ₃ O ₂	-687.80	-754.14	-755.027	-549.80	-666.35	—	-746.4	-755.41
			$\pm 1.75^b$				$\pm 1.4^b$	
C ₉ F ₁₅ H ₃ O ₂	-777.05	-852.47	-853.82	-625.13	-754.52	—	-844.1	-856.18
			$\pm 1.75^b$				$\pm 1.4^b$	

^a DLPNO-CCSD(T).anz represent a DLPNO-CCSD(T) calculation with the corresponding aug-cc-pVnZ basis set, while DLPNO-CBS is the enthalpy obtained at the complete basis set limit.

^b Extrapolated. The uncertainties were based on the root mean square deviation of the extrapolated enthalpies of formation. The methodology used for extrapolation and uncertainties are provided in appendix D.

Importantly, the reported enthalpies of formation of polyfluoroalkyl acids (FTCA), perfluoroalkyl methanol (PFAS-ol) and perfluorinated acids (PFAS) could be used as reference values for future studies focused on reaction energy and pathways, as both FTCA and PFAS-ol act as precursor molecules for the formation of PFAS at atmospheric conditions as illustrated by Ellis *et. al.*²⁷⁶

6.4 Conclusions

In this study, gas phase enthalpies of formation have been predicted for the first time for three classes of linear PFAS molecules, namely PFAS, PFAS-ol and FTCA. These enthalpies of formation are important to understand the thermochemistry of PFAS and their potential precursor molecules. The ccCA method gives the best results overall for the different classes of PFAS molecules with both the atomization and isogyric approaches and is the recommended method to compute enthalpies of formation. The DLPNO-CBS and M06-2X $\Delta H_f^{298\text{ K}}$ s were significantly dependent on the choice of thermochemical (TAE or isogyric) approach employed. The DLPNO-CCSD(T) and M06-2X

$\Delta H_f^{298\text{ K}}$ s determined with a TAE approach were always within a few $\text{kcal}\cdot\text{mol}^{-1}$ of the $\Delta H_f^{298\text{ K}}$ from ccCA and experiments. However, in the case of isogyric reactions, a substantial difference in the $\Delta H_f^{298\text{ K}}$ values are observed for both DLPNO-CBS and M06-2X methods when compared to experimental or ccCA enthalpies. Importantly, the methods selected were validated on linear alkanolic acid and perfluoroalkanes in reporting enthalpies of formation comparable to experiment, showing ccCA being the most credible method. Although, the Joback method gave very good results for the alkanolic acid, Joback predictions for the fluorinated molecules were significantly far from either experiment and/or the ccCA predictions. Based on this study, the ccCA enthalpies of formation obtained can be directly used in other studies focused on the thermochemical properties of PFAS molecules (Gibbs free energies, enthalpies of reaction). Furthermore, the relative energies for PFAS, FTCA and alkanolic acids using ccCA will be useful for future studies on reaction mechanisms, gas phase transportation of PFAS in the atmosphere, PFAS removal and mitigation.

CHAPTER 7

IONIZATION PROPERTIES OF DIATOMIC MOLECULE WITH TIME-DEPENDENT CONFIGURATION INTERACTION

7.1 Introduction

To investigate electron dynamic properties such as light absorption or emission, metal-ligand charge transfer, fluorescence, ionization processes and many others, computational chemistry methods can be instrumental. An ideal approach would be to utilize quantum electrodynamics (QED) theory,^{277,278} which treats both the electromagnetic field and molecular system at a quantum mechanical level, and also include an explicit treatment of time and a fully relativistic Hamiltonian. However, such a method becomes computationally expensive (in terms of CPU time, disk space). Furthermore, in the applications described above, nuclear motion can play an important role, and thus, approaches have been developed that do not make use of the Born-Oppenheimer approximation (BOA), which treats nuclei as static particles.²⁷⁹ As with QED, methods that do not include BOA become very computationally expensive as the system size increases. Thus, other methods need to be employed to investigate light matter interactions.

One of the most popular methods to study electron dynamic properties is time-dependent DFT (TDDFT).²⁸⁰ TDDFT was introduced in 1984 by Runge and Gross and since then has been used in many applications from small organic molecules^{281,282} to transition metal complexes.^{283–285} However, while TDDFT is very cost efficient, it is based on DFT, which means that the choice of exchange-correlation functional is very important. For example, there exist inconsistencies between functionals when comparing properties to experiment, some functionals are based on experimental parameters and thus can be only be applied to specific systems, and the DFT results can not be systematically improved by increasing the basis set size.^{286,287} Moreover, TDDFT is not a method with which electron dynamic properties are calculated. However, there is a form of DFT, real-time TDDFT (RT-TDDFT) that does enable the computation of electron dynamic properties, however, as with TDDFT, the choice of functional is fundamental. On the other hand, wavefunction-based methods have been developed such as time-dependent coupled-cluster (TD-CC)

methods, time-dependent equation of motion coupled-cluster (TD-EOMCC) methods and time-dependent configuration interaction (TDCI) methods, to study molecular systems in the presence of a strong laser pulse.^{25,288–291} However, these methods are more computationally demanding than TDDFT and smaller system are often prioritized. Kristiansen *et. al.* used TD-CC with single and double excitations to investigate the stability of TD-CC for He and Be atoms and the LiH molecule using aug-cc-pVDZ and cc-pVDZ basis sets.²⁸⁹ In 2019, Nascimento and De Prince used the time-dependent version of EOMCC at the second order (TD-EOM-CC2) to investigate the linear absorption of small organic molecules.²⁹⁰ Finally, Ulusoy and co-workers used the time-dependent configuration interaction method (TDCI) to study small molecules: H₂, LiH, CH₄, CH₂O, CH₃CN, and CH₃F in the presence of a laser pulse, determining their transition energies and dipole moments using different basis sets, gauges (as defined in section 2.6) and CI truncations.²⁵

In all of these studies, ionization processes such as those that are above and below the ionization threshold can not be described directly due to the atom-centered basis set. These basis sets can not describe the continuum of states above the ionization energy. However, approximations of the ionization treatment have been made, in particular using a heuristic model or an absorbing potential.^{29,32,292}

In this project, the time-dependent configuration interaction (TDCI) method is used to calculate the ionization rate (Γ) of H₂⁺ and N₂ using a heuristic model (see section 2.6) to simulate the above ionization potential ionization processes. This model has been implemented in the group’s TDCI code..

7.2 Ionization rate of H₂⁺

The ionization rate (Γ) of H₂⁺ has been studied previously using TDCIS with a complex absorbing potential²⁹² and a three-dimensional grid based TD-Schrödinger equation method.²⁹³ In this study, the TDCIS method is used with a heuristic model. A CIS calculation using a modified version of GAMESS US 2018²⁹⁴ is initially done in order to obtain the electronic structure information necessary for the time-propagation. In the TDCIS step, a linearly polarized laser pulse was used with a carrier frequency of 0.043 E_h/ħ (1064 cm⁻¹). The frequency corresponds to what was previously

done in other theoretical studies. The maximum field strength was set to $0.0534 E_h/(ea_0)$, which corresponds to an intensity of $1 \times 10^{14} \text{ W/cm}^2$. The cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ basis sets were used to examine the effect of larger basis set on the ionization properties. The ionization rate was calculated for internuclear distances between 2 and 9 Å (or ~ 3.8 and ~ 17.0 a.u.) with an equilibrium bond distance of 2 a.u. The molecule was aligned along the z-axis while the laser pulse was polarized along the z-axis. The results obtained were compared with those from the study of Zuo *et. al.*,²⁹³ which was based on solving the three-dimensional time-dependent Schrödinger equation (3D-TDSE) and from a study by Krause *et. al.*²⁹² In Krause's study, a complex absorption potential (CAP) was used with TDCI (TDCI-CAP) to simulate the ionization of H_2^+ . Moreover, they added some diffuse functions to an aug-cc-pVTZ basis set to make sure that there was a better overlap between the CAP and the basis set.

First, the ionization rate calculated in this study using TDCIS with the aug-cc-pVTZ basis set are compared with previous theoretical work in Figure 7.1.

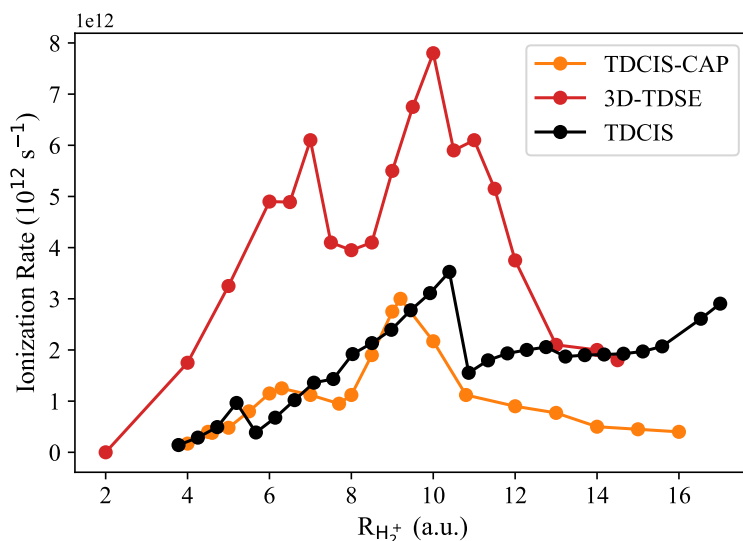


Figure 7.1: Ionization rate of H_2^+ in s^{-1} , where the laser pulse was aligned along the z axis and the carrier frequency is $0.043 E_h/\hbar$. The black line correspond to this study, the orange line was taken from Krause *et. al.*²⁹² and the red line from Zuo *et. al.*²⁹³

For the shortest H_2^+ bond distances (2 to 9 a.u.), the TDCIS results obtained in this study are in qualitative agreement (same overall increases in Γ with the bond distance increasing) with the results obtained by TDCIS-CAP. However, while the results by Krause *et. al.* show a fluctuation

of the ionization rate at $R_{H_2^+} \sim 7.8$ a.u., the results from this TDCIS study show a decrease at ~ 5.8 a.u. and a slight one at $R_{H_2^+} \sim 7.8$ a.u. On the other hand, the Γ obtained by the 3D-TDSE study for bond distances lower than 9 a.u. are higher than the one obtained in the TDCIS and TDCIS-CAP study. Furthermore, as for the TDCIS results obtained here, the Figure 7.1 shows two decreases in the ionization rate for the 3D-TDSE study: a small one at $R_{H_2^+} \sim 6.5$ a.u. and a larger one between the bond distances ~ 7 and ~ 8.5 a.u.

At $R_{H_2^+} = 9$ a.u., the ionization rate using TDCI-CAP starts to decrease until 16 a.u., however, in both the 3D-TDSE and TDCIS studies, the ionization rates start to decrease only at $R_{H_2^+} \sim 11$ a.u. and $R_{H_2^+} \sim 10.5$ a.u., respectively. While for bond distances less than 9 a.u., there was qualitative agreement (same overall change of ionization rate as the distance increases) between the previous theoretical methods and TDCIS. At bond distances greater than ~ 10.5 a.u., this is not the case. Indeed, the ionization rate, when TDCIS is used, increases between the bond distance of ~ 10.8 a.u. and 17 a.u. while for both TDCI-CAP and the 3D-TDSE methods, Γ decreases. The different behavior at larger distances can be explained by the fact that in TDCI-CAP, a medium absorbing basis was placed on a ghost atom between the hydrogen, which helps to recover the missing basis set overlap. The medium absorbing basis corresponds to a standard basis set (aug-cc-pVTZ, in Krause *et al.* study) to which a number of diffuse functions were added. In this study, at larger distances, the basis set overlap between the two hydrogens is missing which could lead to the incorrect behavior of the heuristic model. Indeed, as the distance between the hydrogen increases, more and more states exist below the ionization threshold which lead to a smaller number of states for possible ionization. However, at shorter distances between the two atoms, the basis set overlap is present and that could explain why there is a qualitative agreement between the TDCIS results in this study and the one obtained with TDCI-CAP.

To see the effect of the basis set, the ionization rate for different internuclear distances were calculated with a cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ basis set. The results are shown in Figure 7.2.

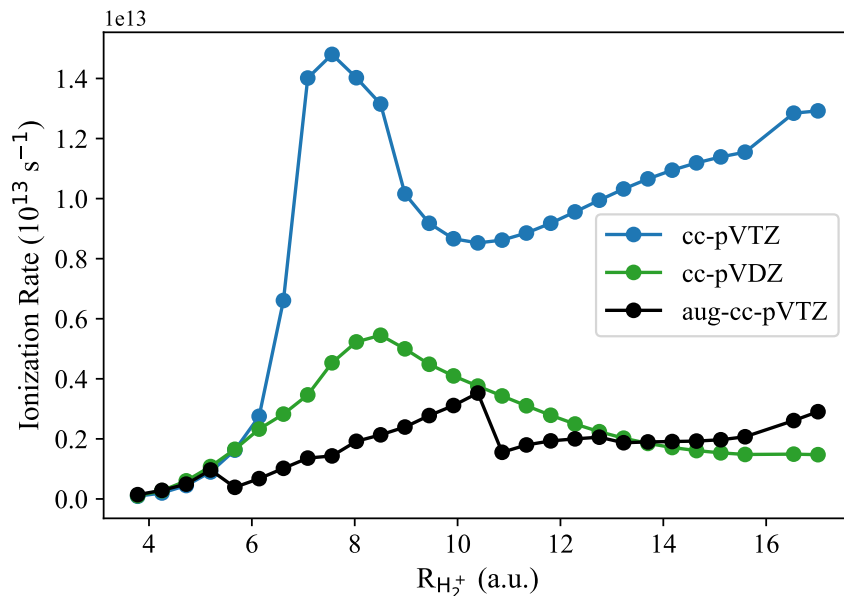


Figure 7.2: Ionization rate of H_2^+ in s^{-1} as a function of the basis set used. The laser pulse was aligned along the z axis and the carrier frequency is $0.043 E_h/\hbar$.

At very short bond distances (3.8 to 5.2 a.u.), the basis sets result in the same ionization rate showing that at these distances, the overlap between the basis function of each hydrogen is enough to correctly describe the ionization process. After 5.2 a.u., the different basis sets show very different behavior. When the cc-pVTZ basis set is used, the maximum ionization rate is $1.45 \times 10^{13} \text{ s}^{-1}$ for a bond distance of ~ 7.9 a.u., while the maximum with the cc-pVDZ basis set is $\sim 0.6 \times 10^{13} \text{ s}^{-1}$ at $R_{H_2^+} \sim 8.2$ a.u. Once the ionization rate maximum has been reached, the results from the cc-pVDZ basis set slowly decreases as the TDCI-CAP results by Krause and co-workers showed.²⁹² On the other hand, the ionization rate with the cc-pVTZ as well as the aug-cc-pVTZ basis set increases until the longest distances. The improvement between cc-pVTZ and aug-cc-pVTZ is quite significant and shows the importance of diffuse functions.

The results obtained for H_2^+ in this study do not perfectly reproduce previous theoretical results. However, it is important to note that the heuristic model is a quick but qualitative method to include ionization effects. In order to obtain more insight about the method, the ionization rate of the N_2 molecule was studied.

7.3 Ionization rate of N₂

In 2007, the ionization rate of N₂ was studied experimentally by Pavičić *et. al.* using a Ti:sapphire laser system.²⁹⁵ The effect of the angle between the incident pulse and the molecule on the ionization rate was investigated. Theoretically, there have been different studies using TD-CIS-CAP, TD-DFT or the molecular strong-field ionization.^{296–298} Sissay and co-workers used real time TD-DFT with the long range corrected PBE functional (LC-PBE) with an aug-cc-pVTZ basis set and added a complex absorption potential (CAP) to capture ionization.²⁹⁷ On the other hand, TDCI was also employed by Hoerner and co-workers where they used a CAP to model the ionization. In their study, different field strengths and their effects on the ionization rate were examined.²⁹⁶ Finally, an earlier study by Kjeldsen *et. al.*, utilized molecular strong-field ionization methods, molecular Ammosov-Delone-Krainov (MO-ADK) model and molecular strong-field approximation (MO-SFA) which are based on tunneling theory.²⁹⁸

To determine the ionization properties of N₂, it is important to obtain the correct set of orbitals and the correct energy order. Since ionization will mostly occur from the Highest Occupied Molecular Orbital (HOMO), it is important for the theory to capture the correct HOMO. For N₂ the valence orbital should have the following order $E_\pi = E_\pi < E_\sigma$, where σ corresponds to the HOMO.

In this project, the change of ionization rate for N₂ is investigated when the angle between the molecular axis and the laser pulse changes. The external field intensity was set to 0.053 a.u. or 10^{14} W/cm⁻¹ and the frequency to 0.057 a.u. or 800 nm which corresponds to the frequency in work by Sissay *et. al.*²⁹⁷ Here, the equilibrium geometry (1.10 Å) was obtained with RCCSD(T) and the aug-cc-pVQZ basis set using Molpro 2020.¹⁴⁷ The molecule was aligned along the z -axis and the ionization rate was calculated as the ratio between the ionization rate when the pulse is not aligned along the z -axis and when it is. All of the single point (MCSCF and CISD) calculations were done using the aug-cc-pVTZ basis set. An active space composed of ten electrons in twelve orbitals was used, which correspond to: two $2s^2$, and the $2p_x^2$, $2p_y^2$ and $2p_z^2$ for the occupied orbitals and the other $2p_x^0$, $2p_y^0$ and $2p_z^0$, one $3s^0$ and one set of $3p_{x,y,z}$ for the virtual orbitals. An initial complete active space self-consistent field (CASSCF) was run to obtain optimized orbitals for a subsequent CISD

calculation. This CISD calculation was carried out to produce the initial CI states coefficients for the TDCI propagation. Finally, a RASSCF calculation was tested using a two holes/two electrons RAS1/RAS3 space and four orbitals in the RAS2 active space (see section 2.5). This calculation corresponds to six electrons in six orbitals.

First, the ionization rate ratio obtained with the standard TDCI steps is compared to experiment and previous theoretical results (TD-CIS-CAP, TD-DFT and MO methods).

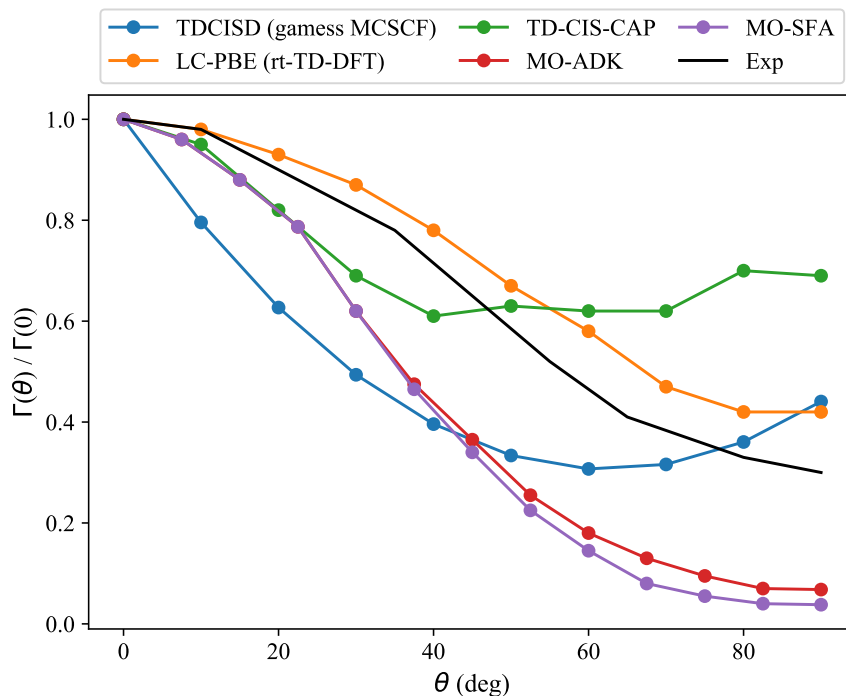


Figure 7.3: Ionization rate of N_2 at different angles between the laser pulse and the molecule axis. The ionization is calculated as a ratio between $\Gamma(\theta)/\Gamma(0^\circ)$ and compared with theoretical (LC-PBE results from Sissay²⁹⁷, MO-SFA and MO-ADK results from Kjeldsen²⁹⁸ and TD-CIS-CAP results from Hoerner²⁹⁶) and experimental data from Pavičić *et. al.*²⁹⁵

Figure 7.3 depicted the ionization rate ratio obtained in this study compared to experiment and prior theoretical results. As shown, the result obtained with TDCISD is not in agreement with experiment. For the smaller angles, (between the laser pulse polarization and the molecule axis) from 0° to 40° , both the experimental and TDCISD ionization rate ratio decreases. However, while the experimental ionization rate still decreases after 40° , the TDCISD ionization starts to decrease much more slowly and at 60° the ionization rate ratio increases. The incorrect behavior at a larger angle can be attributed to an unreasonable set of orbitals. Indeed, in this calculation, the MCSCF

orbitals used for the CISD calculation, showed that the two non-degenerate occupied π orbitals are the HOMO while the σ_z orbital is lower in energy. Since the heuristic model depends on the HOMO energy to define the ionization potential, the fact that the orbital ordering is wrong is likely to lead to incorrect results. The MCSCF orbitals leads to an ionization potential (IP) of ~ 16.12 eV while the experimental value reported by NIST is ~ 15.6 eV.²³⁴ This difference between the calculated and the experimental IP's means that more states are ionized in the calculation than necessary.

The TD-CIS-CAP results by Hoerner *et. al.* show an abrupt increase in the ionization rate at 40° and is overall in poorer agreement with experiment than the results obtained in this study. On the other hand, both molecular strong-field ionization results by Kjeldsen *et. al.* show an overall decrease of the ionization rate ratio which is consistent with experiment. The initial wavefunction in MO-SFA and MO-ADK were taken as an effective one-electron wavefunction taken from a Hartree-Fock approximation. Finally, the real time TD-DFT result by Sissay and co-workers showed the best agreement with experiment. Overall the same shape as the experimental results was obtained with a slightly higher rate ratio.

Since the standard TDCI procedure using GAMESS US was not able to produce the correct ionization rate ratio for larger angles, other procedures have been considered. First, the LC-WB97 functional was used to generate initial guess orbitals for a subsequent MCSCF calculation using GAMESS.²⁹⁹ For the second approach, Molpro 2020 was used to generate the MCSCF orbitals which were then used in GAMESS to run the CISD calculation.¹⁴⁷ Finally, the RASSCF, restrictive active space state interaction (RASSI) and SO-RASSI methods from the OpenMolcas software (v. 21.06) were utilized to generate the CI states coefficient to run the TDCISD simulation.³⁰⁰ The SO coupling in SO-RASSI is included *a posteriori* with the spin-orbit part of the Douglas-Kroll Hamiltonian. The results arising from these different methods are presented in Figure 7.4. Furthermore, Table 7.1 gives the ionization potential and corresponding HOMO for each methods described above.

