

---

MSU Graduate Theses

---

Summer 2018

## NMR Spectroscopic Properties of Nucleotides, and a New Method of Numeric Calculation of Raman Intensities for Organic Molecules

William R. Ehrhardt

Missouri State University, William123@live.missouristate.edu

As with any intellectual project, the content and views expressed in this thesis may be considered objectionable by some readers. However, this student-scholar's work has been judged to have academic value by the student's thesis committee members trained in the discipline. The content and views expressed in this thesis are those of the student-scholar and are not endorsed by Missouri State University, its Graduate College, or its employees.

---

Follow this and additional works at: <https://bearworks.missouristate.edu/theses>

 Part of the [Organic Chemistry Commons](#), and the [Physical Chemistry Commons](#)

### Recommended Citation

Ehrhardt, William R., "NMR Spectroscopic Properties of Nucleotides, and a New Method of Numeric Calculation of Raman Intensities for Organic Molecules" (2018). *MSU Graduate Theses*. 3299.  
<https://bearworks.missouristate.edu/theses/3299>

This article or document was made available through BearWorks, the institutional repository of Missouri State University. The work contained in it may be protected by copyright and require permission of the copyright holder for reuse or redistribution.

For more information, please contact [BearWorks@library.missouristate.edu](mailto: BearWorks@library.missouristate.edu).

**NMR SPECTROSCOPIC PROPERTIES OF NUCLEOTIDES, AND A NEW  
METHOD OF NUMERIC CALCULATION OF RAMAN INTENSITIES FOR  
ORGANIC MOLECULES**

A Master's Thesis

Presented to

The Graduate College of

Missouri State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science, Chemistry

By

William Ryan Ehrhardt

August 2018

Copyright 2018 by William Ryan Ehrhardt

**NMR SPECTROSCOPIC PROPERTIES OF NUCLEOTIDES, AND A NEW  
METHOD OF NUMERIC CALCULATION OF RAMAN INTENSITIES FOR  
ORGANIC MOLECULES**

Chemistry

Missouri State University, August 2018

Master of Science

William Ryan Ehrhardt

**ABSTRACT**

General and accurate computational methodologies are currently lacking for large chemical systems. This is primarily due to the computational expense required to perform calculations on systems with one hundred or more atoms. Calculated spectroscopic properties could aid in the process of elucidating structural features of large biologically relevant molecules if accurate and inexpensive methods are developed. Towards this end the first steps were taken to design a general methodology for predicting NMR chemical shifts of large nucleic acid systems. It was found that HF and semi-empirical methods were not sufficient for optimization of nucleobases, and therefore larger nucleotide or nucleic acid systems. It was also found that there is little difference in performance between DFT methods for prediction of NMR shifts of nucleobases as long as hydrogen bonding requirements are satisfied. To potentially reduce the computational expense of calculating Raman activities, a new and potentially inexpensive numerical method was developed. This method utilizes volume changes as a basis for approximating polarizability changes over the course of molecular vibrations. Raman intensities calculated using this methodology were compared to experimentally obtained Raman intensities by linear regression. While a positive correlation was found further refinement is needed.

**KEYWORDS:** DFT, DNA, Raman, NMR, electronic volume

This abstract is approved as to form and content

---

Matthew R. Siebert, PhD, Committee Chairperson  
Advisory Committee  
Missouri State University

**NMR SPECTROSCOPIC PROPERTIES OF NUCLEOTIDES, AND A NEW  
METHOD OF NUMERIC CALCULATION OF RAMAN INTENSITIES FOR  
ORGANIC MOLECULES**

By

William Ryan Ehrhardt

A Master's Thesis  
Submitted to the Graduate College  
Of Missouri State University  
In Partial Fulfillment of the Requirements  
For the Degree of Master of Science, Chemistry

August 2018

Approved:

---

Matthew R. Siebert, PhD, Committee Chairperson

---

Amanda C. Brodeur, MD PhD, Committee Member

---

Gary A. Meints, PhD, Committee Member

---

G. Alan Schick, PhD, Committee Member

---

Julie J. Masterson, PhD, Dean Graduate College

In the interest of academic freedom and the principle of free speech, approval of this thesis indicates the format is acceptable and meets the academic criteria for the discipline as determined by the faculty that constitute the thesis committee. The content and views expressed in this thesis are those of the student-scholar and are not endorsed by Missouri State University, its Graduate College, or its employees.

## TABLE OF CONTENTS

Chapter 1: Relevance of Research .....	1
Introduction.....	1
Applications of Spectroscopy in Modern Chemistry.....	1
Applications of Computational NMR in Modern Chemistry .....	4
Applications of Vibrational Spectroscopy and Predicted Vibrational Spectra.....	7
Volume Changes as a Replacement for Analytic Raman Intensities.....	9
Chapter 2: Theoretical Basis.....	11
Introduction.....	11
Computational Chemistry .....	11
DNA Structure and Function .....	15
NMR Spectroscopy.....	22
Molecular Vibrations and Polarizability.....	27
Calculating Spectroscopic Intensity.....	31
Chapter 3: Methods.....	38
Programs Utilized .....	38
Model Chemistries .....	38
Calculating an NMR Spectrum.....	43
Calculation and Interpretation of Vibrational Spectra .....	47
Calculating Electronic Volume Changes .....	53
Chapter 4: Results and Discussion.....	55
NMR of Nucleobases .....	55
Volume changes as a Replacement for Analytic Polarizabilities .....	67
Chapter 5: Conclusions and Future Work.....	77
NMR of Nucleobases.....	77
Volume Changes as Polarizability Changes .....	77
References.....	79
Appendices.....	86
Appendix A: Experimental Raman Spectra for Test Set Molecules.....	86
Appendix B: Code to Generate Vibrationally Displaced Structures .....	90

## LIST OF TABLES

Table 1. <sup>13</sup> C chemical shifts (ppm) obtained from chemical literature .....	56
Table 2. Adenine <sup>1</sup> H chemical shifts (ppm) obtained from chemical literature.....	57
Table 3. Cytosine <sup>1</sup> H chemical shifts (ppm) obtained from chemical literature.....	57
Table 4. Thymine <sup>1</sup> H chemical shifts (ppm) obtained from chemical literature.....	57
Table 5. Uracil <sup>1</sup> H chemical shifts (ppm) obtained from chemical literature .....	58
Table 6. Model chemistries excluded as candidates for optimization of larger nucleic acid systems .....	60
Table 7. Linear regression data for nucleobase <sup>1</sup> H chemical shifts predicted using B3LYP and B3P86 functionals.....	63
Table 8. Linear regression data for nucleobase non-exchangeable <sup>1</sup> H chemical shifts predicted using B3LYP and B3P86 functionals and implicit solvation .....	64
Table 9. Top and bottom performing model chemistries using implicit solvation.....	65
Table 10. Error for non-exchangeable <sup>1</sup> H and <sup>13</sup> C nuclei using the five best model chemistries utilizing implicit solvation.....	67
Table 11. Top performing model chemistries using minimum explicit solvation as judged by <sup>1</sup> H chemical shift linear regression.....	68
Table 12. Error for all <sup>1</sup> H and <sup>13</sup> C nuclei using the five best model chemistries and explicit solvation, as judged by the proton R <sup>2</sup> value.....	69
Table 13. Description of vibrations used for experimental comparison.....	71
Table 14. Linear regression data for Figure 25.....	75