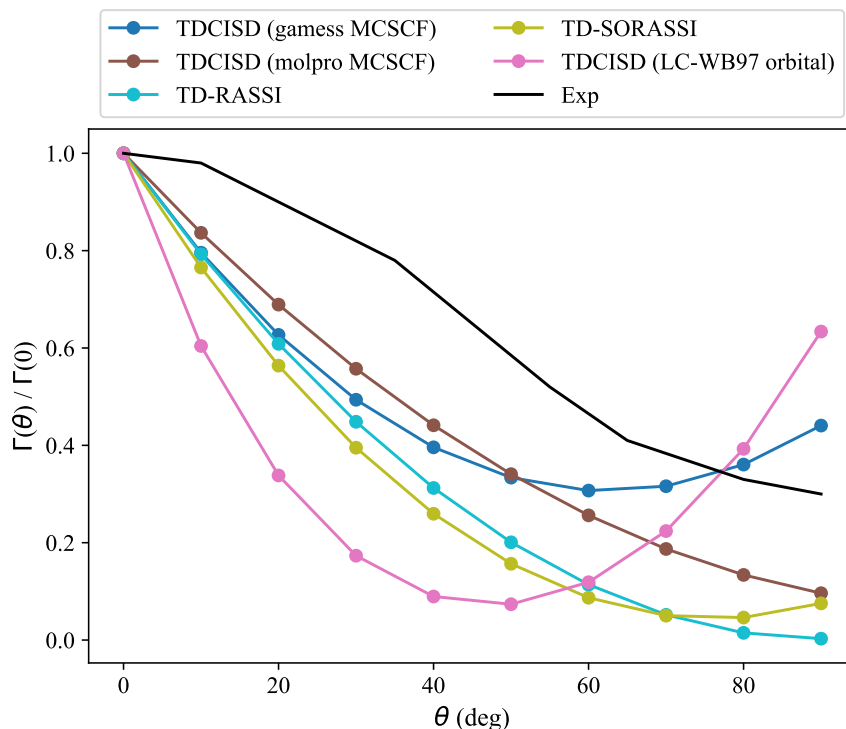


Figure 7.4: Ionization rate of N_2 at different angles between the laser pulse and the molecule axis. The ionization is calculated as a ratio between the ionization when the laser is not aligned with the molecule and when it is ($\Gamma(\theta)/\Gamma(0^\circ)$). Several procedures are reported, each based upon different orbital selections and MCSCF approaches. The experimental data were taken from Pavičić *et. al.*²⁹⁵

Table 7.1: Ionization potential (IP) of N_2 depending on the method used to generate the orbitals. All calculations were done using the aug-cc-pVTZ basis set. The experimental data is taken from the NIST database.²³⁴

Method	Ionization Potential (eV)	HOMO
MCSCF (HF orbitals)	16.12	π
MCSCF (LC-WB97 orbitals)	15.12	π
MCSCF (Molpro)	14.11	σ_z
RASSCF	15.98	π
Exp. ²³⁴	~ 15.5 to ~ 15.7	—

As Table 7.1 shows, the RASSCF method results in the best IP compared to experiment with an error of less than 0.3 eV while the MCSCF orbitals from Molpro lead to the largest error (~ 1.4 eV). However, so far, only Molpro leads to the correct orbital for the HOMO. Moreover, while the use of

DFT orbitals as guess orbitals for MCSCF reduced the IP from 16.12 eV to 15.12 eV, the HOMOs are still the π orbitals.

The ionization rate resulting from these different simulations are presented in Figure 7.4. The use of DFT orbitals as an initial guess did not improve the ionization rate as compared to the HF orbital initial guess. In fact, the overall shape with the DFT orbitals is worse than that of the HF orbitals as compared to experiment. The ionization rate ratio decreases and increases much faster when the LC-WB97 orbitals are used. On the other hand, when Molpro is used to generate the MCSCF orbitals to use in the CISD calculation, the ionization rate is in much better agreement with experiment with overall decreases of the ratio. Even though the ratio is smaller when Molpro is used as compared to the ratio from experiment, the difference between theory and experiment is roughly the same for each angle. The same behavior was observed when the RASSCF and RASSI modules from OpenMolcas were used. When the RASSI states coefficients were propagated, the ionization rates obtained were very similar to the ones obtained with Molpro. In addition, the inclusion of spin-orbit coupling in RASSI (TD-SORASSI) leads to the same results as the calculations that have not included SO coupling. Interestingly, the Molpro MCSCF and both RASSI results are very close to the MO-ADK and MO-SFA results presented in Figure 7.3. They show the same trends and values for the ionization ratio as the angle increases.

7.4 Conclusion

Overall, the heuristic model shows good qualitative results as compared to other theoretical or experimental results. For the N_2 molecule, the ionization rate ratio obtained at the TDCISD depends heavily on the orbital used at the MCSCF level. Indeed, the HF and LC-WB87 orbitals both lead to the incorrect behavior of the ionization rate ratio compared to experiment. On the other hand, when the MCSCF orbitals from Molpro or the RASSCF method are used, the results are in much better agreement with experiment. Overall, the heuristic model provides an efficient way to describe the ionization process in TDCI theory. It is important to note that the heuristic model used in this work was developed for TDCIS and not for TDCISD. A development for TDCISD was given by Coccia *et. al.* and will need to be implemented in the group TDCI code.³⁰¹

For smaller intramolecular distances between the two hydrogens, the TDCIS ionizations rate are in agreement with previous theoretical results. However, for longer distances this is not the case, and with the aug-cc-pVTZ basis set, incorrect behavior for the ionization rate is obtained. Moreover, the importance of the diffuse functions in the basis set was demonstrated for the cc-pVTZ and aug-cc-pVTZ basis set.

CHAPTER 8

FUTURE DIRECTIONS

In this thesis, a number of computational strategies have been employed to investigate the gas phase thermodynamic and spectroscopic properties of PFAS and heavy element complexes. The *rp-ccCA* approach has been successfully applied to the *5d* transition metal, while *f-ccCA* has been gauged against the latest experimental bond dissociation energy by Micheal Morse group. Moreover, ground and excited states potential energy curves of LuF have been characterized by MRCI methods and compared to experiment. Finally, the linear PFAS gas phase enthalpies of formation have been investigated for the first time with both *ab initio* and DFT methods.

In looking forward for the *5d* projects, improvement can be made in the *rp-ccCA* procedure. First, a better inclusion of spin-orbit coupling would be required to obtain more consistent results among different metals. There are many ways to introduce a better SO for the molecule, but among existing methods, an MRCI calculation with the Breit-Pauli Hamiltonian should be considered. This calculation will give insight into the multireference character of the compound and will give a molecular SO coupling energy. Another option is to use the Dirac Hartree-Fock (DHF) Hamiltonian available via the DIRAC software which corresponds to a four-component Hamiltonian. This Hamiltonian allows a full description of both the scalar and spin-orbit relativistic effects which is needed for heavy elements. However, including dynamic correlation through the use of coupled-cluster or MP2 methods will become very expensive for larger molecules as DHF is already expensive as compared to its two-component and pseudopotential counterparts (see Chapter 4). Since *5d* transition metal complexes have the potential to be highly multireference in nature, a more rigorous investigation of the multireference character (more than one Slater determinant to build the wavefunction) might be needed, especially for smaller systems as the number of unpaired electron increases and thus the potential number of possible Slater determinant, also increases.

Finally, DFT shows inconsistent results depending on the functional and the molecule studied. Thus, a more in depth investigation of the effect of functional choice on thermochemical properties may be important. The double-hybrid functional which includes a MP2 correlation energy component

could help when the exact exchange from the H-GGA functional is not enough. As well, the recently developed multiconfiguration paired-DFT (MC-PDFT) theory by Gagliardi group may be useful for some molecules.³⁰² This somewhat recent DFT approach is based on an initial CASSCF or RASSCF calculation to obtain a multireference wavefunction that is used in a DFT calculation.

In the lanthanide study describe herein, lanthanide diatomics have been investigated using wavefunction based methods. As the project on the LnS and LnSe molecules showed, single-reference methods can fail to accurately predict the dissociation energies of some diatomics. Thus, multireference composite schemes will be needed, in particular using a CASSCF or MRCI initial wavefunction. While the multireference path has already been included at the DFT level, almost no work has been done with a wavefunction-based composite approach for heavier elements. However, as with any active space type of methods, the selection of a proper active space will be the bottleneck of the method as explained in Chapter 2.

However, most applications for the lanthanides involve much larger ligands. While it is important to understand the fundamental properties of lanthanides, wavefunction methods that can accurately describe larger lanthanide complex properties are needed. Previously, a more cost-efficient composite approach as compared to ccCA was developed called DLPNO-ccCA.²⁶³ This method was developed for organic molecules but could be modified to study heavier elements. DLPNO-ccCA becomes more and more cost advantageous as compared with ccCA as the size of the molecule increases. However, approximations are made in the DLPNO method and their effects on the thermochemical properties need to be tested.

For the PFAS project, the gas phase enthalpies of formation have been calculated for a set of small to medium linear alkanoic acids, fluoroalkanes and three types of PFAS. Since not much is known in terms of the thermochemical properties of PFAS, more studies are needed. For example, a limited number of experimental papers have been published on possible PFAS degradation pathways involving the molecules from Chapter 6. Further studies should be done to determine if the methods used in Chapter 6 can properly describe the degradation mechanism of these PFAS. Second, as the size of the molecule increases, it will become increasingly difficult to use the standard ccCA

method and thus a more cost efficient method like DLPNO-ccCA needs to be considered. Initial calculations of gas phase $\Delta H_f^{298\text{ K}}$ with DLPNO-ccCA have been conducted for the same set of molecules presented in Chapter 6. Finally, while gas phase properties are important, most PFAS study are conducted in solvent. Among the aqueous properties, pK_a is of paramount importance as it provides insight about the PFAS acidity in water or other solvent. Such a project was begun for PFAS molecule but initial results were not consistent considering different methods, basis sets and implicit solvents. With the knowledge gained about PFAS properties during the study included in this dissertation, it will be easier to pursue properties in aqueous solutions.

Finally, for the TDCI project, more work needs to be done, in particular for transition metal and heavier elements. So far, GAMESS US has been used to generate the CI coefficients and states energies for the TDCI propagation. However, there are constraints from this software especially for lanthanides. First of all, when a Douglas-Kroll-Hess Hamiltonian is used, the basis set size can not exceed a cc-pVTZ with the utilized software package. Secondly, if a pseudopotential (ECP) is used, the highest angular momentum component allowed is "g". It means that if a pseudopotential is build using "h" or higher angular momentum, the ECP can not be used directly. However, diffuse functions play an important role in the description of light absorption properties and the lanthanide pseudopotentials include the "h" angular momentum. While in theory, it would be possible to modify the GAMESS US source code so that it could use larger basis sets, other software options have been investigated, in particular, the RASSCF and RASSI module of the OpenMolcas software. So far, both modules have been tested for N_2 and RASSCF and RASSI are being tested for LuF. Since LuF has been extensively studied by Wilson's group, data are available for the ground and different excited state, which can be compared to the results from RASSCF and RASSI. Furthermore, one advantage of the OpenMolcas software is that it handles symmetry such that each symmetries can be investigated separately. Thus, larger systems can be studied, in particular, for metal-ligand charge transfer complexes. Finally, for the ionization rate, the CISD heuristic model by Coccia *et. al.* needs to be implemented in the developed code.³⁰¹ This model should allow a better description of ionization rates with TDCISD.

BIBLIOGRAPHY

- [1] Schrödinger, E. Quantisierung als Eigenwertproblem. *Ann. Phys.* **1926**, 384, 361–376.
- [2] Heisenberg, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Phys.* **1927**, 43, 172–198.
- [3] Heitler, W.; London, F. Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik. *Zeitschrift für Phys.* **1927**, 44, 455–472.
- [4] Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, 45, 4256–4266.
- [5] Löwdin, P.-O. *Adv. Chem. Phys.*; Advances in Chem. Phys.; 1958; pp 207–322.
- [6] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, 136, B864–B871.
- [7] Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, 140, A1133–A1138.
- [8] Møller, C.; Plesset, M. Note on an approximation for many-electrons systems. *Phys. Rev.* **1934**, 46, 618–622.
- [9] Dunning, T. H.; Peterson, K. A. Use of Møller-Plesset perturbation theory in molecular calculations: Spectroscopic constants of first row diatomic molecules. *J. Chem. Phys.* **1998**, 108, 4761–4771.
- [10] Piecuch, P.; Kowalski, K.; Pimienta, I. S. O.; McGuire, M. J. Recent advances in electronic structure theory: Method of moments of coupled-cluster equations and renormalized coupled-cluster approaches. *Int. Rev. Phys. Chem.* **2002**, 21, 527–655.
- [11] Fermi, E. Un metodo statistico per la determinazione di alcune prioprieta dell’atomo. *Rend. Acad. Dei Lincei* 6, 602–607.
- [12] L.H.Thomas, The calculation of atomic fields. *Math. Proc. Cambridge Philos. Soc.* **1927**, 23, 542–548.
- [13] Gritsenko, O. V.; Mentel, Ł. M.; Baerends, E. J. On the errors of local density (LDA) and generalized gradient (GGA) approximations to the Kohn-Sham potential and orbital energies. *J. Chem. Phys.* **2016**, 144, 204114.
- [14] Furche, F.; Perdew, J. P. The performance of semilocal and hybrid density functionals in 3d transition-metal chemistry. *J. Chem. Phys.* **2006**, 124, 044103.
- [15] Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, 33, 8822–8824.
- [16] Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, 98, 1372–1377.

- [17] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. General performance of density functionals. *J. Phys. Chem. A* **2007**, *111*, 10439–10452.
- [18] Kowalski, K.; Piecuch, P. The method of moments of coupled-cluster equations and the renormalized CCSD[T], CCSD(T), CCSD(TQ), and CCSDT(Q) approaches. *J. Chem. Phys.* **2000**, *113*, 18–35.
- [19] McWeeny, R.; Coulson, C. A. The density matrix in self-consistent field theory - III. Generalizations of the theory. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **1957**, *241*, 239–256.
- [20] Malmqvist, P. A.; Rendell, A.; Roos, B. O. The restricted active space self-consistent-field method, implemented with a split graph unitary group approach. *J. Phys. Chem.* **1990**, *94*, 5477–5482.
- [21] Meyer, H.-D.; Manthe, U.; Cederbaum, L. S. The multi-configurational time-dependent Hartree approach. *Chem. Phys. Lett.* **1990**, *165*, 73–78.
- [22] Hochstuhl, D.; Bauch, S.; Bonitz, M. Multiconfigurational time-dependent Hartree-Fock calculations for photoionization of one-dimensional Helium. *J. Phys. Conf. Ser.* **2010**, *220*, 12019.
- [23] Sato, T.; Ishikawa, K. L. Time-dependent complete-active-space self-consistent-field method for multielectron dynamics in intense laser fields. *Phys. Rev. A* **2013**, *88*, 23402.
- [24] Dalgaard, E.; Monkhorst, H. J. Some aspects of the time-dependent coupled-cluster approach to dynamic response functions. *Phys. Rev. A* **1983**, *28*, 1217–1222.
- [25] Ulusoy, I. S.; Stewart, Z.; Wilson, A. K. The role of the CI expansion length in time-dependent studies. *J. Chem. Phys.* **2018**, *148*, 014107.
- [26] List, N. H.; Melin, T. R. L.; van Horn, M.; Saue, T. Beyond the electric-dipole approximation in simulations of x-ray absorption spectroscopy: Lessons from relativistic theory. *J. Chem. Phys.* **2020**, *152*, 184110.
- [27] Lestrange, P. J.; Hoffmann, M. R.; Li, X. In *Nov. Electron. Struct. Theory Gen. Innov. Strongly Correl. Syst.*; Hoggan, P. E., Ed.; Advances in Quantum Chemistry; Academic Press, 2018; Vol. 76; pp 295–313.
- [28] Klamroth, T. Laser-driven electron transfer through metal-insulator-metal contacts: Time-dependent configuration interaction singles calculations for a jellium model. *Phys. Rev. B* **2003**, *68*, 245421.
- [29] Coccia, E.; Mussard, B.; Labeye, M.; Caillat, J.; Taïeb, R.; Toulouse, J.; Luppi, E. Gaussian continuum basis functions for calculating high-harmonic generation spectra. *Int. J. Quantum Chem.* **2016**, *116*, 1120–1131.
- [30] Awasthi, M.; Vanne, Y. V.; Saenz, A. Non-perturbative solution of the time-dependent Schrödinger equation describing H₂ in intense short laser pulses. *J. Phys. B At. Mol. Opt. Phys.* **2005**, *38*, 3973–3985.

- [31] Scrinzi, A.; Piraux, B. Two-electron atoms in short intense laser pulses. *Phys. Rev. A* **1998**, *58*, 1310–1321.
- [32] Klinkusch, S.; Saalfrank, P.; Klamroth, T. Laser-induced electron dynamics including photoionization: A heuristic model within time-dependent configuration interaction theory. *J. Chem. Phys.* **2009**, *131*, 114304.
- [33] Slater, J. Atomic Shielding Constants. *Phys. Rev.* **1930**, *36*, 57.
- [34] Boys, S. F. Electronic Wave Functions . I . A General Method of Calculation for the Stationary States of Any Molecular. *Proc. R. Soc. Lond. A. Math. Phys. Sci.* **1950**, *200*, 542–554.
- [35] Dunning, T. H. Gaussian basis functions for use in molecular calculations. III. Contraction of (10s6p) atomic basis sets for the first-row atoms. *J. Chem. Phys.* **1971**, *55*, 716–723.
- [36] Raffenetti, R. C. General contraction of Gaussian atomic orbitals: Core, valence, polarization, and diffuse basis sets; Molecular integral evaluation. *J. Chem. Phys.* **1973**, *4452*, 4452–4458.
- [37] Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent molecular-orbital methods. I. Use of gaussian expansions of slater-type atomic orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
- [38] Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
- [39] Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- [40] Pou-Amerigo, R.; Merchan, M.; Netob-Gil, I.; Widmark, P.-O.; Roos, B. O. Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions III. First row transition metal atoms. *Theor. Chim. Acta* **1995**, *92*, 149–181.
- [41] Dirac, P. A. M. The Quantum Theory of the Emission and Absorption of Radiation. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1927**, *114*, 243–265.
- [42] Pauli Jr., W. Zur Quantenmechanik des magnetischen Elektrons. *Zeitschrift für Phys.* **1927**, *43*, 601–623.
- [43] Van Lenthe, E.; Baerends, E. J.; Snijders, J. G. Relativistic total energy using regular approximations. *J. Chem. Phys.* **1994**, *101*, 9783–9792.
- [44] Iliáš, M.; Saue, T. An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation. *J. Chem. Phys.* **2007**, *126*, 064102.
- [45] Douglas, M.; Kroll, N. M. Quantum electrodynamical corrections to the fine structure of Helium. *Ann. Phys. (N. Y.)* **1974**, *155*, 89–155.
- [46] Dolg, M.; Cao, X. Relativistic pseudopotentials: Their development and scope of applications. *Chem. Rev.* **2012**, *112*, 403–480.

- [47] Figgen, D.; Peterson, K. A.; Dolg, M.; Stoll, H. Energy-consistent pseudopotentials and correlation consistent basis sets for the 5d elements Hf-Pt. *J. Chem. Phys.* **2009**, *130*, 164108.
- [48] Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 theory: A general procedure for prediction of molecular energies. *J. Chem. Phys.* **1989**, *90*, 5622–5629.
- [49] Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221–7230.
- [50] Nyden, M. R.; Petersson, G. A. Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions. *J. Chem. Phys.* **1981**, *75*, 1843–1862.
- [51] Martin, J. M.; De Oliveira, G. Towards standard methods for benchmark quality *ab initio* thermochemistry - W1 and W2 theory. *J. Chem. Phys.* **1999**, *111*, 1843–1856.
- [52] Daniel Boese, A.; Oren, M.; Atasoylu, O.; Martin, J. M.; Kállay, M.; Gauss, J. W3 theory: Robust computational thermochemistry in the kJ/mol accuracy range. *J. Chem. Phys.* **2004**, *120*, 4129–4141.
- [53] Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated *ab initio* thermochemistry. *J. Chem. Phys.* **2004**, *121*, 11599–11613.
- [54] Feller, D.; Peterson, K. A.; Dixon, D. A. A survey of factors contributing to accurate theoretical predictions of atomization energies and molecular structures. *J. Chem. Phys.* **2008**, *129*, 204105.
- [55] DeYonker, N. J.; Grimes, T.; Yockel, S.; Dinescu, A.; Mintz, B.; Cundari, T. R.; Wilson, A. K. The correlation-consistent composite approach: Application to the G3/99 test set. *J. Chem. Phys.* **2006**, *125*, 104111.
- [56] DeYonker, N. J.; Peterson, K. A.; Steyl, G.; Wilson, A. K.; Cundari, T. R. Quantitative Computational Thermochemistry of Transition Metal Species. *J. Phys. Chem. A* **2007**, *111*, 11269–11277.
- [57] DeYonker, N. J.; Williams, T. G.; Imel, A. E.; Cundari, T. R.; Wilson, A. K. Accurate thermochemistry for transition metal complexes from first-principles calculations. *J. Chem. Phys.* **2009**, *131*, 024106.
- [58] Williams, T. G.; DeYonker, N. J.; Wilson, A. K. Hartree-Fock complete basis set limit properties for transition metal diatomics. *J. Chem. Phys.* **2008**, *128*, 044101.
- [59] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-set convergence of the energy in molecular Hartree-Fock calculations. *Chem. Phys. Lett.* **1999**, *302*, 437–446.
- [60] Peterson, K. A.; Woon, D. E.; Dunning, T. H. Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the $\text{H}+\text{H}_2 \rightarrow \text{H}_2+\text{H}$ reaction. *J. Chem. Phys.* **1994**, *100*, 7410–7415.

- [61] Laury, M. L.; DeYonker, N. J.; Jiang, W.; Wilson, A. K. A pseudopotential-based composite method: The relativistic pseudopotential correlation consistent composite approach for molecules containing 4d transition metals (Y-Cd). *J. Chem. Phys.* **2011**, *135*, 214103.
- [62] Welch, B. K.; Wilson, A. K. f-ccCA: Developing a the correlation consistent Composite Approach (ccCA) Thermochemical Protocol for Energetic Predictions of the Lanthanide Species. *To be submitted*
- [63] Almeida, N. M.; Melin, T. R.; North, S. C.; Welch, B. K.; Wilson, A. K. *Ab initio* composite strategies and multireference approaches for lanthanide sulfides and selenides. *J. Chem. Phys.* **2022**, *157*, 024105.
- [64] Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. Molecular Orbital Theory of the Electronic Structure of Organic Compounds. V. Molecular Theory of Bond Separation. *J. Am. Chem. Soc.* **1970**, *92*, 4796–4801.
- [65] Hehre, W. J.; Radom, L.; Scheleyer, P.; Pople, J. A. *Ab initio molecular orbital theory*; Wiley, 1986.
- [66] Simonenko, E. P.; Sevast'yanov, D. V.; Simonenko, N. P.; Sevast'yanov, V. G.; Kuznetsov, N. T. Promising ultra-high-temperature ceramic materials for aerospace applications. *Russ. J. Inorg. Chem.* **2013**, *58*, 1669–1693.
- [67] Cortright, R. D.; Davda, R. R.; Dumesic, J. A. Hydrogen from catalytic reforming of biomass-derived hydrocarbons in liquid water. *Nature* **2002**, *418*, 964–967.
- [68] Chase, M. NIST-JANAF Thermochemical Tables, 4th Edition. *J. Phys. Chem. Ref. Data, Monogr.* **1998**, *9*, 1–1952.
- [69] Yungman, V. S.; Glushko, *Thermal constants of substances*; Begell House, 1999.
- [70] Krasnov, K. *Molecular Constants of Inorganic Compounds*, reference ed.; 1979.
- [71] Gurvich, V.; Veyts, I.; Alock, C. B. *Thermodynamics properties of individual substances*; 1994; p 12.
- [72] Krasnov, K.; Morozov, E.; Filippenko, N.; Giricheva, N. Thermodynamic characteristics of hafnium and thorium halides. *Izv. Vyss. Uchebn. Zaved. Pish. Tekhnol* **1973**, *16*, 1500–1503.
- [73] Sorensen, J. J.; Persinger, T. D.; Sevy, A.; Franchina, J. A.; Johnson, E. L.; Morse, M. D. Bond dissociation energies of diatomic transition metal selenides: TiSe, ZrSe, HfSe, VSe, NbSe, and TaSe. *J. Chem. Phys.* **2016**, *145*, 214308.
- [74] Morse, M. D. Predissociation Measurements of Bond Dissociation Energies. *Acc. Chem. Res.* **2019**, *52*, 119–126.
- [75] Sanderson, R. T. The Interrelation of Bond Dissociation Energies and Contributing Bond Energies. *J. Am. Chem. Soc.* **1975**, *97*, 1367–1372.