## LIST OF FIGURES

Figure 1. 1-dimensional <sup>1</sup> H NMR spectrum of Drew-Dickerson dodecamer .....	3
Figure 2. Originally proposed structure and revised structure of aquatolide .....	5
Figure 3. Originally proposed structure and revised structure of nobilisitine A .....	7
Figure 4. Numbering scheme for purines, pyrimidines, and 2'-deoxyribofuranose .....	16
Figure 5. Structures of the nucleotides in DNA and uracil.....	17
Figure 6. Primary structure of AGCT DNA sequence and hydrogen bonding between bases.....	19
Figure 7. Backbone structure of DNA .....	20
Figure 8. Modified nucleotides with mutagenic changes .....	21
Figure 9. Energy difference as a function of external magnetic field strength .....	24
Figure 10. Stretching, bending, and torsions demonstrated through vibrations of the ethene molecule .....	28
Figure 11. Complex vibration of thymine molecule.....	30
Figure 12. Energy level diagram for infrared absorption, Raman scattering, and Rayleigh scattering.....	31
Figure 13. Electronic surfaces at 0.1, 0.2, 0.4, and $0.8 \frac{e^-}{Bohr^3}$ .....	35
Figure 14. Structure of TMS, Input and Output File Text for NMR Calculations .....	45
Figure 15. Example plot of linear regression method of calculating chemical shifts.....	46
Figure 16. Output from Gaussian following a frequency calculation.....	49
Figure 17. Sample output from VEDA program using both mixed and unmixed modes .....	51
Figure 18. Electronic density surface rendered with a low number of triangles for easy visualization .....	54
Figure 19. Optimized structure of guanine with four DMSO molecules.....	59
Figure 20. Optimized structure of uracil with two DMSO molecules.....	64

Figure 21. Optimized structure of thymine with two DMSO molecules.....	65
Figure 22. Optimized structure of cytosine with three DMSO molecules.....	66
Figure 23. Optimized structure of adenine with three DMSO molecules.....	66
Figure 24. Seven molecule test set to calculate volume changes of vibrational modes ...	70
Figure 25. Linear regression for the strongest correlation found using the $\frac{\Delta V_{a,b}}{\Delta ALIE_{a,b}}$ function .....	74
Figure 26. Linear regression for the strongest correlation found using the $\Delta V_{a,b}$ function .....	76
Figure 27. Linear regression for the strongest correlation found using the $\Delta \left( \frac{V}{ALIE} \right)_{a,b}$ function .....	76

## CHAPTER 1: RELEVANCE OF RESEARCH

### Introduction

Quantum chemical calculations on very large systems (100+ atoms) remains a difficult task. This is due to both the computational expense and difficulty in designing accurate methodologies for these large systems. General and accurate methodologies for these large systems, if developed, would allow systematic investigation of structural and spectroscopic features for these systems.

The first project contained herein begins exploring the optimization requirements for nucleotide systems as they pertain to NMR spectra prediction, then begins the process of designing an accurate methodology for predicting NMR chemical shifts of large nucleic acid systems. In future works the recommendations contained herein will be applied to larger nucleic acid systems, and then applied to unknown nucleic acid systems to determine structural properties.

The second project is aimed towards reducing the computational expense in accurate calculation of Raman spectra for organic molecules. Currently, calculated vibrational spectra are difficult to interpret and expensive to produce. A solution that allows for faster calculation would be useful towards determining spectroscopic properties of large molecules.

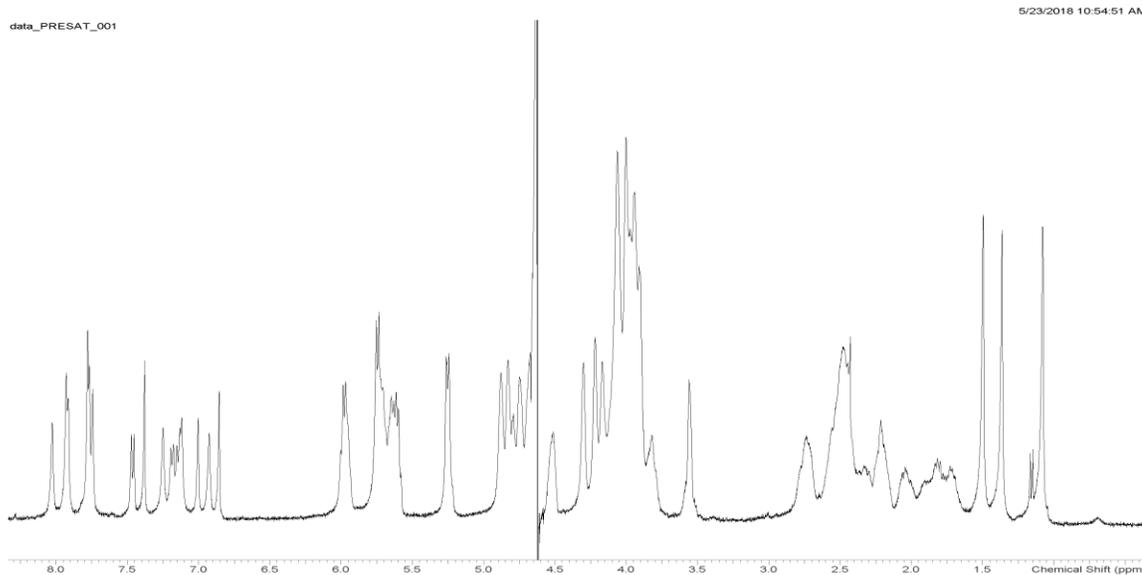
### Applications of Spectroscopy in Modern Chemistry

Spectroscopic methods are of great use to modern synthetic and analytic chemists. When a new molecule is synthesized or isolated, spectroscopic evidence for the structure

is considered the minimum threshold to publish a proposed structure. Additionally, when performing a novel synthesis or isolation with a known compound as the target, a comparison between spectra can be conducted as confirmation. Changes to measured spectra can yield information about a molecule's interaction with a substrate, or after dimerization.<sup>1,2</sup> Two of the most commonly used types of spectroscopy are nuclear magnetic resonance (NMR) spectroscopy and vibrational spectroscopy.

NMR spectroscopy provides useful data for molecules ranging from a few tens to several tens of thousands amu. For small molecules basic information about functional groups and their connectivity can be obtained from 1-dimensional (1-D) NMR. For larger molecules such as proteins or DNA 3-dimensional features can be discerned from 2-dimensional (2-D) spectra.

While NMR is useful, there are several complicating factors to interpretation. Figure 1 demonstrates the limited applicability of 1-D NMR spectroscopy to large systems as individual peaks arising from similar functional groups quickly become unresolvable.<sup>3</sup> Similarly, when several stereo centers exist the expected difference in the experimental spectra between the possible structures may be too subtle to differentiate on a 1-D spectrum. Another source of error in structural assignment is bias on the part of a synthetic chemist. Chemists typically have a product they are aiming to synthesize, and so features that support their conclusion are naturally highlighted and inconsistent results minimized. Natural products pose the opposite problem, in that the structure is completely unknown beforehand and must be proposed from the ground-up. Ideally, several spectroscopic methods should be used to characterize a sample before an accurate assignment can be made.



**Figure 1:** 1-dimensional <sup>1</sup>H NMR spectrum of Drew-Dickerson dodecamer.

Vibrational spectra are of use to a great many chemical endeavors. Infrared spectra are routinely recorded during organic synthesis to provide evidence for the presence of the expected products. Raman spectra may be obtained and provide complementary information to infrared spectra. It is known that certain characteristic peaks in vibrational spectra are indicative of specific functional groups<sup>4</sup>, and therefore can be used to assist in the identification of an unknown substance, a newly synthesized compound, or a compound isolated from a natural source.

Raman spectroscopy has recently found an interesting home in the field of medical diagnostics. By applying this analytical technique to biologically derived fluids, cells, and tissues biochemical markers can be measured and disease states can be analyzed. In particular Raman spectroscopy has been applied to the diagnosis and staging of cancer, detection of diseases that result in significant changes to connective tissues,

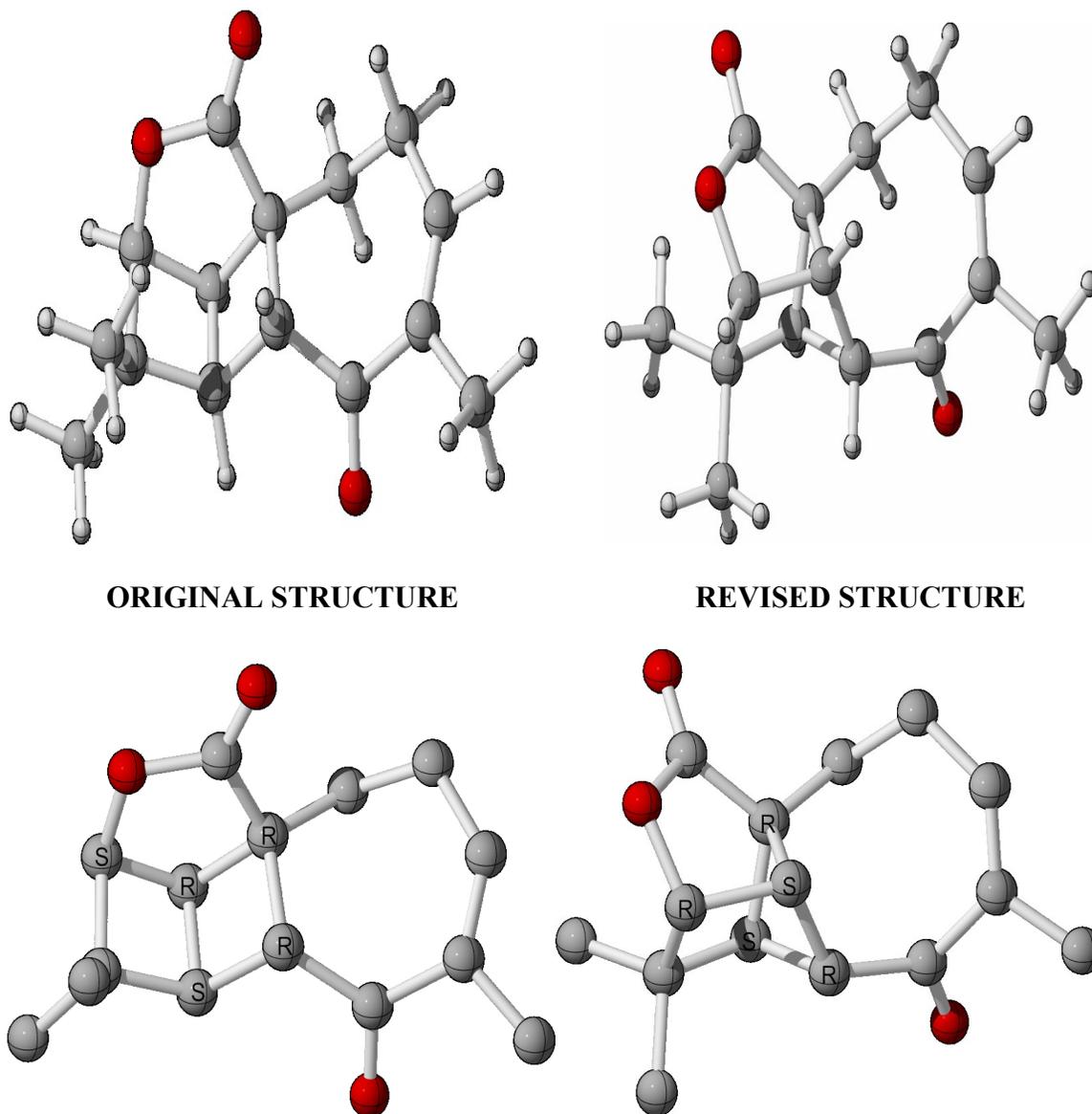
and to quantitatively assess the contents of biofluids.<sup>5</sup> Additional applications of NMR to biological systems are currently an active area of research.

### **Applications of Computational NMR in Modern Chemistry**

Computed NMR spectra can be of assistance in elucidating complex chemical structures. While NMR spectroscopy is a highly versatile methodology, interpretation becomes difficult when complex structures or mixtures are being analyzed. Comparing a predicted NMR spectrum with an experimentally obtained NMR spectrum can help to highlight inconsistencies between a proposed structure based on experimental NMR spectra. As computational power increases and chemical models become increasingly accurate computational methodology will become useful in validating larger molecules and higher order NMR spectra.

There have been instances in which a computed NMR spectrum has led to a revised structural assignment. The chemical aquatolide was originally isolated from *Asteriscus aquaticus* in 1989 and a structural assignment was made on the basis of 1- and 2-dimensional NMR spectroscopy.<sup>6</sup> In 2012 a paper was published revising the structure of this natural product. This need for a revision was initially made on the basis of the inconsistencies between computed and experimental NMR spectra for the originally agreed upon structure. After more 2-dimensional NMR spectra were obtained from samples of the isolated compound, and several computed NMR spectra were compared, a revised structure was proposed. This revised structure was ultimately confirmed by crystallographic methods and can be seen alongside the original structure in Figure 2.<sup>7</sup> The original structure, although incorrect, was in agreement with all collected

experimental data at that time. This example demonstrates the difficulty of determining a chemical structure on the basis of experimentally obtained NMR spectra.



**Figure 2:** Originally proposed structure (left) and revised structure (right) of aquatolide with hydrogens present (top) and with hydrogens removed and stereo centers labeled (bottom). Grey = C, Red = O, Small = H

A second example of the utility of computed NMR can be seen in the structural revision of nobilistine A, the originally proposed and revised structure of which can be

seen in Figure 3. The originally proposed structure had a total of five stereo centers making unambiguous structural assignment difficult.<sup>8</sup> The enantiomer of the proposed structure was synthesized several years later and was confirmed using x-ray crystallography. When the NMR data was compared between the suspected enantiomers, inconsistencies were found, prompting a structural revision of the naturally isolated compound.<sup>9</sup> NMR calculations were performed for sixteen possible stereoisomers to begin the search for an accurate structural revision. Ultimately, a revised structure was proposed on the basis of these calculations. The newly proposed molecule was synthesized and the structure confirmed by crystallographic methods. New measurements confirmed that it was the same structure as was originally isolated. This example shows the utility of NMR calculations for distinguishing between closely related structures, and how it can be used to reduce synthetic work when troubleshooting structures.

While there have been several successes in utilizing computed 1-D NMR spectra, it is more difficult to accurately predict higher order spectra. This is due to both the increased difficulty of interpreting higher order spectra, and that a greater number of features must be accurately predicted to reproduce experimental spectra. If higher order NMR spectra can be accurately predicted, then experimental chemists will likely benefit from its use in determining higher order structural features of large molecules.<sup>11</sup>

Nucleic acids are large molecules to which computational NMR predictions could be of great value. First, computational methodologies that accurately reproduce the 1-D spectra of nucleotides and nucleosides should be identified. Second, these methodologies should be applied to larger DNA molecules, and higher-dimensional spectra should be

reproduced. Finally, methods that have been validated to reproduce experimental results for unmodified structures can be applied to structures that have incorporated mutagenic chemical changes. By following this scheme, structural changes that result from chemical modification of DNA can be better understood and proposed structural changes which result from modification can be validated computationally.