- [76] Wheeler, S. E. Homodesmotic reactions for thermochemistry. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, 2, 204–220.
- [77] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *J. Chem. Phys.* **1998**, 109, 7764–7776.
- [78] Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. *J. Chem. Phys.* **2006**, 125, 144108.
- [79] Chan, B.; Radom, L. W2X and W3X-L: Cost-Effective Approximations to W2 and W4 with kJ mol⁻¹ Accuracy. *J. Chem. Theory Comput.* **2015**, 11, 2109–2119.
- [80] Sylvetsky, N.; Peterson, K. A.; Karton, A.; Martin, J. M. L. Toward a W4-F12 approach: Can explicitly correlated and orbital-based *ab initio* CCSD(T) limits be reconciled? *J. Chem. Phys.* **2016**, 144, 214101.
- [81] Bomble, Y. J.; Vázquez, J.; Kállay, M.; Michauk, C.; Szalay, P. G.; Császár, A. G.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated *ab initio* thermochemistry. II. Minor improvements to the protocol and a vital simplification. *J. Chem. Phys.* **2006**, 125, 064108.
- [82] Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated *ab initio* thermochemistry. III. Additional improvements and overview. *J. Chem. Phys.* **2008**, 128, 114111.
- [83] Manivasagam, S.; Laury, M. L.; Wilson, A. K. Pseudopotential-Based Correlation Consistent Composite Approach (rp-ccCA) for First- and Second-Row Transition Metal Thermochemistry. *J. Phys. Chem. A* **2015**, 119, 6867–6874.
- [84] Craciun, R.; Picone, D.; Long, R. T.; Li, S.; Dixon, D. A.; Peterson, K. A.; Christe, K. O. Third row transition metal hexafluorides, extraordinary oxidizers, and lewis acids: Electron affinities, fluoride affinities, and heats of formation of WF₆, ReF₆, OsF₆, IrF₆, PtF₆, and AuF₆. *Inorg. Chem.* **2010**, 49, 1056–1070.
- [85] Minenkov, Y.; Sliznev, V. V.; Cavallo, L. Accurate Gas Phase Formation Enthalpies of Alloys and Refractories Decomposition Products. *Inorg. Chem.* **2017**, 56, 1386–1401.
- [86] Sevy, A.; Sorensen, J. J.; Persinger, T. D.; Franchina, J. A.; Johnson, E. L.; Morse, M. D. Bond dissociation energies of TiSi, ZrSi, HfSi, VSi, NbSi, and TaSi. *J. Chem. Phys.* **2017**, 147, 084301.
- [87] Schimmelpfennig, B.; Wahlgren, U.; Gropen, O.; Haaland, A. The gas phase structures of tungsten chlorides: density functional theory calculations on WCl₆, WCl₅, WCl₄, WCl₃ and W₂Cl₆. *J. Chem. Soc. Dalton Trans.* **2001**, 1616–1620.
- [88] Thanthiriwatte, K. S.; Vasiliu, M.; Battey, S. R.; Lu, Q.; Peterson, K. A.; Andrews, L.; Dixon, D. A. Gas Phase Properties of MX₂ and MX₄ (X = F, Cl) for M = Group 4, Group 14, Cerium, and Thorium. *J. Phys. Chem. A* **2015**, 119, 5790–5803.

- [89] Fang, Z.; Dixon, D. A. Hydrolysis of ZrCl_4 and HfCl_4 : The initial steps in the high-temperature oxidation of metal chlorides to produce ZrO_2 and HfO_2 . *J. Phys. Chem. C* **2013**, *117*, 7459–7474.
- [90] Fang, Z.; Lee, Z.; Peterson, K. A.; Dixon, D. A. Use of Improved Orbitals for CCSD(T) Calculations for Predicting Heats of Formation of Group IV and Group VI Metal Oxide Monomers and Dimers and UCl_6 . *J. Chem. Theory Comput.* **2016**, *12*, 3583–3592.
- [91] Jiang, W.; Deyonker, N. J.; Wilson, A. K. Multireference character for 3d transition-metal-containing molecules. *J. Chem. Theory Comput.* **2012**, *8*, 460–468.
- [92] Peterson, C.; Penchoff, D. A.; Wilson, A. K. Prediction of Thermochemical Properties Across the Periodic Table: A Review of the correlation consistent Composite Approach (ccCA) Strategies and Applications. *Annu. Rep. Comput. Chem.* **2016**, *12*, 3–45.
- [93] Jorgensen, K. R.; Wilson, A. K. Enthalpies of formation for organosulfur compounds: Atomization energy and hypohomodesmotic reaction schemes via *ab initio* composite methods. *Comput. Theor. Chem.* **2012**, *991*, 1–12.
- [94] Wilson, B. R.; Deyonker, N. J.; Wilson, A. K. Prediction of hydrocarbon enthalpies of formation by various thermochemical schemes. *J. Comput. Chem.* **2012**, *33*, 2032–2042.
- [95] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- [96] Aebbersold, L. E.; Yuwono, S. H.; Schoendorff, G.; Wilson, A. K. Efficacy of Density Functionals and Relativistic Effective Core Potentials for Lanthanide-Containing Species: The Ln_{54} Molecule Set. *J. Chem. Theory Comput.* **2017**, *13*, 2831–2839.
- [97] Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- [98] Peterson, K. A.; Shepler, B. C.; Figgen, D.; Stoll, H. On the spectroscopic and thermochemical properties of ClO , BrO , IO , and their anions. *J. Phys. Chem. A* **2006**, *110*, 13877–13883.
- [99] Perdew, J. P. Jacob’s ladder of density functional approximations for the exchange-correlation energy. **2001**, *1*, 1–20.
- [100] Slater, J. C.; Johnson, K. H. Self-consistent-field $X\alpha$ cluster method for polyatomic molecules and solids. *Phys. Rev. B* **1972**, *5*, 844–853.
- [101] Becke, A. D. Density-functional exchange approximation with correct asymptotic behaviour. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- [102] Perdew, J. P.; Burke, K.; Wang, Y. Generalized gradient approximation for the exchange-correlation hole of a many-electron system. *Phys. Rev. B* **1996**, *54*, 16533–16539.
- [103] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

- [104] Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101.
- [105] Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- [106] Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554–8560.
- [107] Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129–12137.
- [108] Valiev, M.; Bylaska, E.; Govind, N.; Kowalski, K.; Straatsma, T.; Van Dam, H.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T.; de Jong, W. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477–1489.
- [109] Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Others, MOLPRO, version 2015.1, a package of *ab initio* programs. 2015.
- [110] Girichev, G. V.; Petrov, V. M.; Giricheva, N. I.; Utkin, A. N.; Petrova, V. N. An electron diffraction study of the structure and force field of the hafnium tetrachloride molecule. *Zhurnal Strukt. Khimii* **1980**, *22*, 65–68.
- [111] Balabanov, N. B.; Peterson, K. A. A systematic *ab initio* study of the structure and vibrational spectroscopy of HgCl₂, HgBr₂, and HgBrCl. *J. Chem. Phys.* **2003**, *119*, 12271–12278.
- [112] Gross, P.; Hayman, C. Heats of formation of metal halides. Tetrachlorides of vanadium and hafnium. *Trans. Faraday Soc.* **1964**, *60*, 45–49.
- [113] Rossini, F. D.; Wagman, D. D.; William, E. H.; Levine, S.; Jaffe, I. *Selected values of chemical thermodynamic properties*; 1971.
- [114] Almeida, N. M. S.; Melin, T. R. L.; Wilson, A. K. Multireference calculations on the ground and lowest excited states and dissociation energy of LuF. *J. Chem. Phys.* **2021**, *154*, 244304.
- [115] Banerjee, S.; Pillai, M. R. A.; Knapp, F. F. R. Lutetium-177 Therapeutic Radiopharmaceuticals: Linking Chemistry, Radiochemistry, and Practical Applications. *Chem. Rev.* **2015**, *115*, 2934–2974.
- [116] Hacker, S. G. The Spectrum of Arcturus. *Astrophys. J.* **1936**, *83*, 140.
- [117] Sneden, C.; Cowan, J. J.; Lawler, J. E.; Ivans, I. I.; Burles, S.; Beers, T. C.; Primas, F.; Hill, V.; Truran, J. W.; Fuller, G. M.; Pfeiffer, B.; Kratz, K.-L. The Extremely Metal-poor, Neutron Capture-rich Star CS 2289-052: A Comprehensive Abundance Analysis. *Astrophys. J.* **2003**, *591*, 936–953.

- [118] The s-Process in Metal-Poor Stars: Abundances for 22 Neutron-Capture Elements in CS 31062-050*. *Astrophys. J.* **2004**, 605, 462–471.
- [119] Roederer, I. U.; Sneden, C.; Lawler, J. E.; Cowan, J. J. New abundance determinations of cadmium, lutetium, and osmium in the r-process enriched star BD +17 3248. *Astrophys. J. Lett.* **2010**, 714, 123–127.
- [120] Den Hartog, E. A.; Curry, J. J.; Wickliffe, M. E.; Lawler, J. E. Spectroscopic Data for the 6it s6it p 3P_1 level of Lu⁺ for the Determination of the Solar Lutetium Abundance. *Sol. Phys.* **1998**, 178, 239–244.
- [121] Zmbov, K. F.; Margrave, J. L. Mass Spectrometric Studies of Scandium, Yttrium, Lanthanum, and Rare-Earth Fluorides. *Mass Spectrom. Inorg. Chem.* **1968**, 72, 267.
- [122] Kaledin, L. A.; Heaven, M. C.; Field, R. W. Thermochemical Properties (D_0° and IP) of the Lanthanide Monohalides. *J. Mol. Spectrosc.* **1999**, 193, 285–292.
- [123] D’Incan, J.; Effantin, C.; Bacis, R. Electronic spectrum of the LuF molecule. *J. Phys. B At. Mol. Phys.* **1972**, 5, L189–L190.
- [124] Effantin, C.; Wannous, G.; D’Incan, J. Les états électroniques de la molécule LuF. *Can. J. Phys* **1977**, 55, 64.
- [125] Huber, K. P.; Herzberg, G. Molecular Structure. *Phys. Today* **1952**, 5, 33–33.
- [126] Effantin, C.; Wannous, G.; D’Incan, J.; Athenour, C. Rotational analysis of selected bands from the electronic spectrum of the LuF molecule. *Can. J. Phys.* **1976**, 54, 279–294.
- [127] Rajamanickam, N.; Narasimhamurthy, B. LuF molecule: True potential energy curve and the dissociation energy. *Acta Phys. Hungarica* **1984**, 56, 67–71.
- [128] Reddy, R. R.; Reddy, A. R.; Rao, T. V. R. Dissociation Energies of SiTe and LuF. *Indian J. Pure Appl. Phys.* **1985**, 23, 424–425.
- [129] Simon, A. In *Rare Earth Chemistry*; Pöttgen, R., Jüstel, T., Strassert, C. A., Eds.; De Gruyter: Berlin, Boston, 2020; pp 137–156.
- [130] Hamade, Y.; El Sabbahi, A. Theoretical study of the electronic structure of mono-chloride of lanthanum molecule including spin-orbit coupling effect. *J. Mol. Model.* **2018**, 24, 100.
- [131] Chmaisani, W.; El-Kork, N.; Elmoussaoui, S.; Korek, M. Electronic Structure Calculations with the Spin Orbit Effect of the Low-Lying Electronic States of the YbBr Molecule. *ACS Omega* **2019**, 4, 14987–14995.
- [132] Chmaisani, W.; Korek, M. *Ab initio* study of the low-lying electronic states of YbCl molecule including spin-orbit effects. *J. Quant. Spectrosc. Radiat. Transf.* **2018**, 217, 63–72.
- [133] Yamamoto, S.; Tatewaki, H. Electronic spectra of DyF studied by four-component relativistic configuration interaction methods. *J. Chem. Phys.* **2015**, 142, 094312.

- [134] Plokker, M.; van der Kolk, E. Temperature dependent relaxation dynamics of luminescent NaX:Tm²⁺ (X=Cl, Br, I). *J. Lumin.* **2019**, *216*, 116694.
- [135] Assaf, J.; Taher, F.; Magnier, S. Theoretical description of the low-lying electronic states of LuBr located below 41,700 cm⁻¹. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *189*, 421–427.
- [136] South, C.; Schoendorff, G.; Wilson, A. K. Dissociation energy and electronic structure of the low valent lanthanide compound NdF⁺. *Int. J. Quantum Chem.* **2016**, *116*, 791–794.
- [137] Wang, S. G.; Schwarz, W. H. E. Lanthanide Diatomics and Lanthanide Contractions. *J. Phys. Chem.* **1995**, *99*, 11687–11695.
- [138] Küchle, W.; Dolg, M.; Stoll, H. Ab Initio Study of the Lanthanide and Actinide Contraction. *J. Phys. Chem. A* **1997**, *101*, 7128–7133.
- [139] Cooke, S. A.; Krumrey, C.; Gerry, M. C. Pure rotational spectra of LuF and LuCl. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2570–2578.
- [140] Hong, G.; Dolg, M.; Li, L. A comparison of scalar-relativistic ZORA and DKH density functional schemes: monohydrides, monoxides and monofluorides of La, Lu, Ac and Lr. *Chem. Phys. Lett.* **2001**, *334*, 396–402.
- [141] Grimmel, S.; Schoendorff, G.; Wilson, A. K. Gauging the Performance of Density Functionals for Lanthanide-Containing Molecules. *J. Chem. Theory Comput.* **2016**, *12*, 1259–1266.
- [142] Schoendorff, G.; Wilson, A. K. Low valency in lanthanides: A theoretical study of NdF and LuF. *J. Chem. Phys.* **2014**, *140*, 224314.
- [143] Solomonik, V. G.; Smirnov, A. N. Toward Chemical Accuracy in ab Initio Thermochemistry and Spectroscopy of Lanthanide Compounds: Assessing Core-Valence Correlation, Second-Order Spin-Orbit Coupling, and Higher Order Effects in Lanthanide Diatomics. *J. Chem. Theory Comput.* **2017**, *13*, 5240–5254.
- [144] Lu, Q. Development and applications of relativistic correlation consistent basis sets for lanthanide elements and accurate *ab initio* thermochemistry. PhD thesis, Washington State University, 2017.
- [145] Hamade, Y.; Taher, F.; Choueib, M.; Monteil, Y. Theoretical electronic investigation of the low-lying electronic states of the LuF molecule. *Can. J. Phys.* **2009**, *87*, 1163–1169.
- [146] Assaf, J.; Zeitoun, S.; Safa, A.; Nascimento, E. C. M. Ab-initio study of spin-orbit effect on ¹⁷⁵Lu¹⁹F spectroscopy. *J. Mol. Struct.* **2019**, *1178*, 458–466.
- [147] Werner, H. J. et al. The Molpro quantum chemistry package. *J. Chem. Phys.* **2020**, *152*, 144107.
- [148] Knowles, P. J.; Werner, H.-J. An efficient method for the evaluation of coupling coefficients in configuration interaction calculations. *Chem. Phys. Lett.* **1988**, *145*, 514–522.

- [149] Werner, H.-J.; Knowles, P. J. An efficient internally contracted multiconfiguration-reference configuration interaction method. *J. Chem. Phys.* **1988**, *89*, 5803–5814.
- [150] Werner, H.-J.; Knowles, P. J. A second order multiconfiguration SCF procedure with optimum convergence. *J. Chem. Phys.* **1985**, *82*, 5053–5063.
- [151] Werner, H.-J. Matrix-formulated direct multiconfiguration self-consistent field and multiconfiguration reference configuration-interaction methods. *Adv. Chem. Phys.* **1987**, *69*, 1–62.
- [152] Dunham, J. L. The Energy Levels of a Rotating Vibrator. *Phys. Rev.* **1932**, *41*, 721–731.
- [153] Cao, X.; Dolg, M. Segmented contraction scheme for small-core lanthanide pseudopotential basis sets. *J. Mol. Struct. THEOCHEM* **2002**, *581*, 139–147.
- [154] Cao, X.; Dolg, M. Valence basis sets for relativistic energy-consistent small-core lanthanide pseudopotentials. *J. Chem. Phys.* **2001**, *115*, 7348–7355.
- [155] Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- [156] Gulde, R.; Pollak, P.; Weigend, F. Error-Balanced Segmented Contracted Basis Sets of Double- ζ to Quadruple- ζ Valence Quality for the Lanthanides. *J. Chem. Theory Comput.* **2012**, *8*, 4062–4068.
- [157] Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.
- [158] Sekiya, M.; Noro, T.; Koga, T.; Shimazaki, T. Relativistic segmented contraction basis sets with core-valence correlation effects for atoms $_{57}\text{La}$ through $_{71}\text{Lu}$: Sapporo-DK- $n\text{ZP}$ sets ($n = \text{D, T, Q}$). *Theor. Chem. Acc.* **2012**, *131*, 1–8.
- [159] Gomes, A. S. P.; Dylla, K. G.; Visscher, L. Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the lanthanides La-Lu. *Theor. Chem. Acc.* **2010**, *127*, 369–381.
- [160] Martin, J. M. *Ab initio* total atomization energies of small molecules - towards the basis set limit. *Chem. Phys. Lett.* **1996**, *259*, 669–678.
- [161] Bross, D. H.; Peterson, K. A. Composite thermochemistry of gas phase U(VI)-containing molecules. *J. Chem. Phys.* **2014**, *141*, 244308.
- [162] Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.
- [163] T.Saue, et al. DIRAC, a relativistic *ab initio* electronic structure program. 2018; <http://www.diracprogram.org>.
- [164] Bünzli, J.-C. G. Lanthanide Photonics: Shaping the Nanoworld. *Trends Chem.* **2019**, *1*, 751–762.
- [165] Freedman, D.; Emge, T. J.; Brennan, J. G. $(\text{THF})_8\text{Ln}_8\text{E}_6(\text{EPh})_{12}$ cluster reactivity: Systematic control of Ln, E, EPh, and neutral donor ligands. *Inorg. Chem.* **1999**, *38*, 4400–4404.

- [166] Freedman, D.; Sayan, S.; Emge, T. J.; Croft, M.; Brennan, J. G. Heterovalent clusters: $\text{Ln}_4\text{Se}(\text{SePh})_8$ ($\text{Ln}_4 = \text{Sm}_4, \text{Yb}_4, \text{Sm}_2\text{Yb}_2, \text{Nd}_2\text{Yb}_2$). *J. Am. Chem. Soc.* **1999**, *121*, 11713–11719.
- [167] Melman, J. H.; Fitzgerald, M.; Freedman, D.; Emge, T. J.; Brennan, J. G. Chalcogen-rich lanthanide clusters from lanthanide halide starting materials: A new approach to the low-temperature synthesis of LnS_x solids from molecular precursors. 1999.
- [168] Choi, Y. G.; Kim, K. H.; Park, S. H.; Jong, H. Comparative study of energy transfers from Er^{3+} to Ce^{3+} in tellurite and sulfide glasses under 980 nm excitation. *J. Appl. Phys.* **2000**, *88*, 3832–3839.
- [169] Shin, Y. B.; Heo, J.; Kim, S. Modification of the local phonon modes and electron-phonon coupling strengths in Dy^{3+} -doped sulfide glasses for efficient 1.3 μm amplification. **2000**, *317*, 637–641.
- [170] Luo, Y.; Wan, X.; Ito, Y.; Takami, S.; Kubo, M.; Miyamoto, A. A density functional theory calculation on lanthanide monosulfides. *Chem. Phys.* **2002**, *282*, 197–206.
- [171] Sujecki, S. Modelling and Design of Lanthanide Ion-Doped Chalcogenide Fiber Lasers: Progress towards the Practical Realization of the First MIR Chalcogenide Fiber Laser. *Fibers* **2018**, *6*, 25.
- [172] Crane, R. W.; Sójka, Ł.; Furniss, D.; Nunes, J.; Barney, E.; Farries, M. C.; Benson, T. M.; Sujecki, S.; Seddon, A. B. Experimental photoluminescence and lifetimes at wavelengths including beyond 7 microns in Sm^{3+} -doped selenide-chalcogenide glass fibers. *Opt. Express* **2020**, *28*, 12373.
- [173] Golubkov, A. V.; Kazanin, M. M.; Kaminskii, V. V.; Sokolov, V. V.; Solov'ev, S. M.; Trushnikova, L. N. Thermoelectric Properties of SmS_x ($x = 0.8\text{--}1.5$). *Inorg. Mater.* **2003**, *39*, 1251–1257.
- [174] Chess, D. L.; Chess, C. A.; Biggers, J. V.; White, W. B. Processing Ternary Sulfide Ceramics: Powder Preparation, Sintering, and Hot-Pressing. *J. Am. Ceram. Soc.* **1983**, *66*, 18–22.
- [175] Jundale, S.; Lokhande, C. (Photo)electrochemical studies on p- Sm_2S_3 films. *Sol. Energy Mater. Sol. Cells* **1992**, *28*, 151–157.
- [176] Chandrasekharaiah, M.; Gingerich, K. *Chapter 86 Thermodynamic properties of gaseous species*; Elsevier, 1989; Vol. 12; pp 409–431.
- [177] Murad, E.; Hildenbrand, D. L. Dissociation energies of GdO , HoO , ErO , TmO , and LuO ; correlation of results for the lanthanide monoxide series. *J. Chem. Phys.* **1980**, *73*, 4005–4011.
- [178] Gingerich, K. A. Mass-Spectrometric Evidence for the Molecules UC and CeC and Predicted Stability of Diatomic Carbides of Electropositive Transition Metals. *J. Chem. Phys.* **1969**, *50*, 2255–2256.
- [179] Ciccioli, A.; Gigli, G.; Meloni, G.; Testani, E. The dissociation energy of the new diatomic molecules SiPb and GePb . *J. Chem. Phys.* **2007**, *127*, 054303.