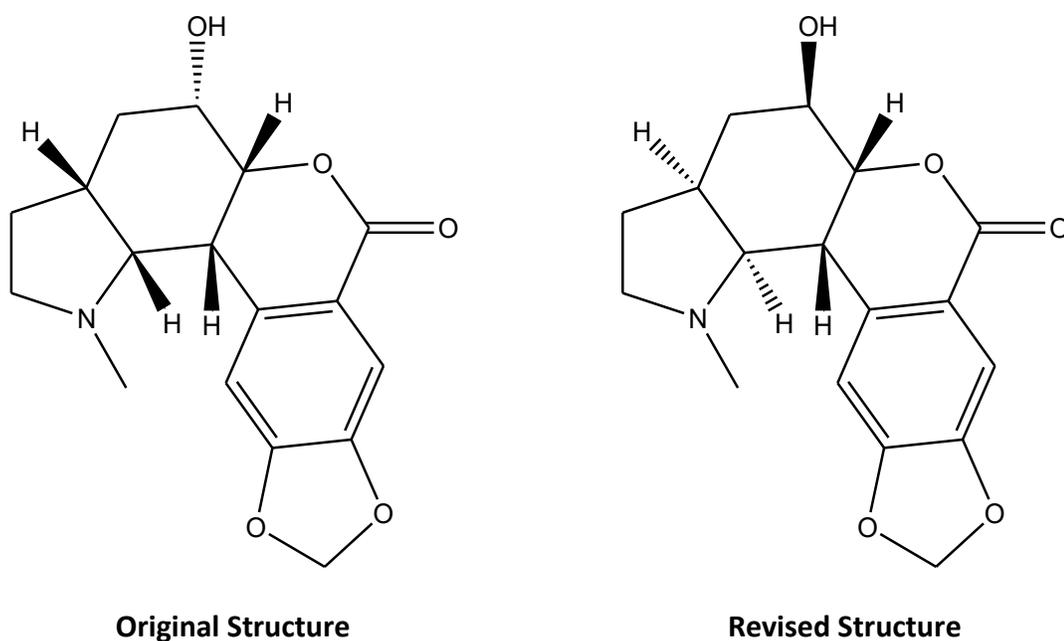


Figure 3: Originally proposed structure (left) and revised structure (right) of nobilisitine A.

### **Applications of Vibrational Spectroscopy and Predicted Vibrational Spectra**

The primary use in calculating a Raman spectrum is to assist in assigning spectral features to specific molecular vibrations. This seemingly straightforward task can actually be quite daunting for even small or moderate sized molecules, often requiring complex synthesis and measurement schemes. A common piece of experimental

evidence for making spectral assignments is isotopic substitution. During this process a chemical is synthesized with isotopes present at a specific location in the molecule. This causes frequencies of peaks in vibrational spectra to shift from a few tens to several hundred wavenumbers depending on the nature of the peak and substitution. Isotopic enrichment can be a difficult process and often require a great deal of effort to ensure enrichment only happens at specific locations in the molecule.

The process of identifying which peaks are due to which vibrations can be aided by the use of computed vibrational spectra. First, a computational methodology that accurately replicates known vibrations for a particular molecule, or set of closely related molecules, should be identified. This can be done by comparison of the calculated vibrational frequencies or intensities with experimental vibrational spectra. If the results for well-defined vibrations are in good agreement with experimental data then it is likely that peaks which are not well defined by experiment are also accurately described.

The process of calculating vibrational features of a molecule is called a frequency calculation. The output of this type of calculation is the normal vibrational modes (normal modes) of a molecule, and information about transitions from the ground vibrational state to the next lowest energy vibrational state. Specifically normal modes are reported as the relative atomic motions during the vibration, and information about the transitions are the energy separation between vibrational energy levels and the expected spectroscopic activities. When these values agree well with experimental data it is an indication that the molecule is accurately described by the computational method.

Computational Raman could also find a home in medical applications by using its predictive power to discern targets for analysis. First it can be used to understand the

origins of the complex spectroscopic features that originate from the sea of biomolecules present in cells and tissue samples. Then potential changes to the vibrational spectra that would result from a disease state can be predicted by modifying the molecule or environment appropriately and repeating the calculation.

While frequency calculations would be useful in elucidating vibrational features of large molecules the computational time required to carry out such a calculation scales poorly with respect to the number of atoms present. Additionally, to accurately predict spectroscopic intensities requires large basis sets that further worsen the scaling of the calculation.

### **Volume Changes as a Replacement for Analytic Raman Intensities**

Raman calculations are being employed on increasingly complex systems, as such it is important to develop computational methods capable of handling this increasing complexity. Due to how poorly frequency calculations scale with molecular size, Raman calculations are limited in usefulness to systems of only a few hundred amu. A method of calculation that extends the applicability of Raman calculations to larger systems would make spectral analysis of these complex mixtures a much less daunting task.

It is known that accurate prediction of Raman intensities requires large basis sets that scale poorly with respect to the number of atoms present. Large basis sets are required because it is molecular polarizability which determines Raman activities, and the most polarizable regions require large basis sets to be adequately described.<sup>12</sup> Knowing this, it would be advantageous to first calculate vibrational frequencies and normal modes

using an inexpensive method, then use an alternative method to calculate spectroscopic intensities.

Electronic volume calculations are a potentially attractive replacement for analytic Raman intensity calculations. Previous work has established that a reliable correlation exists between electronic volume and bulk polarizability.<sup>13</sup> Additionally, it has been shown that bulk polarizabilities correlate well with the ratio of molecular volume, as defined by the volume enclosed by a specific electron density surface, and the average ionization energy enclosed within that volume at calculated equilibrium geometries.<sup>13</sup> It seems reasonable that this relationship, having been established for equilibrium geometries, should be applicable to displaced geometries and polarizability changes.

The proposed procedure requires creation of a structure representative of each vibration of interest followed by calculation of electronic density (performed as a calculation of energy on that structure), then an electronic volume calculation. In effect, this procedure allows frequency calculation utilizing an expensive basis set to be broken into a less expensive frequency calculation (from which frequencies and normal modes can be obtained) followed by several smaller volume calculations from which spectroscopic intensities can be derived. Even if this procedure does not allow for immediate savings in computational resources, the ability to run the jobs in parallel may allow for more efficient use of those computational resources.

## CHAPTER 2: THEORETICAL BASIS

### Introduction

Having demonstrated the utility of computational methods it now becomes important to understand the principles underlying their use. Computational methodology requires knowledge of the experimental procedures and theory, as well as knowledge of the underlying computational concepts and methodologies. In this way it can be daunting to utilize computational methodologies. Luckily, the finer details have been implemented by computational packages, such as the Gaussian<sup>14</sup> products, so we can stick to the broad strokes.

### Computational Chemistry

The theory underlying computational methods can be found in texts such as those authored by McQuarrie<sup>15</sup> and Cramer<sup>16</sup>. The former covers the basic of quantum chemistry, the latter the implementation and theory of computational methods.

The foundation of quantum mechanics is the time-independent Schrödinger equation:

$$\hat{H}\Psi = E\Psi \quad (1)$$

In which the Hamiltonian operator,  $\hat{H}$ , operates on a wave function,  $\Psi$ , and returns the product of the wave function and energy,  $E$ . The complete Hamiltonian operator for a molecule can be written as the sum of kinetic energies for each particle and electrostatic potentials for each pair of particles. It is generally useful to assume the nuclei are

stationary (the Born-Oppenheimer approximation) so that only electrons need be considered. The electronic Hamiltonian for a molecule is then:

$$\hat{H} = \sum_i \nabla_i^2 - \sum_i \sum_k \frac{Z_k}{r_{ik}} + \sum_i \frac{p}{r_i} \quad (2)$$

In which  $\nabla_i^2$  is the Laplacian operator,  $Z_k$  is nuclear charge of nucleus  $k$ ,  $r_{i,k}$  is the distance between electron  $i$  and nucleus  $k$ ,  $r_i$  is the distance between electron  $i$  and the electron density function  $p$ . The first summation occurs over all electrons and accounts for the electron kinetic energy. The second sum accounts for attractive forces between electrons and nuclei. The final sum accounts for electron-electron repulsion.

The electronic Hamiltonian represents a many-body problem that is analytically unsolvable for systems with more than one electron. This description of electronic energy is dependent on the locations of electrons within a molecule as the total electron density must be known to calculate the electron-electron repulsive forces. There are ways to reach increasingly accurate solutions for determining electronic properties, but a wave function that exactly satisfies the Schrodinger equation cannot be found.