- [180] Dolg, M.; Liu, W.; Kalvoda, S. Performance of relativistic density functional and *ab initio* pseudopotential approaches for systems with high-spin multiplicities: Gadolinium diatomics GdX (X=H, N, O, F, P, S, Cl, Gd). *Int. J. Quantum Chem.* **2000**, *76*, 359–370.
- [181] Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.
- [182] Perdew, J. P.; Wang, Y. Erratum: Accurate and simple analytic representation of the electron-gas correlation energy [Phys. Rev. B 45, 13244 (1992)]. *Phys. Rev. B* **2018**, *98*, 079904.
- [183] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- [184] Xu, W.; Ji, W.-X.; Wang, S.-G. The electronic configurations of LnX (Ln=La-Eu, X=O, S, Se, Te): A FON-DFT investigation. *Comput. Theor. Chem.* **2015**, *1068*, 81–87.
- [185] Aebbersold, L. E.; Wilson, A. K. Considering Density Functional Approaches for Actinide Species: The An66 Molecule Set. *J. Phys. Chem. A* **2021**, *125*, 7029–7037.
- [186] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792–794.
- [187] Adeyiga, O.; Suleiman, O.; Dandu, N. K.; Odoh, S. O. Ground-state actinide chemistry with scalar-relativistic multiconfiguration pair-density functional theory. *J. Chem. Phys.* **2019**, *151*, 134102.
- [188] Coppens, P.; Smoes, S.; Drowart, J. Mass spectrometric determination of the dissociation energies of the molecules GeS, ScS, YS, LaS and CeS. *Trans. Faraday Soc.* **1967**, *63*, 2140.
- [189] Smoes, S.; Coppens, P.; Bergman, C.; Drowart, J. Mass spectrometric determination of the dissociation energies of the gaseous rare earth monosulphides. *Trans. Faraday Soc.* **1969**, *65*, 682.
- [190] Bergman, C.; Coppens, P.; Drowart, J.; Smoes, S. Mass spectrometric determination of the dissociation energies of the gaseous rare earth monoselenides and monotellurides. *Trans. Faraday Soc.* **1970**, *66*, 800.
- [191] Nagai, S.-I.; Shinmei, M.; Yokokawa, T. Heats of atomization, dissociation energies and heats of sublimation of several rare earth monochalcogenides. *J. inorg. nucl.* **1974**, *36*, 1904–1905.
- [192] Oakley, M. S.; Gagliardi, L.; Truhlar, D. G. Multiconfiguration Pair-Density Functional Theory for Transition Metal Silicide Bond Dissociation Energies, Bond Lengths, and State Orderings. *Molecules* **2021**, *26*, 2881.
- [193] Sevy, A.; Matthew, D. J.; Morse, M. D. Bond dissociation energies of TiC, ZrC, HfC, ThC, NbC, and TaC. *J. Chem. Phys.* **2018**, *149*, 44306.
- [194] Sevy, A.; Merriles, D. M.; Wentz, R. S.; Morse, M. D. Bond dissociation energies of ScSi, YSi, LaSi, ScC, YC, LaC, CoC, and YCH. *J. Chem. Phys.* **2019**, *151*, 024302.

- [195] Sorensen, J. J.; Tieu, E.; Morse, M. D. Bond dissociation energies of lanthanide sulfides and selenides. *J. Chem. Phys.* **2021**, *154*, 124307.
- [196] Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. Gaussian-1 theory of molecular energies for second-row compounds. *J. Chem. Phys.* **1990**, *93*, 2537–2545.
- [197] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. Gaussian-3X (G3X) theory: Use of improved geometries, zero-point energies, and Hartree-Fock basis sets. *J. Chem. Phys.* **2001**, *114*, 108–117.
- [198] Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory. *J. Chem. Phys.* **2007**, *126*, 084108.
- [199] Parthiban, S.; Martin, J. M. Assessment of W1 and W2 theories for the computational of electron affinities, ionization potentials, heats of formation, and proton affinities. *J. Chem. Phys.* **2001**, *114*, 6014–6029.
- [200] Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *J. Chem. Phys.* **1988**, *89*, 2193–2218.
- [201] Petersson, G. A.; Tensfeldt, T. G.; Montgomery, J. A. A complete basis set model chemistry. III. The complete basis set-quadratic configuration interaction family of methods. *J. Chem. Phys.* **1991**, *94*, 6091–6101.
- [202] Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *J. Chem. Phys.* **1996**, *104*, 2598–2619.
- [203] Montgomery, J. A.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. IV. An improved atomic pair natural orbital method. *J. Chem. Phys.* **1994**, *101*, 5900–5909.
- [204] Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* **1999**, *110*, 2822–2827.
- [205] Thorpe, J. H.; Lopez, C. A.; Nguyen, T. L.; Baraban, J. H.; Bross, D. H.; Ruscic, B.; Stanton, J. F. High-accuracy extrapolated *ab initio* thermochemistry. IV. A modified recipe for computational efficiency. *J. Chem. Phys.* **2019**, *150*, 224102.
- [206] Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical accuracy in *ab initio* thermochemistry and spectroscopy: current strategies and future challenges. *Theor. Chem. Acc.* **2012**, *131*, 1079.
- [207] Dixon, D. A.; Feller, D.; Peterson, K. A. In *Annu Rep Comput Chem*; Wheeler, R. A., Ed.; Annu Rep Comput Chem; Elsevier, 2012; Vol. 8; pp 1–28.
- [208] DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *J. Chem. Phys.* **2006**, *124*, 114104.

- [209] DeYonker, N. J.; Mintz, B.; Cundari, T. R.; Wilson, A. K. Application of the correlation consistent composite approach (ccCA) to third-row (Ga-Kr) molecules. *J. Chem. Theory Comput.* **2008**, *4*, 328–334.
- [210] DeYonker, N. J.; Ho, D. S.; Wilson, A. K.; Cundari, T. R. Computational s-Block Thermochemistry with the Correlation Consistent Composite Approach. *J. Phys. Chem. A* **2007**, *111*, 10776–10780.
- [211] Welch, B. K.; Almeida, N. M. S.; Wilson, A. K. Super ccCA (s-ccCA): an approach for accurate transition metal thermochemistry. *Mol. Phys.* **2021**, *119*, e1963001.
- [212] Sevy, A.; Tieu, E.; Morse, M. D. Bond dissociation energies of FeSi, RuSi, OsSi, CoSi, RhSi, IrSi, NiSi, and PtSi. *J. Chem. Phys.* **2018**, *149*, 174307.
- [213] Merriles, D. M.; Sevy, A.; Nielson, C.; Morse, M. D. The bond dissociation energy of VO measured by resonant three-photon ionization spectroscopy. *J. Chem. Phys.* **2020**, *153*, 24303.
- [214] Merriles, D. M.; Tomchak, K. H.; Ewigleben, J. C.; Morse, M. D. Predissociation measurements of the bond dissociation energies of EuO, TmO, and YbO. *J. Chem. Phys.* **2021**, *155*, 144303.
- [215] Almeida, N. M. S.; McKinlay, R. G.; Paterson, M. J. Excited electronic states of MnO₄: Challenges for wavefunction and density functional response theories. *Chem. Phys.* **2015**, *446*, 86–91.
- [216] McKinlay, R. G.; Almeida, N. M. S.; Coe, J. P.; Paterson, M. J. Excited States of the Nickel Carbonyls Ni(CO) and Ni(CO)₄: Challenging Molecules for Electronic Structure Theory. *J. Phys. Chem. A* **2015**, *119*, 10076–10083.
- [217] Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457–464.
- [218] Ruedenberg, K. A Study of Two-Center Integrals Useful in Calculations on Molecular Structure. II. The Two-Center Exchange Integrals. *J. Chem. Phys.* **1951**, *19*, 1459–1477.
- [219] Ruedenberg, K. The Physical Nature of the Chemical Bond. *Rev. Mod. Phys.* **1962**, *34*, 326–376.
- [220] Ruedenberg, K. Autobiography of Klaus Ruedenberg. *J. Phys. Chem. A* **2010**, *114*, 8490–8495.
- [221] Fang, Z.; Vasiliu, M.; Peterson, K. A.; Dixon, D. A. Prediction of Bond Dissociation Energies/Heats of Formation for Diatomic Transition Metal Compounds: CCSD(T) Works. *J. Chem. Theory Comput.* **2017**, *13*, 1057–1066.
- [222] Jansen, G.; Hess, B. A. Revision of the Douglas-Kroll transformation. *Phys. Rev. A* **1989**, *39*, 6016–6017.
- [223] Lu, Q.; Peterson, K. A. Correlation consistent basis sets for lanthanides: The atoms La-Lu. *J. Chem. Phys.* **2016**, *145*, 054111.

- [224] Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton. *J. Chem. Phys.* **1999**, *110*, 7667–7676.
- [225] de Jong, W. A.; Harrison, R. J.; Dixon, D. A. Parallel Douglas-Kroll energy and gradients in NWChem: Estimating scalar relativistic effects using Douglas-Kroll contracted basis sets. *J. Chem. Phys.* **2000**, *114*, 48–53.
- [226] Balabanov, N. B.; Peterson, K. A. Systematically convergent basis sets for transition metals. I. All-electron correlation consistent basis sets for the 3d elements Sc-Zn. *J. Chem. Phys.* **2005**, *123*, 064107.
- [227] Dolg, M.; Stoll, H.; Preuss, H. Energy-adjusted *ab initio* pseudopotentials for the rare earth elements. *J. Chem. Phys.* **1989**, *90*, 1730–1734.
- [228] Cao, X.; Dolg, M.; Stoll, H. Valence basis sets for relativistic energy-consistent small-core actinide pseudopotentials. *J. Chem. Phys.* **2002**, *118*, 487–496.
- [229] Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements. *J. Chem. Phys.* **2003**, *119*, 11113–11123.
- [230] Lee, T. J.; Taylor, P. R. A diagnostic for determining the quality of single-reference electron correlation methods. *Int. J. Quantum Chem.* **2009**, *36*, 199–207.
- [231] Lee, T. J.; Rice, J. E.; Scuseria, G. E.; Schaefer, H. F. Theoretical investigations of molecules composed only of fluorine, oxygen and nitrogen: determination of the equilibrium structures of FOOF, (NO)₂ and FNNF and the transition state structure for FNNF *cis-trans* isomerization. *Theor. Chim. Acta* **1989**, *75*, 81–98.
- [232] Janssen, C. L.; Nielsen, I. M. B. New diagnostics for coupled-cluster and Møller-Plesset perturbation theory. *Chem. Phys. Lett.* **1998**, *290*, 423–430.
- [233] G. Czack, H. Hein, I. Hinz, H. Bergmann, P. Kuhn,, *Gmelin Handbook of Inorg. Chem.*, 8th ed.; Springer-Verlag: New York, 1983; Chapter 7.
- [234] A. Kramida,; Yu. Ralchenko,; J. Reader,; NIST ASD Team, NIST Atomic Spectra Database (ver. 5.8), [Online]. Available: <https://physics.nist.gov/asd> [2021, October 7]. National Institute of Standards and Technology, Gaithersburg, MD. 2020; <https://doi.org/10.18434/T4W30F>.
- [235] Jiang, W.; Deyonker, N. J.; Determan, J. J.; Wilson, A. K. Toward accurate theoretical thermochemistry of first row transition metal complexes. *J. Phys. Chem. A* **2012**, *116*, 870–885.
- [236] Graber, J. M.; Alexander, C.; Laumbach, R. J.; Black, K.; Strickland, P. O.; Georgopoulos, P. G.; Marshall, E. G.; Shendell, D. G.; Alderson, D.; Mi, Z.; Mascari, M.; Weisel, C. P. Per and polyfluoroalkyl substances (PFAS) blood levels after contamination of a community water supply and comparison with 2013–2014 NHANES. *J. Expo. Sci. Environ. Epidemiol.* **2019**, *29*, 172–182.

- [237] Lai, T. T.; Eken, Y.; Wilson, A. K. Binding of Per- And Polyfluoroalkyl Substances to the Human Pregnane X Receptor. *Environ. Sci. Technol.* **2020**, *54*, 15986–15995.
- [238] Zafeiraki, E.; Gebbink, W. A.; Hoogenboom, R. L. A. P.; Kotterman, M.; Kwadijk, C.; Dassenakis, E.; van Leeuwen, S. P. J. Occurrence of perfluoroalkyl substances (PFASs) in a large number of wild and farmed aquatic animals collected in the Netherlands. *Chemosphere* **2019**, *232*, 415–423.
- [239] Brusseau, M. L. The influence of molecular structure on the adsorption of PFAS to fluid-fluid interfaces: Using QSPR to predict interfacial adsorption coefficients. *Water Res.* **2019**, *152*, 148–158.
- [240] Sunderland, E. M.; Hu, X. C.; Dassuncao, C.; Tokranov, A. K.; Wagner, C. C.; Allen, J. G. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. *J. Expo. Sci. Environ. Epidemiol.* **2019**, *29*, 131–147.
- [241] Loganathan, N.; Wilson, A. K. Adsorption, Structure, and Dynamics of Short- and Long-Chain PFAS Molecules in Kaolinite: Molecular-Level Insights. *Environ. Sci. Technol.* **2022**, *56*, 8043–8052.
- [242] Almeida, N. M. S.; Eken, Y.; Wilson, A. K. Binding of Per- and Polyfluoro-alkyl Substances to Peroxisome Proliferator-Activated Receptor Gamma. *ACS Omega* **2021**, *6*, 15103–15114.
- [243] Ahrens, L.; Yeung, L. W.; Taniyasu, S.; Lam, P. K.; Yamashita, N. Partitioning of perfluorooctanoate (PFOA), perfluorooctane sulfonate (PFOS) and perfluorooctane sulfonamide (PFOSA) between water and sediment. *Chemosphere* **2011**, *85*, 731–737.
- [244] Vento, S. D.; Halsall, C.; Gioia, R.; Jones, K.; Dachs, J. Volatile per- and polyfluoroalkyl compounds in the remote atmosphere of the western Antarctic Peninsula: an indirect source of perfluoroalkyl acids to Antarctic waters? *Atmos. Pollut. Res.* **2012**, *3*, 450–455.
- [245] Roth, J.; Abusallout, I.; Hill, T.; Holton, C.; Thapa, U.; Hanigan, D. Release of Volatile Per- and Polyfluoroalkyl Substances from Aqueous Film-Forming Foam. *Environ. Sci. Technol. Lett.* **2020**, *7*, 164–170.
- [246] Riedel, T. P.; Lang, J. R.; Strynar, M. J.; Lindstrom, A. B.; Offenberg, J. H. Gas-Phase Detection of Fluorotelomer Alcohols and Other Oxygenated Per- and Polyfluoroalkyl Substances by Chemical Ionization Mass Spectrometry. *Environ. Sci. Technol. Lett.* **2019**, *6*, 289–293.
- [247] Giroday, T.; Montero-Campillo, M. M.; Mora-Diez, N. Thermodynamic stability of PFOS: M06-2X and B3LYP comparison. *Comput. Theor. Chem.* **2014**, *1046*, 81–92.
- [248] Montero-Campillo, M. M.; Mora-Diez, N.; Lamsabhi, A. M. Thermodynamic stability of neutral and anionic PFOS: A gas-phase, *n*-octanol, and water theoretical study. *J. Phys. Chem. A* **2010**, *114*, 10148–10155.
- [249] Hidalgo, A.; Giroday, T.; Mora-Diez, N. Thermodynamic stability of neutral and anionic PFOAs. *Theor. Chem. Acc.* **2015**, *134*, 1–15.

- [250] Rayne, S.; Forest, K.; Friesen, K. J. Relative gas-phase free energies for the C₃ through C₈ linear and branched perfluorinated sulfonic acids: Implications for kinetic versus thermodynamic control during synthesis of technical mixtures and predicting congener profile inputs to environmental . *J. Mol. Struct. THEOCHEM* **2008**, *869*, 81–82.
- [251] Rayne, S.; Forest, K. Theoretical studies on the pK_a values of perfluoroalkyl carboxylic acids: Non-helical conformation acidity constants for n-perfluorooctanoic acid (n-PFOA). *Nat. Preced.* **2009**, 1–22.
- [252] Kolesov, V. P.; Kozina, M. P. Thermochemistry of Organic and Organohalogen Compounds. *Russ. Chem. Rev.* **1986**, *55*, 912–928.
- [253] Minenkov, Y.; Wang, H.; Wang, Z.; Sarathy, S. M.; Cavallo, L. Heats of Formation of Medium-Sized Organic Compounds from Contemporary Electronic Structure Methods. *J. Chem. Theory Comput.* **2017**, *13*, 3537–3560.
- [254] Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.
- [255] Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315–2372.
- [256] Rayne, S.; Forest, K. Theoretical studies on the pK_a values of perfluoroalkyl carboxylic acids. *J. Mol. Struct. THEOCHEM* **2010**, *949*, 60–69.
- [257] Ayata, S.; Stefanova, A.; Ernst, S.; Baltruschat, H. The electro-oxidation of water and alcohols at BDD in hexafluoroisopropanol. *J. Electroanal. Chem.* **2013**, *701*, 1–6.
- [258] Paulechka, E.; Kazakov, A. Efficient DLPNO-CCSD(T)-Based Estimation of Formation Enthalpies for C-, H-, O-, and N-Containing Closed-Shell Compounds Validated Against Critically Evaluated Experimental Data. *J. Phys. Chem. A* **2017**, *121*, 4379–4387.
- [259] Mielczarek, D. C.; Nait Saidi, C.; Paricaud, P.; Catoire, L. Generalized Prediction of Enthalpies of Formation Using DLPNO-CCSD(T) *Ab Initio* Calculations for Molecules Containing the Elements H, C, N, O, F, S, Cl, Br. *J. Comput. Chem.* **2019**, *40*, 768–793.
- [260] Bentel, M. J.; Yu, Y.; Xu, L.; Li, Z.; Wong, B. M.; Men, Y.; Liu, J. Defluorination of Per- and Polyfluoroalkyl Substances (PFASs) with Hydrated Electrons: Structural Dependence and Implications to PFAS Remediation and Management. *Environ. Sci. Technol.* **2019**, *53*, 3718–3728.
- [261] Altarawneh, M. A chemical kinetic model for the decomposition of perfluorinated sulfonic acids. *Chemosphere* **2021**, *263*, 128256.
- [262] Ruscic, B.; Pinzon, R. E.; Morton, M. L.; Von Laszewski, G.; Bittner, S. J.; Nijssure, S. G.; Amin, K. A.; Minkoff, M.; Wagner, A. F. Introduction to active thermochemical tables: Several "Key" enthalpies of formation revisited. *J. Phys. Chem. A* **2004**, *108*, 9979–9997.

- [263] Patel, P.; Wilson, A. K. Domain-based local pair natural orbital methods within the correlation consistent composite approach. *J. Comput. Chem.* **2020**, *41*, 800–813.
- [264] Liu, Z.; Goddard, J. D. Predictions of the fluorine NMR chemical shifts of perfluorinated carboxylic acids, $C_nF_{2n+1}COOH$ ($n = 6-8$). *J. Phys. Chem. A* **2009**, *113*, 13921–13931.
- [265] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Investigation of the use of B3LYP zero-point energies and geometries in the calculation of enthalpies of formation. *Chem. Phys. Lett.* **1997**, *270*, 419–426.
- [266] Bukheet Hassan, H. Density Function Theory B3LYP/6-31G** Calculation of Geometry Optimization and Energies of Donor-Bridge-Acceptor Molecular System. *Int. J. Curr. Eng. Technol.* **2014**, *4*, 2342–2345.
- [267] Aprà, E. et al. NWChem: Past, present, and future. *J. Chem. Phys.* **2020**, *152*, 184102.
- [268] Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, 4–9.
- [269] Joback, K.; Reid, R. Estimation of pure-component properties from group-contributions. *Chem. Eng. Commun.* **1987**, *57*, 233–243.
- [270] Zhao, Y.; Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **2008**, *41*, 157–167.
- [271] Kolesov, V. P.; Papina, T. S. Thermochemistry of Haloethanes. *Russ. Chem. Rev.* **1983**, *52*, 425–439.
- [272] Paulechka, E.; Kazakov, A. Critical Evaluation of the Enthalpies of Formation for Fluorinated Compounds Using Experimental Data and High-Level *Ab Initio* Calculations. *J. Chem. Eng. Data* **2019**, *64*, 4863–4874.
- [273] Burcat, A.; Branko, R. Third millennium ideal gas and condensed phase thermochemical database for combustion with updates from active thermochemical tables. *Tech. Rep.* **2005**, ANL-05/20, ANL-05/20 TAE 960.
- [274] Ventura, O. N.; Segovia, M. Density functional computational thermochemistry: Accurate determination of the enthalpy of formation of perfluoropropane from DFT and *ab initio* calculations on isodesmic reactions. *Chem. Phys. Lett.* **2005**, *403*, 378–384.
- [275] Stewart, J. J. Comparison of the accuracy of semiempirical and some DFT methods for predicting heats of formation. *J. Mol. Model.* **2004**, *10*, 6–12.
- [276] Ellis, D. A.; Martin, J. W.; De Silva, A. O.; Mabury, S. A.; Hurley, M. D.; Sulbaek Andersen, M. P.; Wallington, T. J. Degradation of fluorotelomer alcohols: A likely atmospheric source of perfluorinated carboxylic acids. *Environ. Sci. Technol.* **2004**, *38*, 3316–3321.
- [277] Brodsky, S. J.; Drell, S. D. The Present Status of Quantum Electrodynamics. *Annu. Rev. Nucl. Sci.* **1970**, *20*, 147–194.

- [278] McGreevy, J. Quantum electrodynamics in a piece of rock. *Nat. Rev. Phys.* **2019**, *1*, 583–584.
- [279] Ulusoy, I. S.; Nest, M. The multi-configuration electron-nuclear dynamics method applied to LiH. *J. Chem. Phys.* **2012**, *136*, 054112.
- [280] Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.
- [281] Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. Benchmarks for electronically excited states: Time-dependent density functional theory and density functional theory based multireference configuration interaction. *J. Chem. Phys.* **2008**, *129*, 104103.
- [282] Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2420–2435.
- [283] Parac, M.; Grimme, S. Comparison of Multireference MøllerPlesset Theory and Time-Dependent Methods for the Calculation of Vertical Excitation Energies of Molecules. *J. Phys. Chem. A* **2002**, *106*, 6844–6850.
- [284] Maschietto, F.; Campetella, M.; Sanz García, J.; Adamo, C.; Ciofini, I. Chasing unphysical TD-DFT excited states in transition metal complexes with a simple diagnostic tool. *J. Chem. Phys.* **2021**, *154*, 204102.
- [285] Kashar, T. I.; Aal, S. A. Spectral, DFT-TDDFT computational investigation and biological studies of transition metal complexes of dehydroacetic acid schiff base. *J. Iran. Chem. Soc.* **2021**, *18*, 1625–1640.
- [286] Dreuw, A.; Head-Gordon, M. Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The ZincbacteriochlorinBacteriochlorin and BacteriochlorophyllSpheroidene Complexes. *J. Am. Chem. Soc.* **2004**, *126*, 4007–4016.
- [287] Manzhos, S.; Segawa, H.; Yamashita, K. Computational dye design by changing the conjugation order: Failure of LR-TDDFT to predict relative excitation energies in organic dyes differing by the position of the methine unit. *Chem. Phys. Lett.* **2012**, *527*, 51–56.
- [288] Koch, H.; Jørgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **1990**, *93*, 3333–3344.
- [289] Kristiansen, H. E.; Schøyen, Ø. S.; Kvaal, S.; Pedersen, T. B. Numerical stability of time-dependent coupled-cluster methods for many-electron dynamics in intense laser pulses. *J. Chem. Phys.* **2020**, *152*, 71102.
- [290] Nascimento, D. R.; DePrince, A. E. A general time-domain formulation of equation-of-motion coupled-cluster theory for linear spectroscopy. *J. Chem. Phys.* **2019**, *151*, 204107.
- [291] Greenman, L.; Ho, P. J.; Pabst, S.; Kamarchik, E.; Mazziotti, D. A.; Santra, R. Implementation of the time-dependent configuration-interaction singles method for atomic strong-field processes. *Phys. Rev. A - At. Mol. Opt. Phys.* **2010**, *82*, 1–12.

- [292] Krause, P.; Sonk, J. A.; Schlegel, H. B. Strong field ionization rates simulated with time-dependent configuration interaction and an absorbing potential. *J. Chem. Phys.* **2014**, *140*, 174113.
- [293] Zuo, T.; Bandrauk, A. D. Charge-resonance-enhanced ionization of diatomic molecular ions by intense lasers. *Phys. Rev. A* **1995**, *52*, 2511–2514.
- [294] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery Jr, J. A. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- [295] Pavičić, D.; Lee, K. F.; Rayner, D. M.; Corkum, P. B.; Villeneuve, D. M. Direct measurement of the angular dependence of ionization for N₂, O₂, and CO₂ in intense laser fields. *Phys. Rev. Lett.* **2007**, *98*, 1–4.
- [296] Hoerner, P.; Lee, M. K.; Schlegel, H. B. Angular dependence of strong field ionization of N₂ by time-dependent configuration interaction using density functional theory and the Tamm-Dancoff approximation. *J. Chem. Phys.* **2019**, *151*, 54102.
- [297] Sissay, A.; Abanador, P.; Mauger, F.; Gaarde, M.; Schafer, K. J.; Lopata, K. Angle-dependent strong-field molecular ionization rates with tuned range-separated time-dependent density functional theory. *J. Chem. Phys.* **2016**, *145*, 094105.
- [298] Kjeldsen, T. K.; Madsen, L. B. Strong-field ionization of N₂: Length and velocity gauge strong-field approximation and tunnelling theory. *J. Phys. B. At. Mol. Opt.* **2004**, *37*, 2033–2044.
- [299] Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- [300] Aquilante, F. et al. Modern quantum chemistry with [Open]Molcas. *J. Chem. Phys.* **2020**, *152*, 214117.
- [301] Coccia, E.; Assaraf, R.; Luppi, E.; Toulouse, J. *Ab initio* lifetime correction to scattering states for time-dependent electronic-structure calculations with incomplete basis sets. *J. Chem. Phys.* **2017**, *147*, 14106.
- [302] Li Manni, G.; Carlson, R. K.; Luo, S.; Ma, D.; Olsen, J.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration Pair-Density Functional Theory. *J. Chem. Theory Comput.* **2014**, *10*, 3669–3680.
- [303] Ruscic, B. Uncertainty quantification in thermochemistry, benchmarking electronic structure computations, and Active Thermochemical Tables. *Int. J. Quantum Chem.* **2014**, *114*, 1097–1101.
- [304] Feller, D.; Bross, D. H.; Ruscic, B. Enthalpy of Formation of C₂H₂O₄ (Oxalic Acid) from High-Level Calculations and the Active Thermochemical Tables Approach. *J. Phys. Chem. A* **2019**, *123*, 3481–3496.

APPENDIX A

5d DFT ENTHALPIES OF FORMATION

Experimental determination of equilibrium constants

As an alternative to calorimetric measurement, equilibrium constants can be used to extract the enthalpy of formation via vapor pressure measurements. One of the methods that relates the change in enthalpy to the vapor pressure is the second law method which is based on the vant Hoff equation:

$$\Delta H^\circ = RT^2 \frac{d \ln(K_p)}{dT} \quad (\text{A.1})$$

for which the vapor pressure at different temperatures is needed. The enthalpy is obtained by calculating the slope of $\ln(K_p)$ vs $1/T$. This method requires a large number of vapor pressure measurements over a large temperature range.

Another approach is the third law method which is based on the Gibbs energy function

$$\Phi_T^\circ = (G_T^\circ - H_T^\circ) / T \quad (\text{A.2})$$

where the standard state Gibbs energy (G_T°) and enthalpy (H_T°) are determined for one temperature. The enthalpy of reaction is then obtained by calculating $\Delta H_r^\circ / T$.

$$\frac{\Delta H_r^\circ}{T} = -R \ln(K_p) - \Delta \Phi_T^\circ \quad (\text{A.3})$$

The main advantage of the third law approach over the second law approach is that the vapor pressure is needed for only one temperature which reduces the final error due to experimental measurement. Moreover, if the vapor pressure data does not lead to a straight line with the second law method, a large error bar can result. In Gurvichs handbook,⁷¹ the $\Delta H_f^{298 \text{ K}}$ reported were calculated using experimental vapor pressure and the use of the third law method while NIST-JANAF⁶⁸ reported data calculated using both the second law and third law methods.

DFT enthalpies of formation

Table A.1: Enthalpies of formation in kcal·mol⁻¹ for 5d compounds with DFT.

Molecules	SVWN	BP86	BLYP	PW91	PBE	TPSS	M06-L	B3P86	X3LYP	B97-1	B3LYP	PBE0	BHLYP	TPSSH	Exp
HfF ₄	-514.35	-433.79	-426.14	-442.37	-438.74	-427.85	-417.98	-422.65	-412.47	-416.34	-411.94	-407.87	-375.36	-417.12	-399.1
HfCl	41.87	57.97	62.00	56.25	56.69	56.96	56.40	58.11	61.58	57.93	62.04	60.86	67.86	57.79	7
HfCl ₂	-67.80	-34.89	-26.71	-38.19	-37.13	-34.58	-39.34	-35.14	-28.25	-34.91	-27.31	-31.04	-19.01	-33.15	-76
HfCl ₄	-282.75	-222.64	-203.34	-230.58	-228.66	-226.72	-239.64	-224.69	-207.22	-222.57	-205.62	-219.92	-193.39	-224.00	-212.90 ± 0.3
HfBr ₂	-42.35	-10.77	-2.02	-13.96	-12.64	-12.22	-3.64	-11.18	13.67	9.31	14.55	9.00	4.33	-16.50	-6.9
HfI ₂	—	19.89	—	16.45	17.70	18.51	23.42	19.40	27.57	20.92	28.52	22.47	35.36	19.49	-8.9
TaCl ₅	-287.65	-203.91	-175.62	-214.01	-211.61	-205.92	-219.43	-197.63	-172.74	-192.42	-170.92	-192.81	-148.37	-198.50	-182.80 ± 2
WCl	109.22	124.34	134.72	120.96	121.23	122.92	133.54	123.28	133.95	135.72	133.79	122.16	137.00	123.37	132.30 ± 10
WCl ₂	—	49.61	62.81	45.89	46.66	48.88	50.93	49.75	62.25	59.83	62.69	50.83	68.51	50.57	-3.00 ± 25
WCl ₄	-155.76	-89.02	-69.17	-95.84	-94.58	-89.19	-94.21	-83.97	-66.14	-83.13	-64.74	-79.53	-47.26	-83.58	-80.3 ± 7.9
WCl ₅	-216.16	-128.07	-101.85	-137.14	-135.19	-127.24	-133.54	-117.58	-94.37	-115.34	-92.68	-111.21	-64.25	-118.13	-98.60 ± 7.88
HgBr	10.46	20.97	25.24	19.37	19.79	19.55	22.98	20.46	24.01	22.56	24.53	20.14	24.74	19.79	24.49 ± 9
HgBr ₂	-51.93	-24.19	-13.32	-27.53	-26.69	-26.94	-19.32	-26.18	-17.29	-22.97	-16.10	-26.49	-15.81	-26.70	-20.50 ± 3
HgCl	6.52	17.79	22.27	16.13	16.39	16.71	15.99	17.42	21.25	19.07	21.73	17.10	22.31	17.07	18.75 ± 2.3
HgCl ₂	-63.95	-34.47	-23.21	-37.85	-37.42	-36.11	-40.34	-36.16	-26.75	-34.28	-25.65	-36.33	-24.32	-35.63	-34.96 ± 1.5
HgI	—	25.29	—	23.69	24.12	23.90	25.49	24.72	28.08	26.49	28.62	24.26	28.64	24.03	31.90 ± 1
HgI ₂	—	-11.61	—	-15.06	-14.15	-14.61	-11.52	-13.82	-5.20	-10.68	-3.97	-14.53	-4.34	-14.63	-3.86 ± 0.5

Table A.2: Enthalpies of formation in kcal·mol⁻¹ for 5*d* compounds with SO-DFT.

Molecules	SVWN	BP86	BLYP	PW91	PBE	TPSS	M06-L	B3P86	X3LYP	B97-1	B3LYP	PBE0	BHLYP	TPSSH	Exp
HfF ₄	-514.35	-433.79	-426.14	-442.37	-438.74	-427.85	-417.98	-422.65	-412.47	-416.34	-411.94	-407.87	-375.36	-417.12	-399.1
HfCl	41.87	57.97	62.00	56.25	56.69	56.96	56.40	58.11	61.58	57.93	62.04	60.86	67.86	57.79	7
HfCl ₂	-67.80	-34.89	-26.71	-38.19	-37.13	-34.58	-39.34	-35.14	-28.25	-34.91	-27.31	-31.04	-19.01	-33.15	-76
HfCl ₄	-282.75	-222.64	-203.34	-230.58	-228.66	-226.72	-239.64	-224.69	-207.22	-222.57	-205.62	-219.92	-193.39	-224.00	-212.90 ± 0.3
HfBr ₂	-42.35	-10.77	-2.02	-13.96	-12.64	-12.22	-3.64	-11.18	13.67	9.31	14.55	9.00	4.33	-16.50	-6.9
HfI ₂	—	19.89	—	16.45	17.70	18.51	23.42	19.40	27.57	20.92	28.52	22.47	35.36	19.49	-8.9
TaCl ₅	-287.65	-203.91	-175.62	-214.01	-211.61	-205.92	-219.43	-197.63	-172.74	-192.42	-170.92	-192.81	-148.37	-198.50	-182.80 ± 2
WCl	109.22	124.34	134.72	120.96	121.23	122.92	133.54	123.28	133.95	135.72	133.79	122.16	137.00	123.37	132.30 ± 10
WCl ₂	—	49.61	62.81	45.89	46.66	48.88	50.93	49.75	62.25	59.83	62.69	50.83	68.51	50.57	-3.00 ± 25
WCl ₄	-155.76	-89.02	-69.17	-95.84	-94.58	-89.19	-94.21	-83.97	-66.14	-83.13	-64.74	-79.53	-47.26	-83.58	-80.3 ± 7.9
WCl ₅	-216.16	-128.07	-101.85	-137.14	-135.19	-127.24	-133.54	-117.58	-94.37	-115.34	-92.68	-111.21	-64.25	-118.13	-98.60 ± 7.88
HgBr	10.46	20.97	25.24	19.37	19.79	19.55	22.98	20.46	24.01	22.56	24.53	20.14	24.74	19.79	24.49 ± 9
HgBr ₂	-51.93	-24.19	-13.32	-27.53	-26.69	-26.94	-19.32	-26.18	-17.29	-22.97	-16.10	-26.49	-15.81	-26.70	-20.50 ± 3
HgCl	6.52	17.79	22.27	16.13	16.39	16.71	15.99	17.42	21.25	19.07	21.73	17.10	22.31	17.07	18.75 ± 2.3
HgCl ₂	-63.95	-34.47	-23.21	-37.85	-37.42	-36.11	-40.34	-36.16	-26.75	-34.28	-25.65	-36.33	-24.32	-35.63	-34.96 ± 1.5
HgI	—	25.29	—	23.69	24.12	23.90	25.49	24.72	28.08	26.49	28.62	24.26	28.64	24.03	31.90 ± 1
HgI ₂	—	-11.61	—	-15.06	-14.15	-14.61	-11.52	-13.82	-5.20	-10.68	-3.97	-14.53	-4.34	-14.63	-3.86 ± 0.5

Isogyric and atomization approaches for the hafnium complexes

Table A.3: Enthalpies of formation in $\text{kcal}\cdot\text{mol}^{-1}$ of the molecule involved in reaction 3.2 to 3.4. The enthalpies calculated with the reactions are using only experimental $\Delta H_f^{298\text{ K}}$ unless noted and all energies were calculated using rp-ccCA.

Molecules	$\Delta H_f^{298\text{ K}}$ (TAE)	$\Delta H_f^{298\text{ K}}$ (Reac. 3.2)	$\Delta H_f^{298\text{ K}}$ (Reac. 3.3)	$\Delta H_f^{298\text{ K}}$ (Reac. 3.4)	Exp.	Theory ⁸⁵
HfCl ₄	-210.18	-260.14 -211.14 ^a	-217.33	-255.00 -206.00 ^a	-212.90	
HfCl ₂	-27.00	-28.75	—	—	-76.00	-27.1 ± 3.3
HfBr ₂	-0.98	—	-2.46	—	-6.90	4.2 ± 5.8
HfI ₂	38.22	—	—	33.60	-8.90	40.5 ± 4.9

^a Results obtained using theoretical enthalpies of formation for the known complex.

APPENDIX B

LuF MRCI CURVES AND 95% CONFIDENCE LIMIT

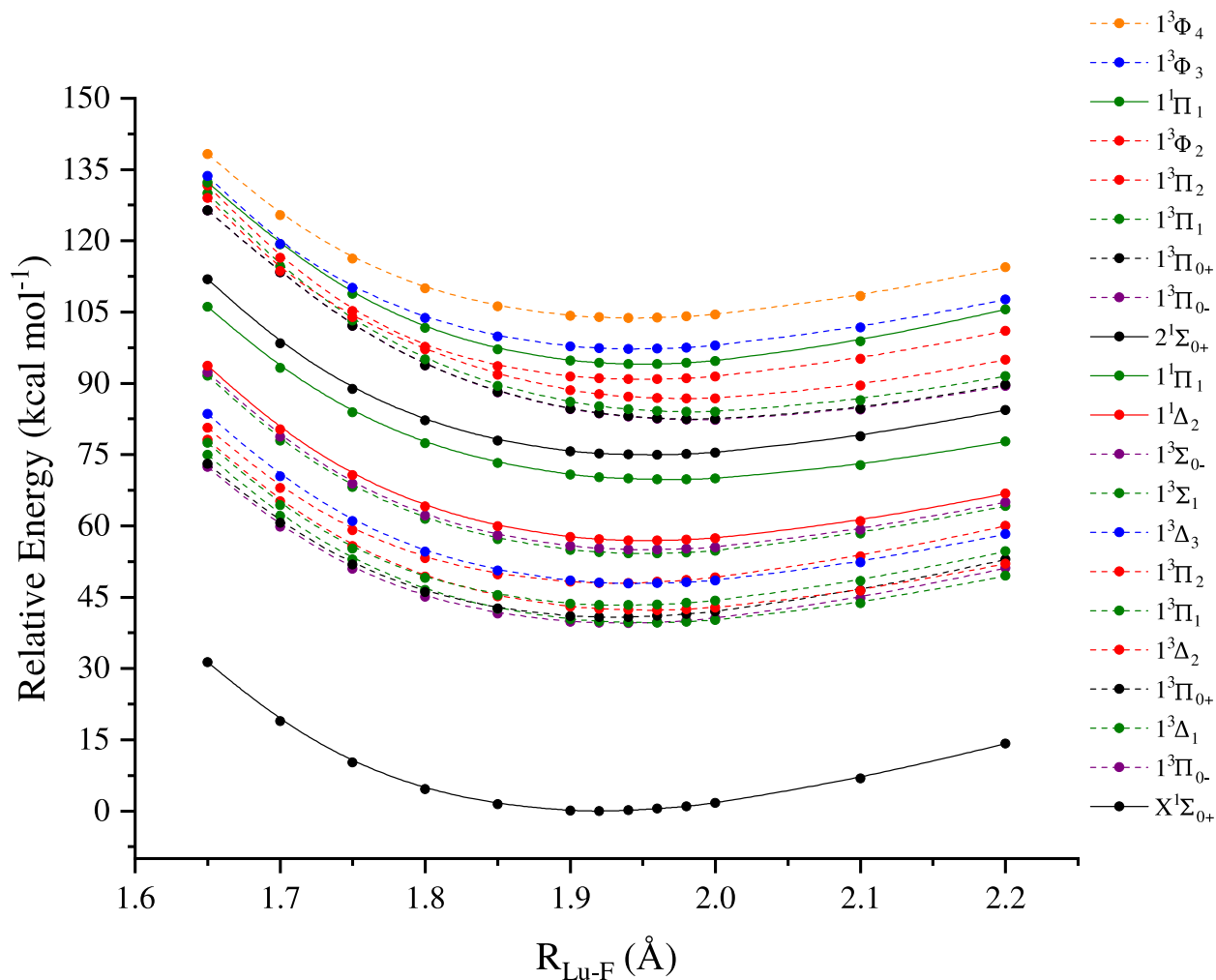


Figure B.1: Spin-orbit MRCI PECs of LuF with respect to the Lu-F distance.

Analysis of the 95% confidence limit using different extrapolation schemes

A type B evaluation can be used to assess the uncertainty of thermochemical properties and was described for composite schemes in the work by Ruscic.³⁰³ The type B evaluation allows for a rational assessment of errors for each contribution (of a composite approach) to an overall energy prediction. The errors are added together as uncertainties in order to obtain a cumulative overall absolute error of the final computational estimate. For example, in prior work, the errors have been determined for the terms including spin-orbit corrections, scalar relativistic effects, core-valence and zero-point energy (ZPE) corrections, allowing for an informed assessment of the propagation of errors in

computations. Since the dissociation energy calculations from section 4.3 were not performed with composite methods, 95% confidence intervals have been calculated using different extrapolation schemes (see Table B.1). The 95% confidence limit has been determined for the CCSD(T) and CR-CCSD(T) predictions (Table B.1) as suggested by Ruscic.³⁰³ Confidence intervals estimate the propagation of errors that arise from different estimations of the dissociation energy at CBS, i.e., there is a 95% probability that the values predicted contain a true estimate of the dissociation energy. For this purpose, four extrapolation schemes were considered as presented by Feller *et. al.* in his type B evaluation of oxalic acid.³⁰⁴ Following Fellers example, four widely used CBS extrapolation schemes were selected and analyzed concerning the overall errors.

The scheme selected are the following: a mixed exponential/gaussian by Peterson,⁶⁰

$$E_n = E_{CBS} + B e^{-(n-1)} + C e^{-(n-1)^2} \quad (\text{B.1})$$

a two-point extrapolation by Martin¹⁶⁰

$$E_n = E_{CBS} + \frac{B}{(n + 0.5)^4} \quad (\text{B.2})$$

a three-point extrapolation by Martin¹⁶⁰

$$E_n = E_{CBS} + \frac{B}{(n + 0.5)^4} + \frac{C}{(n + 0.5)^6} \quad (\text{B.3})$$

and a two-point extrapolation by Halkier *et. al.*⁵⁹:

$$E_n = E_{CBS} + B e^{-1.63n} \quad (\text{B.4})$$

where B and C are constants determined in the scheme and n is the basis set level ($n = \text{D, T, Q}$), E_n represent the energy for each basis set level, and E_{CBS} represents the energy at the CBS limit.

Table B.1: Calculation of the 95 % confidence limit of LuF dissociation energy ($\text{kcal}\cdot\text{mol}^{-1}$) using four different extrapolation schemes at CCSD(T)/FC-subval/ DKH3 level.

Scheme	D_0 (dz)	D_0 (tz)	D_0 (qz)	D_0 CBS
B.1	164.16	169.25	169.96	170.4
B.2	164.16	169.25	169.96	170.4
B.3	164.16	169.25	169.96	170.2
B.4	164.16	169.25	169.96	170.1
95% confidence limit				0.1

Table B.2: Spin-orbit MRCI potential energy curves data ($\text{kcal}\cdot\text{mol}^{-1}$): distances, relative energies and states.

Distances (Å)	$X^1\Sigma_0^+$	$1^3\Pi_{0-}$	$1^3\Delta_1$	$1^3\Pi_{0+}$	$1^3\Delta_2$	$1^3\Pi_1$	$1^3\Pi_2$	$1^3\Delta_3$	$1^3\Sigma_1$	$1^1\Sigma_0^-$	$1^1\Delta_2$
2.20	14.19	51.15	49.48	52.95	52.06	54.67	60.00	58.26	64.17	64.98	66.76
2.10	6.88	44.83	43.71	46.44	46.30	48.43	53.54	52.27	58.32	59.14	60.96
2.00	1.65	40.54	40.14	41.95	42.75	44.22	49.09	48.49	54.70	55.51	57.38
1.98	0.98	40.05	39.82	41.41	42.43	43.75	48.56	48.12	54.37	55.18	57.06
1.96	0.47	39.72	39.66	41.04	42.27	43.44	48.19	47.92	54.21	55.02	56.91
1.94	0.14	39.56	39.67	40.83	42.29	43.30	47.99	47.89	54.23	55.04	56.94
1.92	0.00	39.59	39.89	40.82	42.52	43.35	47.98	48.07	54.46	55.26	57.18
1.90	0.08	39.83	40.32	41.02	42.96	43.61	48.18	48.47	54.92	55.71	57.63
1.85	1.39	41.51	42.49	42.59	45.16	45.37	49.76	50.57	57.19	57.98	59.89
1.80	4.61	45.07	46.53	46.03	49.27	49.05	53.21	54.58	61.44	62.21	64.08
1.75	10.23	50.98	52.89	51.84	55.73	55.16	59.06	61.01	68.16	68.91	70.69
1.70	18.88	59.85	62.11	60.61	65.12	64.37	67.93	70.43	77.97	78.70	80.31
1.65	31.28	72.41	74.89	73.07	78.08	77.43	80.61	83.57	91.61	92.32	93.64
1.60	48.33	89.55	92.13	90.11	95.43	95.17	98.06	101.31	109.97	110.67	111.55
1.50	100.82	141.94	144.53	142.32	147.77	148.94	151.49	154.88	165.28	165.39	165.97

Table B.2: Table (continued)

Distances (Å)	$2^1\Sigma_{0+}$	$1^3\Pi_{0-}$	$1^3\Pi_{0+}$	$1^3\Pi_1$	$1^3\Pi_2$	$1^3\Phi_2$	$1^1\Pi_1$	$1^3\Phi_3$	$1^3\Phi_4$
2.20	84.32	89.43	89.66	91.54	94.98	100.97	105.54	107.64	114.42
2.10	78.78	84.49	84.65	86.40	89.55	95.11	98.87	101.69	108.38
2.00	75.38	82.30	82.41	84.00	86.84	91.41	94.68	97.89	104.47
1.98	75.08	82.33	82.43	83.98	86.76	91.06	94.28	97.51	104.06
1.96	74.95	82.56	82.64	84.16	86.87	90.87	94.07	97.29	103.82
1.94	74.99	82.99	83.07	84.55	87.19	90.86	94.06	97.25	103.75
1.92	75.23	83.65	83.72	85.17	87.74	91.05	94.27	97.41	103.88
1.90	75.69	84.57	84.63	86.04	88.54	91.45	94.73	97.79	104.22
1.85	77.96	88.11	88.15	89.47	91.79	93.57	97.13	99.82	106.17
1.80	82.15	93.78	93.81	95.03	97.12	97.63	101.65	103.75	110.00
1.75	88.77	102.07	102.10	103.21	103.94	105.21	108.81	110.05	116.21
1.70	98.43	113.33	113.33	114.62	113.59	116.43	119.21	119.32	125.37
1.65	111.87	126.40	126.40	129.95	129.02	131.59	132.27	133.58	138.20
1.60	129.98	144.01	144.01	149.76	149.22	151.53	150.05	152.77	155.56
1.50	184.59	197.12	197.12	202.61	208.08	208.13	208.66	209.85	209.98

APPENDIX C

LnS AND LnSe *f*-ccCA ENERGIES AND MULTIREFERENCE DIAGNOSTICS

Table C.1: Total atomization terms and dissociation energy obtained with *f*-ccCA using PW91-DKH3 geometries, along with all-electron basis sets. Theoretical and experimental dissociation energies are in kcal·mol⁻¹.