Due to the fact that we cannot derive a wave function that satisfies the Schrodinger equation we must find an alternate way to derive electronic properties. The first step of this process is specification of a basis-set. A basis-set is a set of basis functions, typically Gaussian functions, which are collectively used to build electron orbitals for an atom. After a basis set is specified, molecular orbitals can be determined by the linear combination of atomic orbitals approximation (LCAO) as follows.

$$\Psi = \sum_i a_i \Phi_i \quad (3)$$

In which  $\Psi$  is a molecular orbital,  $a$  is coefficient that can be optimized to reduce energy, and  $\Phi$  is an atomic orbital. The summation is taken over all atomic orbitals,  $i$ , in a molecule. After an initial guess of the coefficients is made the energy of the system can be calculated. The coefficients can then be changed to reduce the energy until a set of convergence criteria are met.

The energy of a molecular system under the LCAO approximation can be represented as

$$E = \frac{\int (\sum_i a_i \Phi_i) \hat{H} (\sum_j a_j \Phi_j) d\tau}{\int (\sum_i a_i \Phi_i) (\sum_j a_j \Phi_j) d\tau} = \frac{\sum_i \sum_j \int a_i \hat{H} a_j d\tau}{\sum_i \sum_j \int a_i a_j d\tau} \quad (4)$$

In which each integral of the form  $\int a_i \hat{H} a_j d\tau$  in the numerator is abbreviated  $H_{ij}$  and is referred to as a resonance integral, the integral in the denominator of the form  $\int a_i a_j d\tau$  is abbreviated  $S_{ij}$  and is referred to as an overlap integral, and  $d\tau$  indicates the integration should be carried out over all spatial coordinates. Solving for the energy of the molecular orbitals in a system is performed by solving for each value of  $E$  in the following secular equation:

$$\begin{bmatrix} H_{11} - ES_{11} & \cdots & H_{1N} - ES_{1N} \\ \vdots & \ddots & \vdots \\ H_{N1} - ES_{N1} & \cdots & H_{NN} - ES_{NN} \end{bmatrix} = 0 \quad (5)$$

The calculation of electronic properties from wave functions is typically referred to as wave function theory (WFT), but another body of thought exists in calculation of electronic properties. The Hohenberg-Kohn theorems show that electronic energy can be determined using only a function of electron density.<sup>17</sup> This was the beginning of density functional theory (DFT), which seeks to determine electronic properties using only the electron density function as opposed to a wave function.

Density functional theory uses an electron density function,  $p(\tau)$  to find the portions of electronic energy that are not possible to solve analytically. The molecular orbitals and density function are related in that the electron density is the sum of each squared occupied molecular orbital. The energy expression for a molecular system can, for simplicity, be written partially in terms of wave function theory as previously discussed, and partially in terms of the total electron density. The energy of a system would then be as follows:

$$E[p(\tau)] = \sum_i -\frac{1}{2} \int \Psi_i \nabla_i^2 \Psi_i d\tau - \sum_i \int \Psi_i \sum_k \frac{Z_k}{r_{ik}} \Psi_i d\tau + \frac{1}{2} \sum_j \sum_i \int \Psi_i \left[ \int \frac{p(\tau_j)}{r_{i,j}} d\tau_j \right] \Psi_i d\tau_i + E_{XC}[p(\tau)] \quad (6)$$

In which  $p(\tau_j)$  is the electron density function of electron  $j$  as a function of spatial coordinates,  $p(\tau)$  is the total electron density, and  $r_{i,j}$  is the distance between the electrons  $i$  and  $j$ . The first and second summations are identical to that of the molecular Hamiltonian, the third term is the classical Coulombic repulsion between the charge density and itself, and the final term is the exchange-correlation energy. The exchange-correlation energy term theoretically should account for the energy difference between a classical system and quantum mechanical system.

The coulomb-repulsion terms in equations 4 and 6 give rise to integrals of the following form:

$$\sum_i \sum_j \sum_k \sum_l \int a_i a_j \frac{1}{r} a_k a_l dr \quad (7)$$

These so called two-electron integrals scale as  $N^4$  with  $N$  being the number of basis functions. It is for this reason that DFT and WFT methods are only practicable for systems of several tens of atoms, or a few hundred if these integrals are parameterized instead of explicitly calculated.

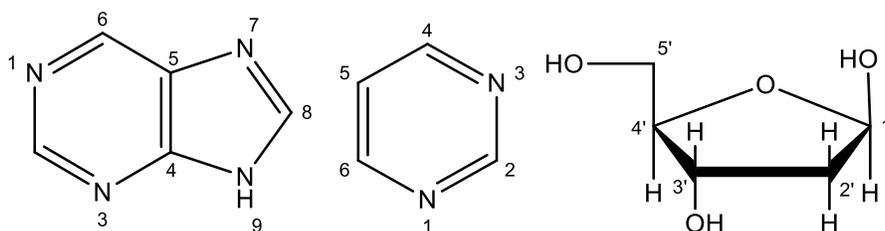
## **DNA Structure and Function**

Deoxyribonucleic Acid (DNA) is a polymeric molecule used by cells to carry genetic information. The information contained within DNA is transcribed and translated into the primary structure of proteins, which are then used to make tertiary proteins that carry out the various functions necessary to maintain life. Although mutations in DNA are vital to continued survival through evolutionary processes, mutations have the potential to be harmful. Cells have adapted to the potential of DNA mutation by creating recognition and repair processes to ensure the integrity of DNA from generation to generation. It is, therefore, important to understand the structure and function of DNA, as well as how chemical damage influences the properties of a DNA molecule.

The structure of DNA, as it pertains to this thesis, can be understood in three parts; the primary, secondary, and tertiary structure. The primary structure is defined as the sequence of individual nucleotides that create the larger DNA molecule. Secondary structure can be understood as the complimentary pairing to a second strand of DNA. The tertiary structure describes the helical nature of a double-stranded DNA molecule (dsDNA).

A DNA molecule's primary structure is its sequence of nucleotides. A nucleotide is itself made of three parts: 2'-deoxyribose, a nitrogenous base, and a phosphate group

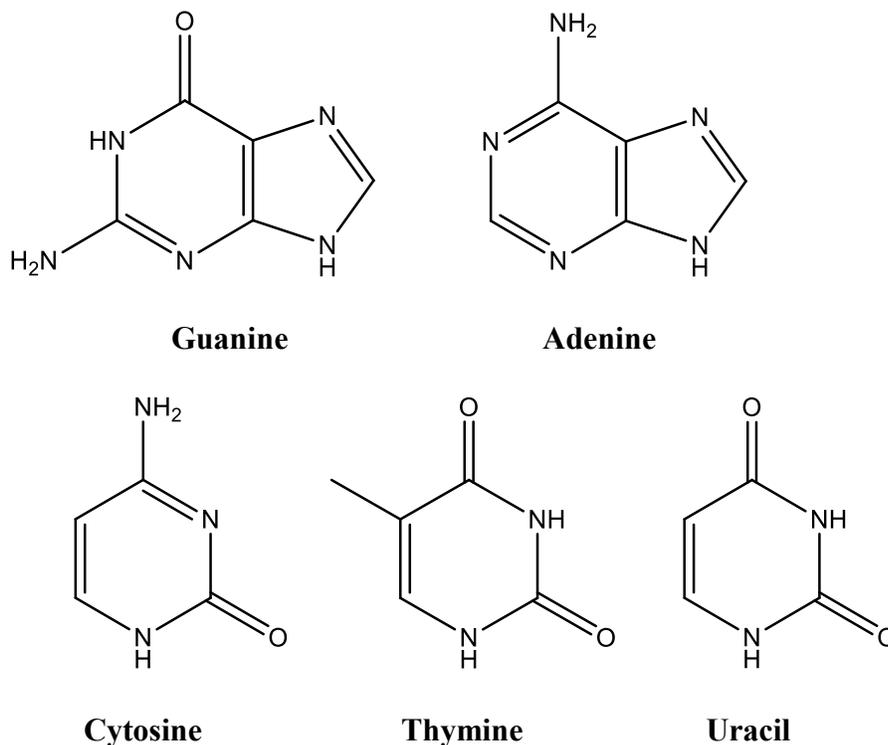
connecting it to the other nucleotides in sequence. The numbering schemes and structures are available in Figure 4. A nitrogenous base is bonded to the 1' carbon in deoxyribose, which is then bonded to other nucleotides via a phosphodiester bond at the 5' and 3' positions. The type of nitrogenous base at each position within the DNA molecule determines the protein coding information contained within the DNA molecule. The structures of the four nitrogenous bases found in DNA; cytosine, thymine, adenine, and guanine are shown in Figure 5. The former two are pyrimidine derivatives, the latter purine. DNA can be made to contain unique information by varying the type of nitrogenous base present in each position on the DNA molecule.