	Δ MP2 _{CBS}	Δ HF _{CBS}	Δ CV	Δ CC	Δ SO	<i>f</i> -ccCA D ₀	Exp. D ₀ ¹⁹⁵	Exp. D ₀
Category I								
PrS	45.10	69.48	1.74	6.32	-2.66	118.69	120.61	112 ± 6 ^a
PrSe	39.84	56.79	2.76	8.64	-4.82	102.35	103.68	-
SmS	50.88	46.75	-0.62	-4.92	1.17	92.70	92.52	-
SmSe	47.49	35.60	0.86	-3.60	-1.32	78.66	80.62	-
GdS	49.22	79.46	0.63	-4.55	-3.08	121.06	121.80	124 ± 6 ^a 125 ± 4 ^b
GdSe	44.24	67.96	1.59	-2.40	-5.22	105.77	106.20	102 ± 5 ^a 103 ± 4 ^c
LuS	38.34	88.25	1.22	-3.57	-3.91	119.69	119.70	120 ± 6 ^a 120 ± 4 ^b
LuSe	36.95	77.39	0.63	-1.94	-6.12	106.51	106.10	99 ± 6 ^a 100 ± 4 ^c
Category II								
NdS	29.34	88.35	1.77	-1.50	-0.22	116.73	111.15	120 ± 6 ^a 112 ± 4 ^b
NdSe	24.08	76.29	2.83	0.73	-3.84	99.21	94.53	91 ± 6 ^a 91 ± 4 ^c
EuS	50.93	42.54	0.41	-5.43	4.44	92.35	87.90	86 ± 6 ^a 86 ± 4 ^b
EuSe	43.97	37.39	1.51	-2.95	2.77	82.34	76.50	66 ± 6 ^a 72 ± 4 ^c
Category III								
TbS	-30.65	155.93	-0.66	9.14	-1.05	132.13	122.00	—
TbSe	-35.40	144.38	-24.57	11.17	-3.19	92.02	106.10	—
ErS	-68.77	130.69	-0.96	28.42	0.18	89.01	98.20	99 ± 4 ^b
ErSe	-181.13	5.43	32.59	44.24	-1.06	-100.29	82.70	—

^a Reference¹⁷⁶

^b Reference¹⁸⁹

^c Reference¹⁹⁰

Table C.2: Total atomization terms and dissociation energy obtained with f -ccCA using CCSD(T)/DKH3 geometries, along with all-electron basis sets. Theoretical and experimental dissociation energies are in kcal·mol⁻¹.

	Δ MP2 _{CBS}	Δ HF _{CBS}	Δ CV	Δ CC	Δ SO	f -ccCA D ₀	Exp. D ₀ ¹⁹⁵	Exp. D ₀
Category I								
PrS	44.75	69.67	1.66	6.46	-2.66	118.59	120.61	112 ± 6 ^a
PrSe	39.57	56.94	2.69	8.77	-4.82	102.29	103.68	-
SmS	50.33	47.08	-0.64	-4.75	1.17	92.63	92.52	-
SmSe	43.68	41.24	0.64	-2.44	-1.32	81.41	80.62	-
GdS	49	79.58	0.62	-4.46	-3.08	121.04	121.80	124 ± 6 ^a 124.5 ± 3.5 ^b
GdSe	43.84	68.21	1.54	-2.21	-5.21	105.74	106.20	102 ± 5 ^a 103 ± 4 ^c
LuS	38.14	88.31	1.23	-3.51	-3.91	119.62	119.70	120 ± 6 ^a 120.2 ± 3.5 ^b
LuSe	37	77.38	0.63	-1.95	-1.24	106.51	106.10	99 ± 6 ^a 100 ± 4 ^c
Category II								
NdS	29.36	88.34	1.78	-1.51	-0.22	116.53	111.15	120 ± 6 ^a 112.2 ± 3.5 ^b
NdSe	24.25	76.22	2.87	0.65	-3.84	99.28	94.53	91 ± 6 ^a 91 ± 4 ^c
EuS	49.64	43.5	0.36	-4.98	4.44	92.43	87.90	86 ± 6 ^a 85.9 ± 3.5 ^b
EuSe	43.65	37.69	1.47	-2.84	2.77	82.4	76.50	66 ± 6 ^a 72 ± 4 ^c
Category III								
TbS	—	—	—	—	-1.05	—	122.00	—
TbSe	—	—	—	—	-3.19	—	106.10	—
ErS	—	—	—	—	1.08	—	98.20	99.2 ± 3.5 ^b
ErSe	—	—	—	—	-1.06	—	82.70	—

^a Reference¹⁷⁶

^b Reference¹⁸⁹

^c Reference¹⁹⁰

Table C.3: Multireference considerations for lanthanide sulfides and selenides using PW91-DKH3 geometries. T_1 and D_1 are common diagnostics, $|T_{1\max}|$ and $|T_{2\max}|$ are the absolute values of the largest amplitudes, and S^2 is the spin contamination.

	T_1	D_1	$ T_{1\max} $	$ T_{2\max} $	Spin contamination (S^2)
PrS	0.03	0.10	0.12	—	0.030
NdS	0.03	0.08	0.07	—	0.010
SmS	0.04	0.11	0.13	0.11	0.010
EuS	0.03	0.10	0.13	—	0.001
GdS	0.04	0.11	0.13	0.2	0.030
TbS	0.03	0.10	0.13	—	0.007
ErS	0.03	0.10	0.13	0.06	0.002
LuS	0.03	0.09	0.08	—	0.003
PrSe	0.03	0.10	0.13	—	0.040
NdSe	0.03	0.09	0.07	—	0.010
SmSe	0.03	0.10	0.13	0.11	0.010
EuSe	0.03	0.12	0.15	—	0.001
GdSe	0.03	0.10	0.12	0.19	0.040
TbSe	0.03	0.10	0.13	—	0.009
ErSe	0.05	0.15	0.21	0.1	0.010
LuSe	0.03	0.10	0.09	—	0.003

Table C.4: Multireference considerations for lanthanide sulfides and selenides using CCSD(T)/ECP geometries. T_1 and D_1 are common diagnostics, $|T_{1\max}|$ and $|T_{2\max}|$ are the absolute values of the largest amplitudes, and S^2 is the spin contamination.

	T_1	D_1	$ T_{1\max} $	$ T_{2\max} $	Spin contamination (S^2)
PrS	0.03	0.1	0.12	—	0.03
NdS	0.03	0.10	0.08	—	0.01
SmS	0.04	0.11	0.13	0.11	0.004
EuS	0.03	0.11	0.13	—	0.001
GdS	0.04	0.11	0.13	0.2	0.005
TbS	—	—	—	—	—
ErS	—	—	—	—	—
LuS	0.03	0.1	0.08	—	0.003
PrSe	0.03	0.1	0.13	—	0.04
NdSe	0.03	0.09	0.07	—	0.01
SmSe	0.03	0.1	0.13	0.1	0.003
EuSe	0.03	0.11	0.15	—	0.001
GdSe	0.04	0.11	0.12	0.2	0.004
TbSe	—	—	—	—	—
ErSe	—	—	—	—	—
LuSe	0.03	0.1	0.09	—	0.003

APPENDIX D

PFAS ENTHALPIES OF FORMATION, EXTRAPOLATED ENTHALPIES AND OPTIMIZED GEOMETRIES

Methodology used to determine the estimated enthalpies

As the computational costs rise substantially with the increase in molecule size, the enthalpies were estimated for larger molecules simply to provide qualitative energy estimates. This was possible due to the observed (somewhat) systematic change in energy with respect to increasing molecule size. The process used is as follows for each (DLPNO and ccCA) method:

- The enthalpy differences between consecutive molecules of increasing size are calculated.
- The enthalpy difference is calculated as follows:

$$\Delta(\Delta H_f^\circ) = \Delta H_f^\circ(\text{CF}_3-(\text{CF}_2)_{n+1}-\text{COOH}) - \Delta H_f^\circ(\text{CF}_3-(\text{CF}_2)_n-\text{COOH})$$

where n is the number of carbon atoms in the smaller of the two molecules in comparison.

- The average enthalpy difference is determined.
- This difference is added to the last DLPNO or ccCA calculated or estimated enthalpy.

The uncertainty is determined using the root mean square deviation of the enthalpy difference $\Delta(\Delta H_f^\circ)$ as compared to the computed enthalpic difference.

Table D.1: ccCA energy differences calculated between subsequent PFAS molecules.

	ccCA (kcal·mol ⁻¹)	$\Delta(\Delta H_f^\circ)$ (kcal·mol ⁻¹)
C ₂ F ₃ HO ₂	-247.51	—
C ₃ F ₅ HO ₂	-347.84	-100.33
C ₄ F ₇ HO ₂	-444.94	-97.1
C ₅ F ₉ HO ₂	-543.78	-98.84

As an example, for the PFAS with ccCA, the average enthalpic difference $\Delta(\Delta H_f^\circ)$ for PF₄EtA to PF₅PeA molecule is -98.76 kcal·mol⁻¹, resulting from $(-100.33 + (-97.10) + (-98.84))/3$. Then the

PFHxA enthalpy of formation is estimated as:

$$\Delta H_f^{\text{ccCA}}(\text{C}_6\text{F}_{11}\text{HO}_2) = \Delta H_f^{\text{ccCA}}(\text{C}_5\text{F}_9\text{HO}_2) + (-98.76)$$

which here gives:

$$\Delta H_f^{\text{ccCA}}(\text{C}_6\text{F}_{11}\text{HO}_2) = -543.78 + (-98.76) = -642.54 \text{ kcal} \cdot \text{mol}^{-1}$$

For PFHpA and PFOA, the same formula is used with the average difference in enthalpy values building upon PFHxA enthalpy and gives -741.29 and -840.05 kcal·mol⁻¹, respectively.

The uncertainty is calculated using the root mean square deviation formula:

$$RMSE = \sqrt{\frac{\sum (x_e - x_0)^2}{n}}$$

where x_e is the average enthalpic difference (in this example, $x_e = -98.76 \text{ kcal} \cdot \text{mol}^{-1}$), x_0 represents the actual enthalpy difference ($\Delta(\Delta H_f^\circ)$) of a given molecule and n is the number of known values (in this example, $n=3$). Using the three calculated enthalpy variation, the uncertainty obtained is 1.32 kcal·mol⁻¹.

Enthalpies of formation from experiment

Table D.2: Experimental enthalpies of formation (kcal·mol⁻¹) for the molecules involved in the different isogyric reactions. All the data were taken from the ATcT database.²⁶²

Molecules	ΔH_f°
CH ₄	-17.81 ± 0.01
CF ₄	-223.09 ± 0.06
HF	-65.18 ± 0.01
CF ₃ H	-166.32 ± 0.10
F ₂	0.0
CO ₂	-94.04 ± 0.01
CH ₃ OH	-47.97 ± 0.04
CH ₃ CH ₂ OH	-56.07 ± 0.05

Relative computational time used per method

The following calculation were run on an Intel® Xeon® Gold 6148 CPU @ 2.40 GHz nodes using 10 processor and 10 Gb of global memory at the Institute for Cyber-Enabled Research (iCER) at Michigan State University.

Table D.3: Total CPU time in second for different calculations.

Molecules	B3LYP ^a	M06-2X ^a	MP2/aug-cc-pVQZ ^b	DLPNO-CCSD(T)/aug-cc-pVQZ ^c
C ₃ H ₆ O ₂	150.1	131.8	4519	16695
C ₃ F ₈	141.1	157.8	14673	67107
C ₃ F ₅ HO ₂	203.2	196.3	17141	56432

^a Calculation with NWChem 6.1.

^b The MP2 calculation (using Molpro 2020) with the aug-cc-pVQZ basis set corresponds to the most expensive step of the ccCA methodology.

^c The DLPNO-CCSD(T) calculation (using ORCA 4.2.1) with the aug-cc-pVQZ basis set corresponds to the most expensive step of the DLPNO-CBS method.

Figures showing comparisons of enthalpies of formation using a number of different methods

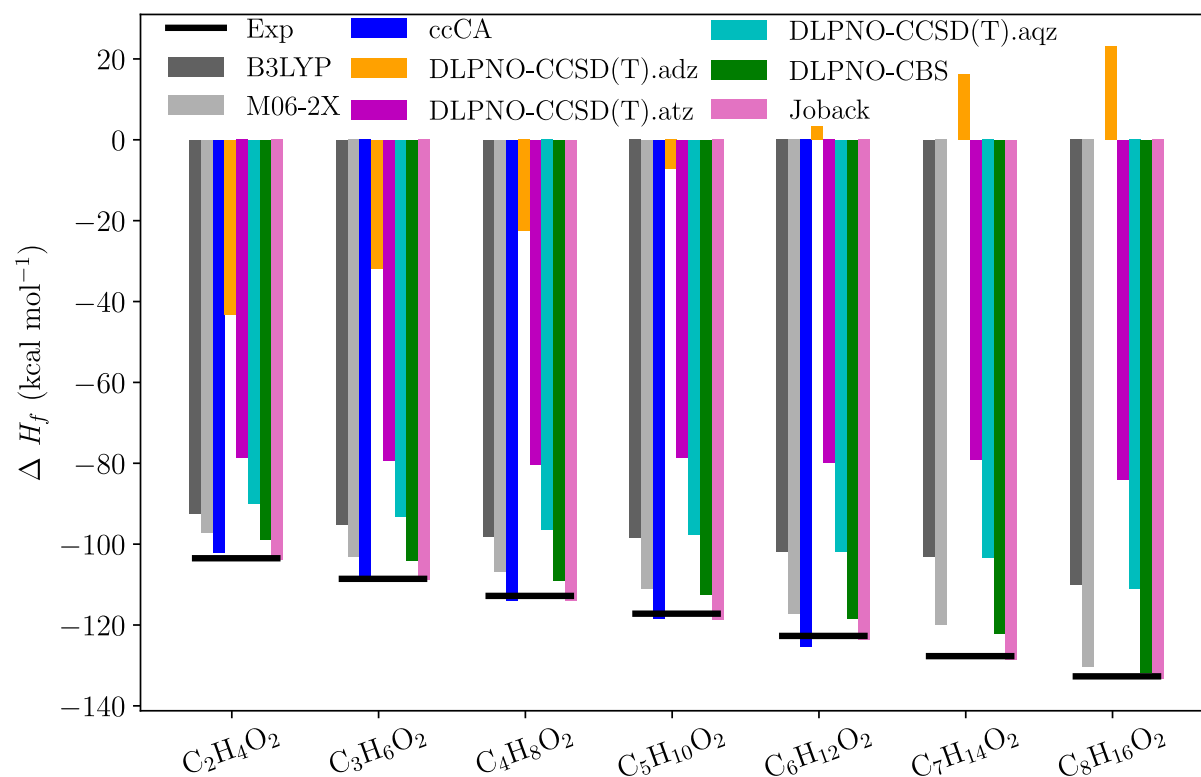


Figure D.1: ΔH_f° (in kcal·mol⁻¹) for alkanic acid determined using the total atomization approach.

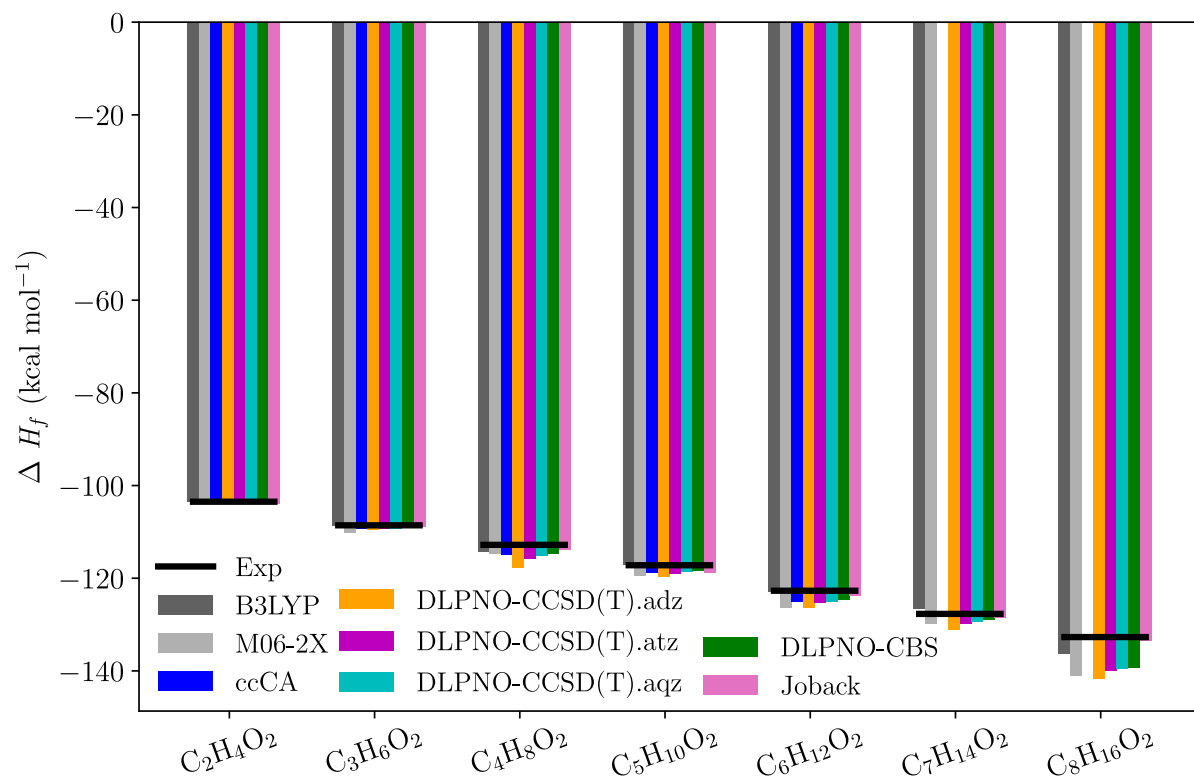


Figure D.2: ΔH_f° (in $\text{kcal}\cdot\text{mol}^{-1}$) for alkanolic acid determined using the homodesmotic approach.

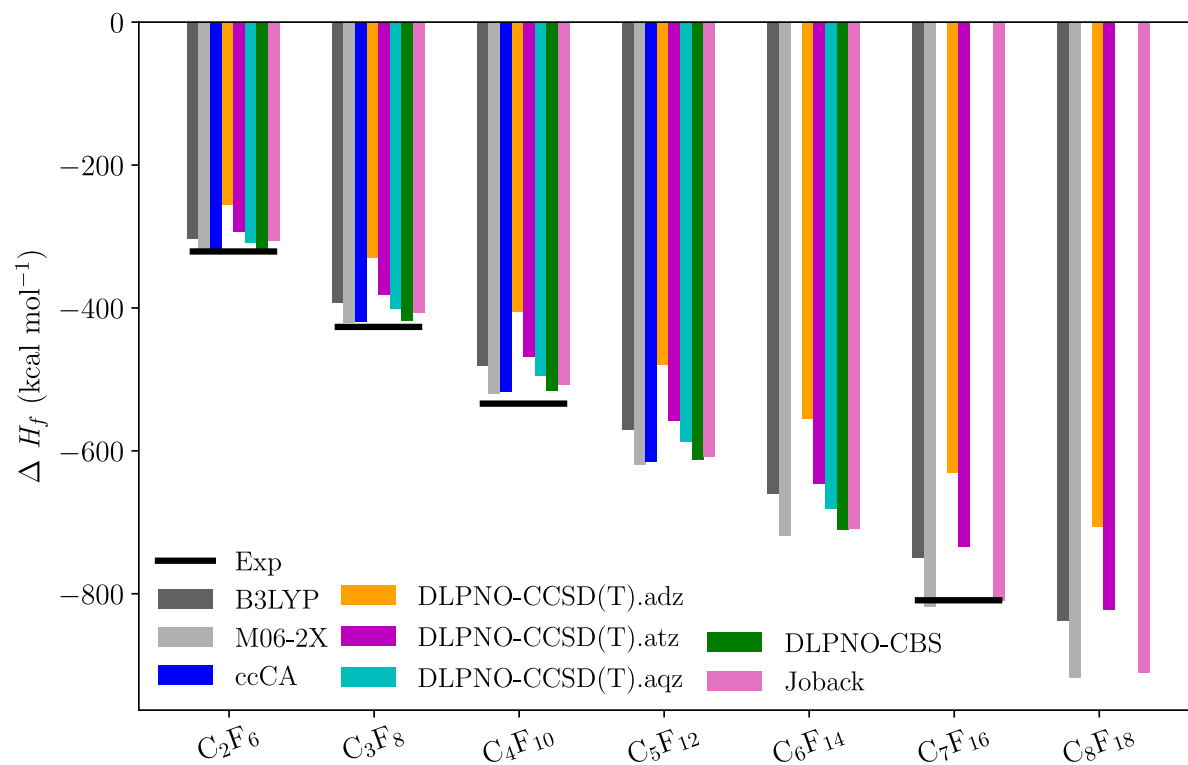


Figure D.3: ΔH_f° (in $\text{kcal}\cdot\text{mol}^{-1}$) of the perfluoroalkanes determined using the total atomization approach.

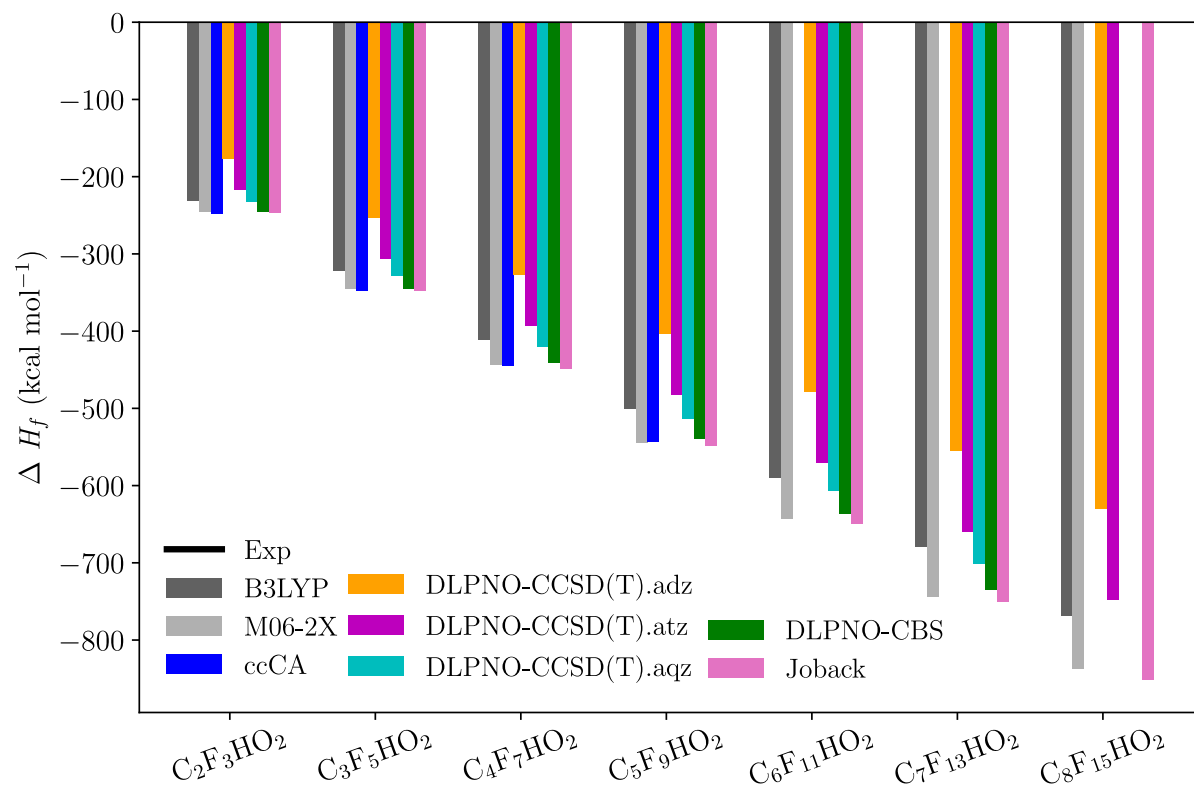


Figure D.4: ΔH_f° (in kcal·mol⁻¹) of PFAS determined using the total atomization approach.

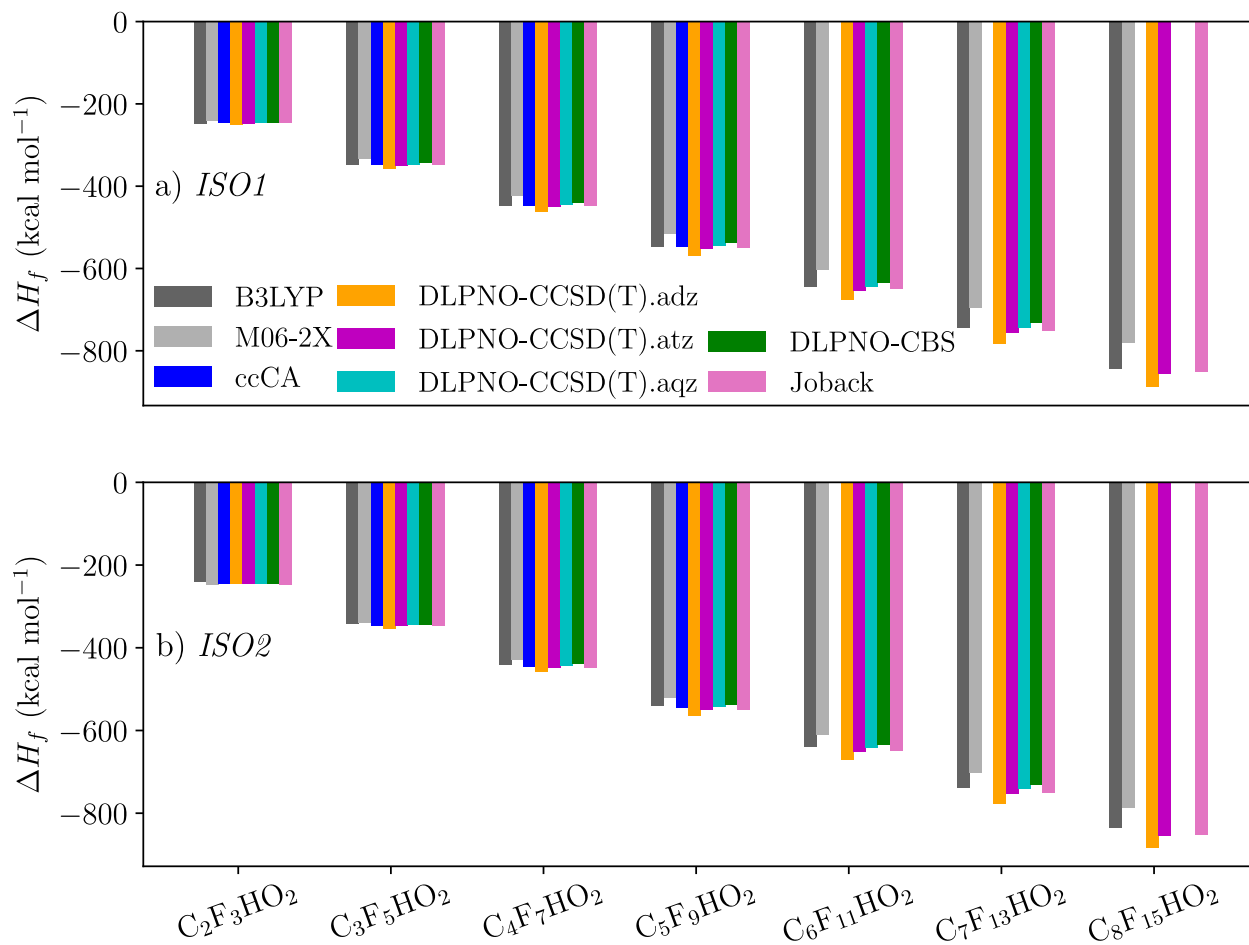


Figure D.5: Comparison between ΔH_f° (in kcal·mol⁻¹) of PFAS using a) ISO1 and b) ISO2.

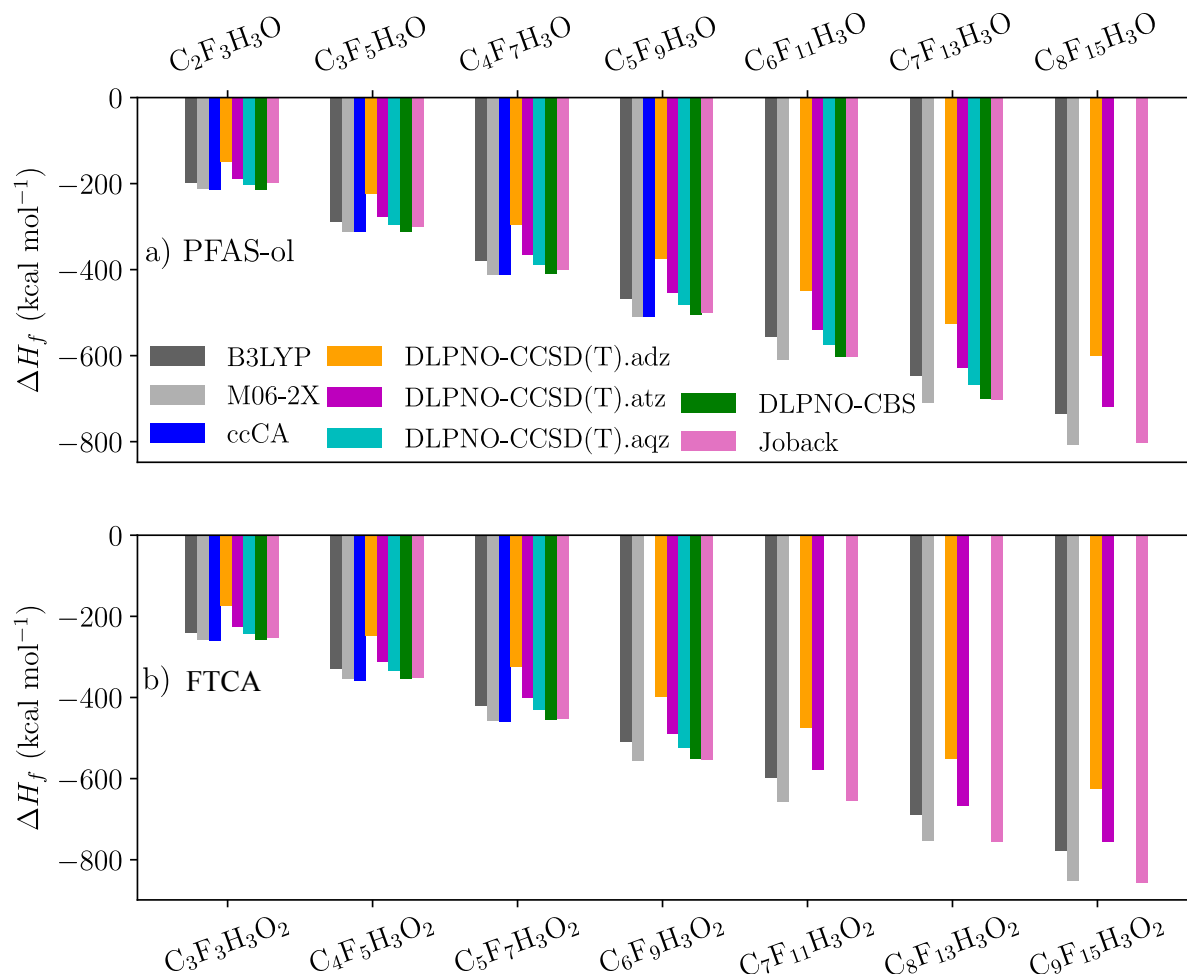


Figure D.6: ΔH_f° (in $\text{kcal}\cdot\text{mol}^{-1}$) of PFAS using a) PFAS-ol and b) FTCA using the total atomization approach.

Optimized Geometries for all molecules used in this study

For all molecules examined in this study (see chapter 6), the geometries were optimized at the B3LYP/6-31G level of theory using NWChem 6.1. All geometries are given in angstrom.

Perfluoroalkanes

Table D.4: Optimized geometry of C₂F₆ in Å.

Energy (a.u) C ₂ F ₆ : -675.0870054			
C	-0.00000000	0.00000000	0.76576260
C	0.00000000	-0.00000000	-0.76576260
F	-1.24922451	-0.33472870	1.23724806
F	0.33472870	1.24922451	1.23724806
F	0.91449581	-0.91449581	1.23724806
F	-0.91449581	0.91449581	-1.23724806
F	-0.33472870	-1.24922451	-1.23724806
F	1.24922451	0.33472870	-1.23724806

Table D.5: Optimized geometry of C₃F₈ in Å.

Energy (a.u) C ₃ F ₈ : -912.796896			
C	1.30638971	0.00000000	-0.226443
C	0.00000000	0.00000000	0.583208
F	1.37649802	-1.11993526	-1.026296
F	2.38134446	0.00000000	0.630762
F	1.37649802	1.11993526	-1.026296
C	-1.30638971	0.00000000	-0.226443
F	0.00000000	1.13296665	1.391197
F	0.00000000	-1.13296665	1.391197
F	-1.37649802	1.11993526	-1.026296
F	-2.38134446	0.00000000	0.630762
F	-1.37649802	-1.11993526	-1.026296

Table D.6: Optimized geometry of C₄F₁₀ in Å.

Energy (a.u) C ₄ F ₁₀ : -1150.531818			
C	-0.38885413	1.96353729	0.118689
C	0.43455372	0.65984611	0.081888
F	-1.14375914	2.08298267	-1.029202
F	0.46201798	3.04112295	0.202805
F	-1.22647822	1.98016924	1.209434
C	-0.35510884	-0.61609574	-0.276545
F	1.01108155	0.49107963	1.340543
F	1.44889688	0.82221960	-0.855142
C	0.34874992	-1.94869439	0.051246
F	-0.61768497	-0.59713154	-1.641759
F	-1.56748635	-0.60017771	0.412643
F	1.63558784	-1.94209673	-0.444402
F	-0.34306535	-2.98712152	-0.527882
F	0.39303344	-2.15564171	1.410260

Table D.7: Optimized geometry of C₅F₁₂ in Å.

Energy (a.u) C ₅ F ₁₂ : -1388.253608			
C	0.13730111	-2.61575727	-0.133275
C	-0.59055179	-1.26443920	0.031563
F	0.63499045	-2.75098152	-1.408072
F	-0.74706740	-3.64285089	0.102456
F	1.17668844	-2.70979824	0.768258
C	0.32740337	-0.01970537	0.057558
F	-1.30286801	-1.30990607	1.223786
F	-1.48738692	-1.13958635	-1.028021
C	-0.38808834	1.32640050	-0.204662
F	1.31722315	-0.17805847	-0.910969
F	0.93894816	0.03928237	1.308756
F	-0.64007996	1.43177627	-1.567116
C	0.39631023	2.58385520	0.227204
F	-1.60065997	1.33241569	0.482542
F	1.68423783	2.53345750	-0.263327
F	0.44155255	2.68531322	1.597903
F	-0.22455654	3.70439504	-0.274242

Table D.8: Optimized geometry of C₆F₁₄ in Å.

Energy (a.u) C ₆ F ₁₄ : -1625.975301			
C	-3.25238023	-0.35429955	0.044396
C	-1.92292632	0.33947282	0.411798
F	-3.46104087	-0.32395166	-1.314505
F	-4.29405459	0.29879940	0.661620
F	-3.23794159	-1.66638191	0.468748
C	-0.64297001	-0.43386905	0.011361
F	-1.91063221	0.52773110	1.788359
F	-1.91236147	1.58346058	-0.216759
C	0.65074175	0.42119819	-0.013189
F	-0.83303525	-0.97092457	-1.259998
F	-0.46990348	-1.47770363	0.917274
F	0.64741563	1.18074045	-1.181083
C	1.96640928	-0.39311617	0.036679
F	0.63617364	1.28249650	1.081443
F	1.84056161	-1.50303088	-0.796967
F	2.16602766	-0.82710836	1.341405
C	3.23208345	0.37483392	-0.401734
F	3.22598809	0.59361386	-1.759383
F	4.34784408	-0.36397163	-0.082655
F	3.30249417	1.58828652	0.249749

Table D.9: Optimized geometry of C₇F₁₆ in Å.

Energy (a.u) C ₇ F ₁₆ : -1863.696959			
C	-0.20658419	3.90961435	-0.210473
C	0.42941863	2.59169121	0.282103
F	-0.07647680	4.03633894	-1.573577
F	0.42526500	4.97384645	0.390133
F	-1.54515250	3.94485439	0.119550
C	-0.34383182	1.30476744	-0.096786
F	0.52104622	2.65841497	1.666910
F	1.71348917	2.52023052	-0.255151
C	0.48250620	-0.00470631	0.013293
F	-0.78779819	1.42999450	-1.411235
F	-1.45063339	1.20553745	0.743101
F	1.32067789	-0.08257775	-1.095625
C	-0.37109252	-1.29964883	0.077185
F	1.25972338	0.05922426	1.166608
F	-1.42821293	-1.18168576	-0.822199
F	-0.88873503	-1.41653753	1.365186
C	0.39937714	-2.59991998	-0.259275
F	0.56618198	-2.66869492	-1.636943
C	-0.28554970	-3.90650097	0.196538
F	1.65290585	-2.55058965	0.348086
F	-1.60393227	-3.91911933	-0.207920
F	-0.23413581	-4.03465887	1.564747
F	0.36071983	-4.98166328	-0.368163

Table D.10: Optimized geometry of C₈F₁₈
in Å.

Energy (a.u) C ₈ F ₁₈ : -2101.418614			
C	-0.09319889	4.56133266	0.286812
C	0.39818412	3.24375672	-0.351103
F	-1.46973298	4.62982497	0.238529
F	0.42737440	5.62754255	-0.409396
F	0.31535464	4.65280380	1.596673
C	-0.31382544	1.96344991	0.150037
F	1.76195588	3.13328007	-0.085204
F	0.21102035	3.34320215	-1.724202
C	0.43869841	0.64028653	-0.155852
F	-1.57135527	1.90958073	-0.446658
F	-0.47560174	2.06310729	1.530063
F	0.96659267	0.72012245	-1.441496
C	-0.44513763	-0.63497305	-0.073643
F	1.48217390	0.51485311	0.757229
F	-1.21540824	-0.70430784	-1.231166
F	-1.28858909	-0.51817889	1.027842
C	0.35142122	-1.96013387	0.066830
F	1.46609354	-1.90098871	-0.766378
C	-0.44743647	-3.23658880	-0.294067
F	0.78310891	-2.07072846	1.386702
F	-0.53000271	-3.32757420	-1.678045
F	-1.73361721	-3.12672555	0.231490
C	0.15532895	-4.55858919	0.228538
F	1.49552782	-4.62924906	-0.088830
F	-0.49457329	-5.61978957	-0.357632
F	0.01082251	-4.65660336	1.592534

Table D.11: Optimized geometry of $\text{C}_2\text{F}_3\text{O}_2\text{H}$ in Å.

Energy (a.u) $\text{C}_2\text{F}_3\text{O}_2\text{H}$: -526.630195			
C	-0.53529140	-0.24437170	-0.030642
C	0.83201048	0.43509830	0.003914
F	-1.51266726	0.66661670	-0.352853
F	-0.83578942	-0.78934420	1.206833
F	-0.56085533	-1.26334498	-0.965459
O	1.79893341	-0.49422979	0.225600
O	1.00880665	1.63781145	-0.136133
H	2.69224573	-0.08842291	0.266355

Table D.12: Optimized geometry of $\text{C}_3\text{FF}_5\text{O}_2\text{H}$ in Å.

Energy (a.u) $\text{C}_3\text{F}_5\text{O}_2\text{H}$: -764.354871			
C	0.28831345	-1.04321695	0.217696
C	-0.58287330	0.14848741	-0.213594
C	0.14205667	1.49884927	-0.229122
O	0.56233219	1.80831321	1.027048
H	1.02531714	2.67387645	1.057332
O	0.30055780	2.18476341	-1.228330
F	0.54416523	-1.01209194	1.568245
F	-0.32892170	-2.23583625	-0.082275
F	1.49499267	-0.98621770	-0.461122
F	-1.67953887	0.21867080	0.651293
F	-1.04639461	-0.11080789	-1.496686

Table D.13: Optimized geometry of $42\text{ F}_7\text{O}_2\text{H}$ in Å.

Energy (a.u) $\text{C}_4\text{F}_7\text{O}_2\text{H}$: -1002.074868			
C	-0.12319170	1.79135874	-0.005065
C	0.52297118	0.39132879	0.002471
C	-0.43889071	-0.80934370	0.007300
C	0.25489725	-2.17672206	-0.004335
F	0.86643583	2.74901757	0.011643
F	-0.88773222	1.97311220	-1.136451
F	-0.92243585	1.96910686	1.102969
F	1.32285351	0.31455017	1.140391
F	1.32389430	0.30474124	-1.134449
F	-1.26802516	-0.70959974	-1.115035
F	-1.24345064	-0.72166556	1.148740
O	1.46780836	-2.34941199	0.004135
O	-0.68538434	-3.15598151	-0.021163
H	-0.27874006	-4.05009963	-0.025806

Table D.14: Optimized geometry of $\text{C}_5\text{F}_9\text{O}_2\text{H}$ in Å.

Energy (a.u) $\text{C}_5\text{F}_9\text{O}_2\text{H}$: -1239.798277			
C	-2.36930776	-0.29505450	-0.016240
C	-1.05618149	0.51443306	0.036420
F	-2.54138860	-0.87239311	-1.252510
F	-3.43493848	0.54162255	0.226544
F	-2.35833130	-1.28539809	0.944322
C	0.23912900	-0.32682021	0.049835
F	-1.08903435	1.29485466	1.186231
F	-1.03747890	1.35062426	-1.078565
C	1.54184004	0.42903685	-0.309918
F	0.09478553	-1.38550006	-0.845811
F	0.39872862	-0.86184155	1.330769
C	2.79797660	-0.35292838	0.105987
F	1.50908821	1.68080769	0.303124
F	1.58083284	0.61014459	-1.689265
O	3.26549730	0.07326262	1.305931
H	4.05872226	-0.43474132	1.585294
O	3.27500440	-1.24952192	-0.577863

Table D.15: Optimized geometry of $\text{C}_6\text{F}_{11}\text{O}_2\text{H}$ in Å.

Energy (a.u) $\text{C}_6\text{F}_{11}\text{O}_2\text{H}$: -1477.519992			
C	-2.99065966	-0.14340199	0.361136
C	-1.67404883	0.54461105	-0.058984
F	-4.04710036	0.70502611	0.119503
F	-2.97742510	-0.45153690	1.701580
F	-3.17833574	-1.30087302	-0.364572
C	-0.42943594	-0.37544365	-0.095275
F	-1.44136307	1.58285166	0.842193
F	-1.86484481	1.08069955	-1.326610
C	0.92979547	0.36654914	-0.098871
F	-0.51206286	-1.16068622	-1.242813
F	-0.47013542	-1.20872589	1.021200
C	2.14846045	-0.46712659	-0.569508
F	1.19381729	0.80511602	1.200313
F	0.83078489	1.48360787	-0.926155
C	3.48424617	0.18987666	-0.185666
F	2.10151675	-0.57403320	-1.955970
F	2.04822255	-1.74469072	-0.019648
O	3.95993743	-0.31599025	0.979767
H	4.80454462	0.11220271	1.241589
O	4.01066617	1.06401213	-0.862092

Table D.16: Optimized geometry of $\text{C}_7\text{F}_{13}\text{O}_2\text{H}$ in Å.

Energy (a.u) $\text{C}_7\text{F}_{13}\text{O}_2\text{H}$: -1715.242896			
C	-3.63189312	-0.65274505	-0.076633
C	-2.38830729	0.16569608	0.333077
F	-3.49497135	-1.96436697	0.327342
F	-3.81383423	-0.61972941	-1.439359
F	-4.74844321	-0.11991885	0.525812
C	-1.02754294	-0.46943957	-0.044785
F	-2.43011749	0.33273625	1.712009
F	-2.49065673	1.41378085	-0.278635
C	0.17598539	0.51053821	-0.027085
F	-1.13707909	-1.00728667	-1.325363
F	-0.77453335	-1.50297885	0.855165
C	1.56625683	-0.17398980	0.031830
F	0.04692534	1.34045057	1.083571
F	0.11739232	1.28462751	-1.183466
C	2.76566318	0.71954285	-0.370452
F	1.54418566	-1.29033017	-0.798601
F	1.79145667	-0.59759099	1.345188
F	2.55625903	1.99421341	0.147033
F	2.81854005	0.80437670	-1.762029
C	4.11379884	0.20651092	0.161788
O	4.40790771	-0.98877366	-0.414428
H	5.26674364	-1.34781427	-0.101333
O	4.80491527	0.80683601	0.971056

Table D.17: Optimized geometry of $\text{C}_8\text{F}_{15}\text{O}_2\text{H}$ in Å

Energy (a.u) $\text{C}_8\text{F}_{15}\text{O}_2\text{H}$: -1952.963328			
C	-1.30819935	-0.91880901	-0.672732
C	-0.55826923	0.10340399	0.223537
C	0.98804703	-0.04625292	0.217964
F	-1.00673047	-0.04873080	1.532744
F	-0.88110424	1.38306242	-0.221288
C	1.75838305	1.19529698	0.735815
F	1.39872523	-0.30000160	-1.088644
F	1.31824161	-1.13677298	1.017696
C	3.22957515	0.94429190	1.155131
F	1.08429439	1.71660850	1.838836
F	1.76343063	2.16140216	-0.271851
C	4.02989887	2.25062747	1.283615
F	3.81745226	0.09497147	0.218123
F	3.23295671	0.31164273	2.394109
O	4.68657350	2.53437985	0.131026
O	4.03630305	2.92674524	2.304219
C	-2.81371952	-1.08158866	-0.349941
F	-0.70323025	-2.16560190	-0.534677
F	-1.18875978	-0.50655175	-1.998560
C	-3.65434337	-1.75271606	-1.457966
F	-3.36335619	0.17740947	-0.114137
F	-2.93826217	-1.85249164	0.799314
F	-4.89698351	-2.07001557	-0.959471
F	-3.04138177	-2.91059904	-1.888467
F	-3.81705656	-0.90618776	-2.529598
H	5.18619343	3.37770778	0.197495

Table D.18: Optimized geometry of $\text{C}_2\text{F}_3\text{H}_3\text{O}$ in Å.