**Figure 4:** Numbering scheme for purines, pyrimidines, and 2'-deoxyribofuranose, in order from left to right.

The secondary structure of DNA is the pairing with a complementary second strand. DNA is usually found in nature as two complementary strands wrapped around each other in a right-handed helical fashion and held together by hydrogen bonds between the nitrogenous bases. The strands are antiparallel, in that the direction from the 5' carbon to the 3' carbon is reversed relative to the opposite strand. The nitrogenous bases in the center are paired purine to pyrimidine, and associate as Watson-Crick pairings (A-T, and C-G). The hydrogen bonding scheme between complimentary bases and connectivity scheme which results in the primary structure of DNA can be seen in

Figure 6. The second strand of a dsDNA molecule can be a separate strand or a complimentary section of the same strand.



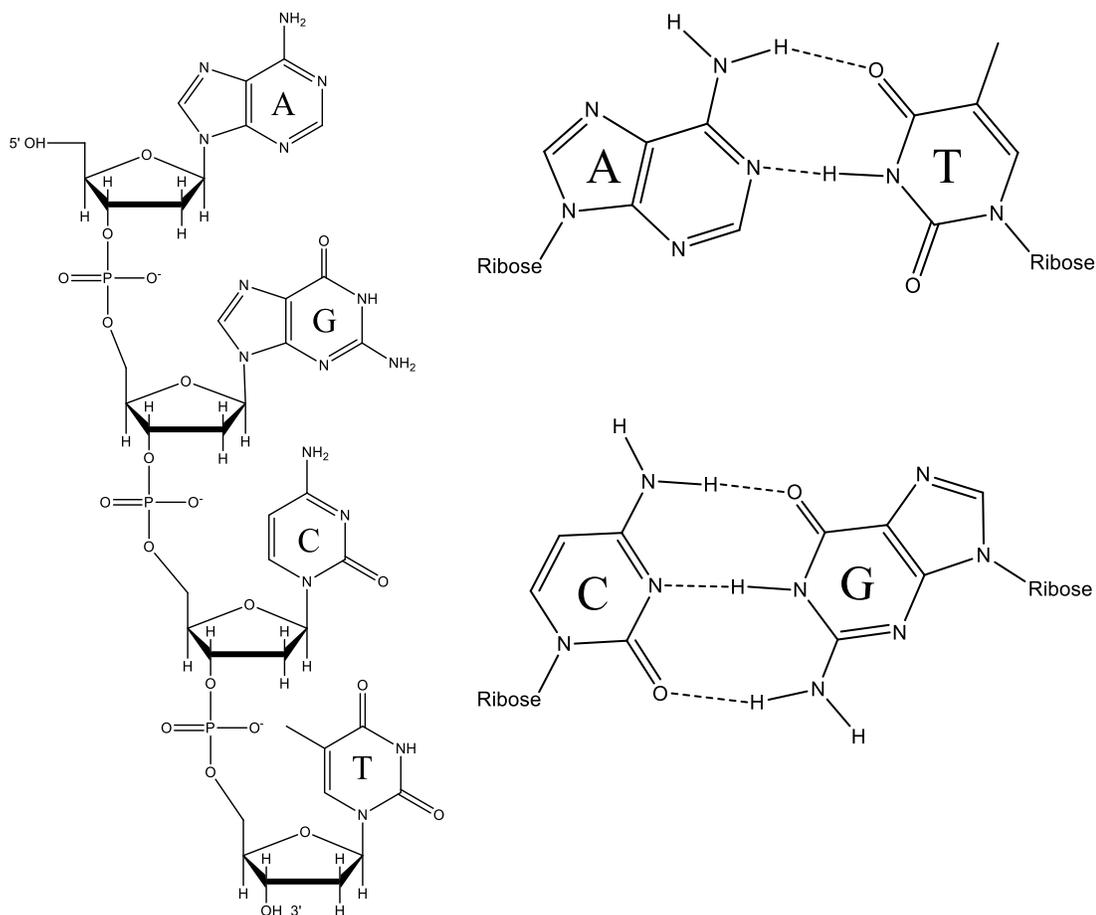
**Figure 5:** Structures of the nucleotides in DNA and Uracil.

The tertiary structure of a DNA molecule describes the right-handed helix assumed by dsDNA. The two strands wind around each other at a rate of 10 – 12 base pairs per turn. In biological systems the most common form of dsDNA is B-DNA with a rise of roughly 10 bases per turn, a diameter of 20 angstroms, and the 2' carbon puckered in the 5' direction. Under dehydrating conditions dsDNA can assume a conformation with 11-12 base pairs per turn, a diameter of 26 angstroms, and the 2' sugar puckered in the 3' direction. This tighter coiled conformation is called A-DNA. The third conformer of dsDNA is Z-DNA. Z-DNA forms a left-handed helix with a rise of 12 bases per turn,

a diameter of 18 angstroms, a 3' sugar pucker for purines and 2' sugar pucker for pyrimidines. This form is only assumed by alternating purine-pyrimidine sequences at high salt concentrations. B-DNA is the most common in biological systems and was the first form described by Watson, Crick, Wilkins, and Franklin.<sup>17</sup> In dsDNA the helix is formed by the winding of the phosphate backbone around the bases. The distance between phosphate groups is asymmetric with respect to the direction of the helix, resulting in alternating larger (major) and smaller (minor) interphosphate distances. The presence of a major and minor groove is common in all DNA conformers. Figure 7 shows the presence of the major and minor grooves as they exist in B-DNA. A-T and C-G base pairs are essentially equal dimensions, so the tertiary structure of a DNA molecule is unchanged by its primary structure so long as Watson-Crick base pairing remains intact.<sup>17</sup>

Several factors stabilize the tertiary structure of dsDNA. The most commonly understood factor that stabilizes DNA is the presence of hydrogen bonds between purine-pyrimidine base pairs in the center of the helix. Hydrogen bonds, although weak individually, create an activation energy that must be overcome before the two strands of DNA will separate. Ionic interactions can also stabilize the helical structure by shielding like-charges along the phosphate backbone. Cations such as  $Mg^{2+}$ ,  $Na^+$ , and  $Li^+$  act to shield negative charges. Divalent cations act to shield anionic interactions to a much higher degree than monovalent cations. It is for this reason magnesium is often required as a co-factor for reactions with anionic species, including DNA molecules. The final factor that stabilizes the tertiary structure of DNA is stacking interactions. Stacking interactions are the forces that result from the stacked base-pairs in the center of a DNA

double helix. These are typically thought of as interactions between the  $\pi$ -orbitals of the aromatic rings.<sup>18</sup>



**Figure 6:** Primary structure of AGCT DNA sequence and hydrogen bonding between bases. When listing DNA sequence convention is to list in the 5' to 3' direction.

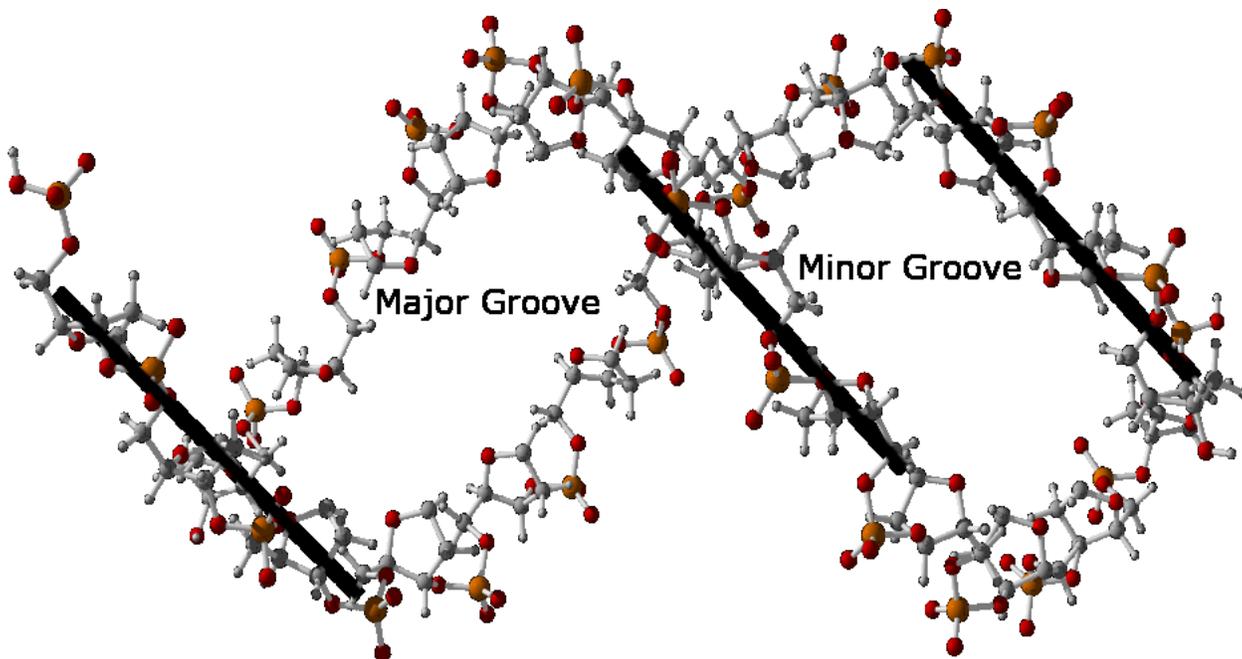
The Drew-Dickerson Dodecamer is a dsDNA molecule of sequence

CGCGAATTCGCG first characterized by Horace R. Drew and Richard E. Dickerson.<sup>19</sup>

This system has become a standard model for examining the dynamics of B-DNA.

Extensive computational work has been performed to examine the dynamics of the

system; however, the work has been primarily limited to force-field methods due to the large size of the system.



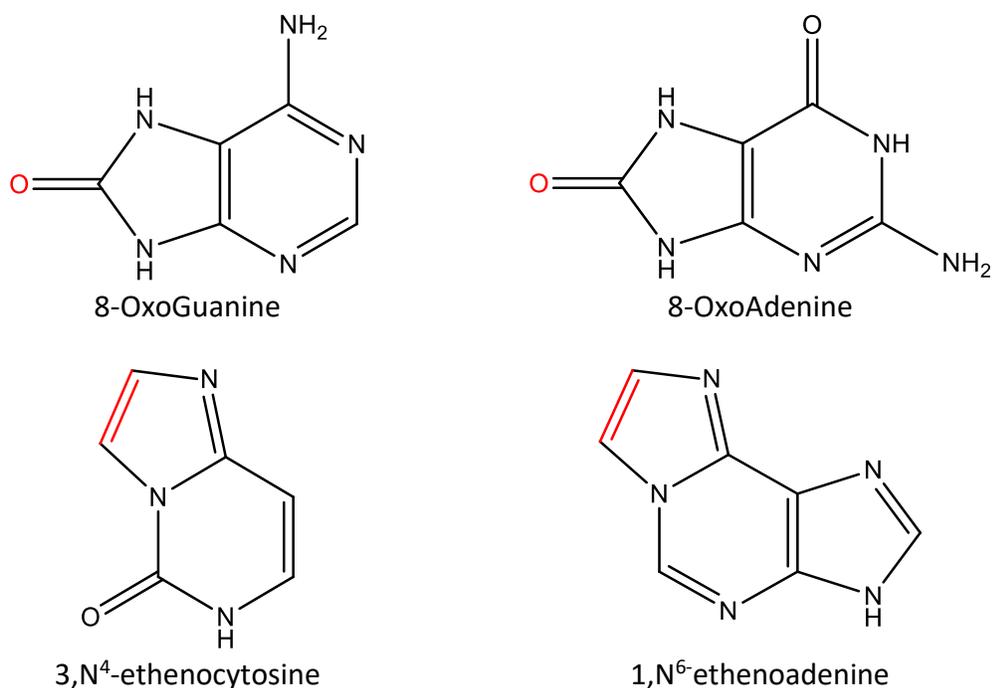
**Figure 7:** Backbone structure of DNA. Bases have been removed and bars added for ease of distinction between the major and minor grooves. Orange = P, Grey = C, Red = O, Blue = N, Small = H.

8-Oxo-2'-DeoxyAdenosine (8-oxo-dA) and 8-Oxo-2'-DeoxyGuanosine (8-oxo-dG) are DNA lesions known to be products of reactive oxygen species (ROS) reacting with nitrogenous bases in DNA. ROS in cells exist as oxygen containing radicals including hydroxyl radical, triplet oxygen, or superoxides. ROS can be generated by exposure to UV radiation, excess heat, or as a normal function of oxygen metabolism during respiration. Reactions of DNA with ROS are potentially mutagenic and have been studied extensively.<sup>20, 21, 22</sup>

The ethene adducts 3,N<sup>4</sup>-ethenocytidine ( $\epsilon$ dC) and 1,N<sup>6</sup>-ethenoadenosine ( $\epsilon$ dA) are highly mutagenic. The structures of these two molecules can be seen in the bottom

row of Figure 8. The creation of these DNA lesions has been shown to be byproducts of lipid peroxidation as the result of oxidative stress, and industrial carcinogens such as vinyl chloride. They have also been implicated as potential tumor-causing agents in metal storage disorders such as Wilson's Disease. In HeLa cells EdA has been shown to have greater mutagenic potential than 8-oxo-dG.<sup>23</sup> The mutagenic potential from EdA results from its potential to cause large deletions during replication or repair of the genome, as well as missense mutations up or down sequence from the lesion site. EdC is primarily responsible for replacement of cytidine with thymidine or adenosine during replication.