Energy (a.u) $\text{C}_2\text{F}_3\text{H}_3\text{O}$: -452.636565			
C	-0.00315507	-0.39077246	-0.000274
C	0.78074644	0.89697556	0.003417
F	0.89388047	-1.46075774	0.015533
F	-0.79336453	-0.53031316	-1.127246
F	-0.82038764	-0.52002108	1.108618
O	-0.16755908	1.97822270	-0.015365
H	1.43125786	0.88965363	-0.879277
H	1.40720260	0.89867991	0.903346
H	0.29069736	2.83995793	0.002983

Table D.19: Optimized geometry of $\text{C}_3\text{F}_5\text{H}_3\text{O}$ in Å.

Energy (a.u) $\text{C}_3\text{F}_5\text{H}_3\text{O}$: -690.361195			
C	-0.58099113	-0.38560486	-0.021367
C	0.00897042	-1.63159231	0.602301
F	-1.73492008	-0.06630623	0.722065
F	-0.98024519	-0.62398016	-1.339931
C	0.30469937	0.86793118	-0.028080
O	1.22849192	-1.93570932	-0.098157
H	-0.74578364	-2.42140467	0.501619
H	0.17257582	-1.42819393	1.666328
H	1.68273799	-2.69927359	0.305861
F	0.85813512	1.04580143	1.233030
F	-0.44029271	1.99112524	-0.328264
F	1.32448499	0.77531037	-0.943318

Table D.20: Optimized geometry of $\text{C}_4\text{F}_7\text{H}_3\text{O}$ in Å.

Energy (a.u) $\text{C}_4\text{F}_7\text{H}_3\text{O}$: -928.083437			
C	-0.61225897	-0.96886330	-0.096218
C	-0.27725431	-2.35183665	0.424669
F	-1.73220841	-0.51770629	0.638630
F	-0.98888448	-1.01948637	-1.440409
C	0.49204072	0.09902870	0.043087
O	0.88483253	-2.81748157	-0.284665
H	-1.15819096	-2.97877056	0.238460
H	-0.11561906	-2.27421221	1.504425
H	1.30482217	-3.56504868	0.181245
F	1.12177075	-0.08684868	1.283111
C	0.02860450	1.56911680	-0.002462
F	1.43365261	-0.06838838	-0.962741
F	-0.66668690	1.90755241	1.137334
F	-0.77944871	1.78749273	-1.098750
F	1.12527425	2.39991620	-0.096944

Table D.21: Optimized geometry of $\text{C}_5\text{F}_9\text{H}_3\text{O}$ in Å.

Energy (a.u) $\text{C}_5\text{F}_9\text{H}_3\text{O}$: -1165.807779			
C	-0.48840389	-1.78359531	-0.165904
C	0.15430647	-3.05037513	0.390942
F	-1.75388328	-1.64728250	0.444234
F	-0.69118385	-1.88830458	-1.543750
C	0.30229962	-0.48178166	0.093929
O	1.50215003	-3.23936730	-0.067487
H	-0.42746311	-3.89076557	0.008772
H	0.06854941	-3.02478560	1.481370
H	2.13168244	-2.70585503	0.458361
F	0.88792910	-0.59006822	1.369031
C	-0.50043622	0.83611147	0.057824
F	1.34237453	-0.39141979	-0.832627
F	-1.23853059	0.93836956	1.232689
F	-1.37449422	0.80101903	-1.026488
C	0.34056742	2.12357600	-0.072550
F	1.35749447	2.12687186	0.861086
F	-0.46025185	3.22135745	0.147738
F	0.88862842	2.23022774	-1.329385

Table D.22: Optimized geometry of $\text{C}_6\text{F}_{11}\text{H}_3\text{O}$ in Å.

Energy (a.u) $\text{C}_6\text{F}_{11}\text{H}_3\text{O}$: -1403.526963			
C	-0.43656497	-2.35359356	-0.214025
C	0.05997694	-3.64663951	0.403230
F	-1.75302054	-2.15963853	0.259540
F	-0.49758405	-2.46132606	-1.604563
C	0.37145901	-1.08002392	0.131342
O	1.41731904	-3.83992058	-0.030496
H	-0.61736050	-4.43535227	0.052239
H	-0.02640511	-3.55915500	1.491364
H	1.75881938	-4.70692510	0.259738
F	0.74280535	-1.16087349	1.478863
C	-0.38316900	0.26240890	-0.059502
F	1.52228218	-1.03991509	-0.649136
F	-1.26688690	0.43189106	1.007177
F	-1.11832121	0.20655150	-1.241674
C	0.52267878	1.51615890	-0.130396
F	1.52166029	1.41322988	0.836880
C	-0.19672874	2.86269753	0.099674
F	1.11409764	1.56901662	-1.387124
F	-0.56526691	3.00640026	1.417463
F	-1.32213749	2.94940242	-0.692680
F	0.65167093	3.89687765	-0.228015

Table D.23: Optimized geometry of $C_7F_{13}H_3O$ in Å.

Energy (a.u) $C_7F_{13}H_3O$: -1641.248795			
C	-0.33387637	-3.01566777	-0.302795
C	0.16013568	-4.29095841	0.352489
F	-1.70900959	-2.90912612	0.002166
F	-0.21262409	-3.09269195	-1.691219
C	0.34874258	-1.70686794	0.163357
O	1.57014965	-4.40058391	0.087989
H	-0.41282495	-5.11326948	-0.093928
H	-0.07128380	-4.23246798	1.420525
H	1.99422581	-5.03472417	0.696537
F	0.57206189	-1.80957963	1.542689
C	-0.46515436	-0.40656737	-0.077473
F	1.57493066	-1.57082503	-0.478805
F	-1.47082552	-0.32721190	0.885393
F	-1.05102320	-0.47002030	-1.339078
C	0.36763522	0.90162883	-0.007918
F	1.26830325	0.80840237	1.051476
C	-0.46477115	2.19155573	0.196788
F	1.08515309	1.03324362	-1.195215
F	-0.84852228	2.26481993	1.530973
F	-1.60657783	2.12930434	-0.600063
C	0.26532315	3.50710705	-0.149451
F	1.50308178	3.54286410	0.458049
F	-0.47746934	4.57462677	0.302704
F	0.42911455	3.63647233	-1.508827

Table D.24: Optimized geometry of $\text{C}_8\text{F}_{15}\text{H}_3\text{O}$ in Å.

Energy (a.u) $\text{C}_8\text{F}_{15}\text{H}_3\text{O}$: -1878.972851			
C	-0.42610450	-3.71424454	-0.381060
C	0.00617135	-4.99434989	0.328177
F	-1.81041769	-3.55362906	-0.156906
F	-0.22686145	-3.81491168	-1.759413
C	0.28012007	-2.42718277	0.104092
O	1.42148116	-5.22453278	0.252419
H	-0.47344440	-5.82006302	-0.200168
H	-0.37274615	-4.96137128	1.354065
H	1.90254750	-4.69965026	0.923644
F	0.44619881	-2.53885283	1.495333
C	-0.44938990	-1.09084062	-0.182390
F	1.55222908	-2.37075776	-0.467804
F	-1.49554235	-0.95960697	0.729623
F	-0.97191960	-1.13914134	-1.470924
C	0.44601023	0.17440097	-0.076218
F	1.27824952	0.03840160	1.032883
C	-0.33570853	1.50750326	0.074242
F	1.22979037	0.25409227	-1.224255
F	-0.77617219	1.60951746	1.392600
F	-1.44462139	1.47631520	-0.767502
C	0.48151155	2.77889103	-0.263099
F	1.76222614	2.64800098	0.271988
C	-0.10867579	4.10132265	0.272700
F	0.57747287	2.88923786	-1.644644
F	0.02421137	4.17984138	1.639259
F	-1.44476880	4.19581800	-0.055833
F	0.56106028	5.16232905	-0.292528

Table D.25: Optimized geometry of $\text{C}_3\text{F}_3\text{H}_3\text{O}_2$ in Å.

Energy (a.u) $\text{C}_3\text{F}_3\text{H}_3\text{O}_2$: -565.950117				
C	0.13161792	-1.03758235	-0.139207	
C	0.89056693	0.23180937	0.131176	
F	0.90114204	-2.15813696	0.109414	
F	-0.33205401	-1.11520001	-1.440286	
F	-1.00537329	-1.13753731	0.697079	
C	0.12267266	1.53487966	-0.116946	
H	1.77943709	0.25055740	-0.500550	
H	1.22150351	0.20787566	1.175847	
O	0.65828128	2.51006545	-0.625646	
O	-1.17539930	1.59318015	0.322564	
H	-1.51597836	0.75117855	0.692885	

Table D.26: Optimized geometry of $\text{C}_4\text{F}_5\text{H}_3\text{O}_2$ in Å.

Energy (a.u) $\text{C}_4\text{F}_5\text{H}_3\text{O}_2$: -803.668961			
C	-0.72874996	0.32773422	-0.075512
C	-0.60029676	-0.94095734	0.752863
F	-1.91975143	0.98711300	0.274008
F	-0.79424813	0.02987934	-1.447400
C	0.37990563	1.37564566	0.098118
C	0.08373242	-2.15083910	0.099111
H	-1.61603271	-1.26681598	0.984614
H	-0.10666918	-0.70757072	1.703719
F	0.48092182	1.75069707	1.421546
F	0.17734377	2.48728550	-0.670324
F	1.61673551	0.82124319	-0.268494
O	-0.42786288	-3.26190016	0.147510
O	1.31544684	-1.97855591	-0.472285
H	1.64176098	-1.05610270	-0.504369

Table D.27: Optimized geometry of $\text{C}_5\text{F}_7\text{H}_3\text{O}_2$ in Å.

Energy (a.u) $\text{C}_5\text{F}_7\text{H}_3\text{O}_2$: -1041.395285			
C	-0.84860910	-0.34419311	0.293154
C	-0.64099103	-1.74929211	0.817225
F	-1.80223790	0.30947526	1.084167
F	-1.38807409	-0.39345468	-1.025950
C	0.40684427	0.54307329	0.239363
C	0.21634482	-2.70240208	-0.027178
H	-1.64262798	-2.18450282	0.926717
H	-0.18931067	-1.69427543	1.808230
F	0.96919475	0.55077286	1.517724
C	0.23783076	2.00797402	-0.199827
F	1.31048388	-0.06295635	-0.638330
F	-0.55146519	2.69510078	0.694693
F	-0.34221287	2.07498432	-1.448306
F	1.47137414	2.61646468	-0.254752
O	1.04794355	-3.43938898	0.484382
O	-0.05740496	-2.77140838	-1.368461
H	-0.68703453	-2.08976978	-1.684468

Table D.28: Optimized geometry of $\text{C}_6\text{F}_9\text{H}_3\text{O}_2$ in Å.

Energy (a.u) $\text{C}_6\text{F}_9\text{H}_3\text{O}_2$: -1279.116693			
C	-0.84889070	-1.15990300	-0.117282
C	-0.55127420	-2.49034160	0.542836
F	-2.19956574	-0.83883037	0.072012
F	-0.64908818	-1.26946284	-1.523024
C	-0.00754371	0.03520938	0.377826
C	0.83580945	-3.10234497	0.299971
H	-1.31201174	-3.18978322	0.173034
H	-0.68057172	-2.39103986	1.620484
F	-0.01839932	0.00732125	1.773833
C	-0.45653595	1.44458526	-0.069151
F	1.30793437	-0.17804477	-0.051966
F	-1.57406889	1.80228911	0.677315
F	-0.80996189	1.40642062	-1.416578
C	0.59684144	2.56029212	0.099307
F	1.12412865	2.53143381	1.372952
F	0.00417367	3.78483008	-0.106393
F	1.61631499	2.40905270	-0.811973
O	1.48279994	-3.60602623	1.207848
O	1.28102368	-3.15198024	-0.995489
H	0.72318785	-2.65445830	-1.629725

Table D.29: Optimized geometry of $C_7F_{11}H_3O$ in Å.

Energy (a.u) $C_7F_{11}H_3O_2$: -1516.838354			
C	-0.85028538	-1.82999908	-0.197482
C	-0.59960122	-3.17177207	0.459722
F	-2.20633986	-1.49888676	-0.075388
F	-0.57662776	-1.92414709	-1.591633
C	-0.02547929	-0.64862937	0.358870
C	0.79333967	-3.79329120	0.283480
H	-1.34531624	-3.85905298	0.039899
H	-0.78592729	-3.08544950	1.530086
F	-0.11319466	-0.69709659	1.750798
C	-0.44667788	0.77140103	-0.096002
F	1.30728619	-0.86649462	-0.007679
F	-1.58806076	1.13170097	0.618970
F	-0.76487587	0.73796672	-1.451414
C	0.62674365	1.86754300	0.112881
F	1.23102325	1.67690230	1.354044
C	0.10333620	3.31927491	0.070905
F	1.58721958	1.73508291	-0.883073
F	-0.64155655	3.60338246	1.191552
F	-0.67759155	3.51280565	-1.049337
F	1.16888133	4.18816674	0.022164
O	1.38525060	-4.31617202	1.217711
O	1.30806556	-3.82889802	-0.986379
H	0.78984515	-3.31760910	-1.642834

Table D.30: Optimized geometry of $\text{C}_8\text{F}_{13}\text{H}_3\text{O}_2$ in Å.

Energy (a.u) $\text{C}_8\text{F}_{13}\text{H}_3\text{O}_2$: -1754.560016			
C	-0.85047557	-2.58828266	-0.125185
C	-0.40864247	-3.89652338	0.498077
F	-2.21426372	-2.39019019	0.128511
F	-0.70443209	-2.66117333	-1.539602
C	-0.09445961	-1.32980337	0.355586
C	1.01440396	-4.38125821	0.185548
H	-1.12113266	-4.65474591	0.148463
H	-0.49893244	-3.82436798	1.581796
F	-0.03821321	-1.38386427	1.748817
C	-0.69771729	0.04137353	-0.045414
F	1.21004394	-1.41632380	-0.143397
F	-1.79297732	0.28664076	0.780015
F	-1.13560620	-0.02863183	-1.364848
C	0.28376347	1.23856778	0.071614
F	1.03334120	1.09010010	1.235979
C	-0.39303524	2.63065677	0.114572
F	1.14476998	1.20138559	-1.023477
F	-0.91951118	2.82467078	1.385783
F	-1.43121699	2.65702741	-0.815173
C	0.53661065	3.82360884	-0.195777
F	1.68659192	3.74109627	0.560516
F	-0.11085674	4.99843321	0.110240
F	0.87566212	3.84592280	-1.528415
O	1.73962345	-4.84397412	1.055320
O	1.40750962	-4.36834987	-1.127476
H	0.78198093	-3.91141246	-1.728173

Table D.31: Optimized geometry of $\text{C}_9\text{F}_{15}\text{H}_3\text{O}_2$ in Å.

Energy (a.u) $\text{C}_9\text{F}_{15}\text{H}_3\text{O}_2$: -1992.281669			
C	-0.73436135	-3.21299826	-0.308445
C	-0.45435667	-4.51733995	0.409544
F	-2.11930614	-3.01703835	-0.392501
F	-0.25080030	-3.29213122	-1.645239
C	-0.11859255	-1.95048144	0.334648
C	1.00305754	-4.99872386	0.452863
H	-1.05970454	-5.27899638	-0.098535
H	-0.80403776	-4.44065544	1.438941
F	-0.40261279	-1.99732262	1.700015
C	-0.60747086	-0.58239894	-0.208736
F	1.26807055	-2.03716352	0.167882
F	-1.86900102	-0.33381964	0.327900
F	-0.71524785	-0.66180484	-1.594089
C	0.31877191	0.61730980	0.133880
F	0.76075331	0.47572383	1.446222
C	-0.35472648	2.00960488	-0.000271
F	1.41617376	0.57371401	-0.722473
F	-1.16596834	2.20995839	1.114466
F	-1.15316289	2.01191687	-1.141797
C	0.63173573	3.19850019	-0.106003
F	1.67250782	3.00267137	0.799924
C	0.02093584	4.58686568	0.182982
F	1.15173349	3.22564757	-1.394273
F	-0.26198362	4.72835404	1.521238
F	-1.13925480	4.75733533	-0.543057
F	0.91761583	5.56485844	-0.180400
O	1.49900292	-5.45119632	1.475518
O	1.70013454	-4.99479523	-0.727333
H	1.23689615	-4.54408733	-1.464353

Alkanoic acids

Table D.32: Optimized geometry of $\text{C}_2\text{H}_4\text{O}_2$ in Å

Energy (a.u) $\text{C}_2\text{H}_4\text{O}_2$: -228.994383			
C	-1.20514248	-0.74159178	-0.052916
C	-0.04243103	0.22511560	0.002617
H	-2.06302767	-0.24441591	-0.504836
H	-1.47995174	-1.07697372	0.954553
H	-0.95211894	-1.62682680	-0.648580
O	1.14179268	-0.28998979	0.518491
H	1.07231902	-1.23009643	0.780969
O	-0.07874733	1.39342884	-0.361237

Table D.33: Optimized geometry of $\text{C}_3\text{H}_6\text{O}_2$ in Å

Energy (a.u) $\text{C}_3\text{H}_6\text{O}_2$: -268.300026			
C	-1.92454839	-0.02660567	-0.406620
C	-0.53153137	0.57871072	-0.603189
H	-1.96324501	-1.04803807	-0.796177
H	-2.68003926	0.57307385	-0.923789
H	-2.18035513	-0.07170193	0.655754
C	0.55027624	-0.17565094	0.152660
H	-0.51577193	1.62798031	-0.271751
H	-0.26271072	0.59034153	-1.670246
O	0.36494156	-1.15606458	0.863232
O	1.84073605	0.31878791	0.001762
H	1.88568872	1.10382912	-0.581618

Table D.34: Optimized geometry of $\text{C}_4\text{H}_8\text{O}_2$ in Å

Energy (a.u) $\text{C}_4\text{H}_8\text{O}_2$: -307.615109			
C	-2.66800919	-0.65731751	0.110805
C	-1.41454835	0.23025285	0.123773
H	-2.63993055	-1.39907206	0.919006
H	-2.76125385	-1.20200609	-0.837255
H	-3.57502509	-0.05698477	0.240757
C	-0.12657691	-0.58432021	-0.056735
H	-1.35763889	0.78976154	1.065615
H	-1.47766535	0.98466594	-0.670084
C	1.11862971	0.26147275	-0.048152
H	-0.14490903	-1.14685799	-1.000731
H	-0.02484715	-1.34142866	0.733405
O	2.25471698	-0.51036186	-0.212046
O	1.17906989	1.48677302	0.083970
H	3.05369634	0.06079351	-0.203567

Table D.35: Optimized geometry of $C_5H_{10}O_2$ in Å.

Energy (a.u) $C_5H_{10}O_2$: -346.906738			
C	-3.33648383	-0.32266396	-0.155345
C	-2.09592057	0.53697296	0.132826
H	-3.36802932	-0.63020275	-1.207728
H	-4.25838427	0.22885264	0.059536
H	-3.34029299	-1.23200354	0.458623
C	-0.77689368	-0.20026266	-0.162603
H	-2.10419298	0.85793908	1.184828
H	-2.13473226	1.45511211	-0.470219
C	0.46438427	0.67296619	0.134925
H	-0.75808371	-0.51160100	-1.215755
H	-0.74906020	-1.12592199	0.433812
C	1.78584276	0.01984154	-0.227790
H	0.48059456	0.93698737	1.203139
H	0.41077323	1.60784057	-0.428872
O	2.62568482	0.50402633	-0.977606
O	2.03398911	-1.21823310	0.354760
H	1.30818076	-1.52275047	0.935888

Table D.36: Optimized geometry of $\text{C}_6\text{H}_{12}\text{O}_2$ in Å.

Energy (a.u) $\text{C}_6\text{H}_{12}\text{O}_2$: -386.212432			
C	-4.00303909	-0.79269640	-0.010894
C	-2.77904811	0.13652041	-0.019101
H	-4.01265446	-1.42539881	0.885614
H	-4.00176233	-1.45527425	-0.885611
H	-4.93781883	-0.22092580	-0.026331
C	-1.44237249	-0.62771272	0.002878
H	-2.82247830	0.81086831	0.848414
H	-2.81107268	0.78011719	-0.910164
C	-0.21852557	0.30409488	-0.006057
H	-1.39881281	-1.30315209	-0.865832
H	-1.40986412	-1.27185351	0.895507
C	1.11037716	-0.46358864	0.014399
H	-0.25394220	0.97892084	0.858553
H	-0.24435329	0.94927363	-0.893367
C	2.32692510	0.44601211	0.005107
H	1.17344533	-1.13803661	-0.853792
H	1.16415647	-1.10674383	0.906661
O	2.29771957	1.67081073	-0.010666
O	3.55757877	-0.20153518	0.016432
H	3.47709126	-1.17721152	0.027882

Table D.37: Optimized geometry of $C_7H_{14}O_2$ in Å.

Energy (a.u) $C_7H_{14}O_2$: -425.514935			
C	-4.69063917	-0.43740267	-0.035839
C	-3.44429918	0.44818562	0.120216
H	-4.69167330	-1.25096879	0.700409
H	-4.72929301	-0.89170136	-1.033789
H	-5.61096004	0.14093169	0.103923
C	-2.12745429	-0.32600104	-0.074047
H	-3.44612646	0.91331336	1.117043
H	-3.48723381	1.27440660	-0.604470
C	-0.87538909	0.55449077	0.087540
H	-2.12217575	-0.78690523	-1.073464
H	-2.08610893	-1.15602346	0.648172
C	0.43937765	-0.21902401	-0.117621
H	-0.87735464	1.01177960	1.088808
H	-0.91889220	1.38574164	-0.631582
C	1.68473444	0.67931925	0.065501
H	0.45408015	-0.65905779	-1.123788
H	0.46622148	-1.06314017	0.589514
C	3.00309211	-0.01493815	-0.225585
H	1.70931228	1.06834636	1.094575
H	1.62964415	1.54098441	-0.604725
O	3.84624818	0.38498665	-1.020074
O	3.24521386	-1.18642177	0.483326
H	2.51686892	-1.42639165	1.090899

Table D.38: Optimized geometry of $\text{C}_8\text{H}_{16}\text{O}_2$ in Å.

Energy (a.u) $\text{C}_8\text{H}_{16}\text{O}_2$: -464.830081			
C	-5.34311333	-0.77028532	0.011380
C	-4.09541294	0.11551284	0.156500
H	-5.34830154	-1.57017617	0.762556
H	-5.37763662	-1.24357243	-0.978035
H	-6.26372082	-0.18855394	0.136080
C	-2.77849433	-0.66403757	-0.013517
H	-4.10239798	0.60131032	1.143433
H	-4.13262124	0.92706298	-0.585176
C	-1.52386078	0.21603864	0.131268
H	-2.77047388	-1.14952623	-1.001654
H	-2.74074778	-1.47665318	0.728338
C	-0.20841073	-0.56440900	-0.042951
H	-1.53122469	0.69866032	1.120526
H	-1.56345347	1.02988487	-0.608732
C	1.04092435	0.32276395	0.096187
H	-0.20152006	-1.04909813	-1.031299
H	-0.16626961	-1.37576792	0.699470
C	2.35065837	-0.48239588	-0.077353
H	1.03996752	0.80459279	1.083966
H	1.01497495	1.12633604	-0.649626
C	3.58715593	0.36105807	0.064851
H	2.35898933	-0.94124867	-1.074854
H	2.41461924	-1.28175020	0.666176
O	3.63679207	1.36293319	-0.890303
H	4.44835328	1.90260515	-0.768822
O	4.47685230	0.22845799	0.909292