24, 25



**Figure 8:** Modified nucleotides with mutagenic changes in red.

Canonical nucleotides can also be considered to be lesioned or damaged bases if they result in a mismatched pair or result in a nucleotide that would not normally be in

the molecule. The usual instance of this is the appearance of the uracil base in DNA, despite it normally only appearing in RNA. This can result from deamination of cytosine, which replaces the  $\text{NH}_2$  group with an oxygen atom. Uracil is less likely to pair appropriately and so is conducive to mutagenesis. For this reason, 2'-deoxyuridine can be considered as a lesioned base when present in the DNA molecule.<sup>26</sup>

Understanding how cellular DNA repair mechanisms recognize damaged sequences remains an active area of research. An understanding of how the structure of DNA is distorted in response to damage may assist in determining how repair mechanisms are initiated by cells. It has been shown, based on crystallographic data, that incorporation of a  $\epsilon\text{C}$  or  $\epsilon\text{A}$  base produces a significant shift in positioning of the lesion base.<sup>27, 28</sup> The larger space required by the etheno-adduct, as well as the disruption of hydrogen bonding, causes a 3' sugar pucker despite the rest of the helix existing as B-DNA. 8-oxo-dG has been shown via NMR methods to significantly distort the sugar-phosphate conformation when incorporated into the GAATTC site of the Drew-Dickerson Dodecamer.<sup>29</sup>

### **NMR Spectroscopy**

Nuclear Magnetic Resonance (NMR) is the phenomenon in which a nucleus having a magnetic moment absorbs or releases radiation while under an external magnetic field. NMR spectroscopy is a technique that can be used to determine structural features of individual molecules by utilizing the phenomena of NMR. The information that follows can be found in greater detail in works by Jacobsen<sup>30</sup> and McQuarrie.<sup>15</sup>

Spin is a quantum mechanical property possessed by fermions, such as electrons, and ensembles of fermions, such as atomic nuclei. We can define a spin operator,  $\hat{S}_z$  to operate on a function and return the spin value of the function as in the following way:

$$\hat{S}_z\alpha = \hbar m_s \alpha = \frac{1}{2} \hbar \alpha \quad (8)$$

In which  $m_s$  is the spin quantum number, and  $\frac{1}{2}\hbar$  is the angular momentum in the Z-direction of the particle described by wave function  $\alpha$ . For electrons the spin quantum number is limited to values of  $\pm\frac{1}{2}$ . For nuclei, which are ensembles of several particles of spin  $\pm\frac{1}{2}$ , the angular momentum may reach values of higher magnitude.

Calculating the energy of an NMR active nucleus in a magnetic field can be done using the corresponding Schrodinger equation:

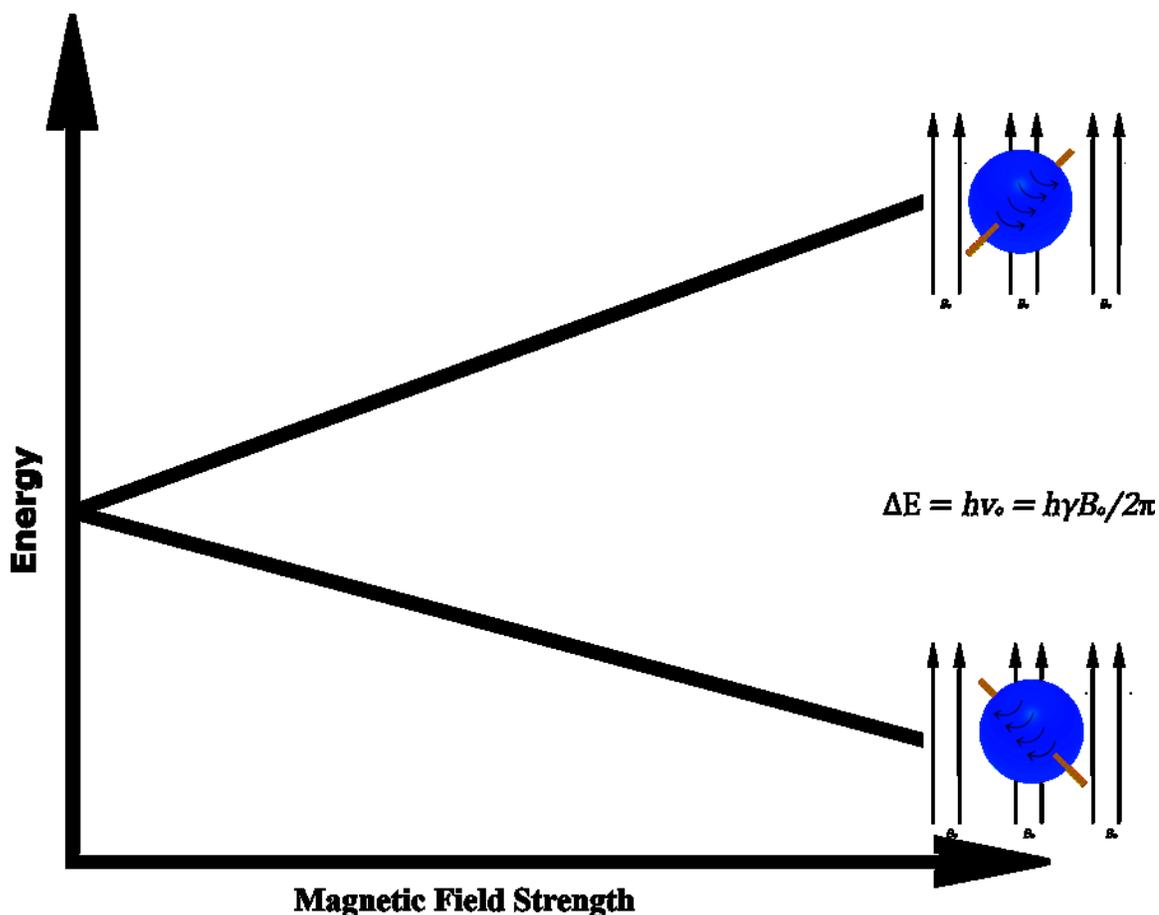
$$\hat{H}\Psi = V\Psi = -\gamma B_z \hat{I}_z \Psi = -\gamma B_z m_s \hbar \Psi = E\Psi \quad (9)$$

In which  $\hat{H}$  is the spin Hamiltonian,  $V$  is the magnetic potential energy of the nucleus in a magnetic field of magnitude  $B_z$ , and  $\gamma$  is the magnetogyric ratio of the nucleus,  $\hat{I}_z$  is the operator to solve for angular momentum in the Z direction,  $m_s$  is the spin quantum number, and  $\hbar$  is the reduced Planck constant.

In the absence of an external magnetic field there is no energy difference between spin quantum numbers. However, when an external magnetic field is present the spin states become non-degenerate. The higher-energy spin states are denoted with a more negative spin quantum number. The energy change due to an NMR active nucleus changing spin states in an external magnetic field is described by the following equation:

$$\Delta E = \Delta m_s \hbar \gamma B_z = \hbar \gamma B_z \quad (10)$$

The energy difference between spin states may be probed spectroscopically. Typical NMR spectrometers utilize magnetic fields with magnitudes ranging from less than 1 Tesla to an excess of 20 Tesla. The corresponding Larmor frequency therefore corresponds to the radio portion of the electromagnetic spectrum, but can be altered as in Figure 9. By emitting a radio frequency pulse across a sample nuclei can be flipped from low-energy spin-states to high-energy. By measuring the response as nuclei relax to equilibrium an NMR spectrum can be collected for a given sample.



**Figure 9:** Energy difference as a function of external magnetic field strength.

Electrons in a molecule generate their own local magnetic field and act to shield nuclei from an external magnetic field. The local changes in the magnetic field experienced by a nucleus allow information about the chemical environment of each nucleus to be obtained.

An NMR spectrum is organized as a plot of chemical shift versus intensity, and the value of the chemical shift gives meaningful information about the environment occupied by a nucleus. The change in Larmor frequency relative to a standard is referred to as a chemical shift. Chemical shifts can be calculated (in parts-per-million) using the following equation in which  $\nu$  is a measured frequency:

$$\delta = \frac{\nu_{sample} - \nu_{reference}}{\nu_{reference}} * 10^6 \quad (11)$$

Chemical shifts arise due to changes in the external magnetic field by nearby electrons. The magnetic field introduced by the spin of an electron acts to shield a nucleus from the external field. A higher chemical shift value typically corresponds to a lower electron density in the immediate vicinity. A nucleus is said to be deshielded from the external magnetic field and rest downfield on an NMR spectrum if it has a higher chemical shift.

Spin-spin coupling provides information about the relative position of nuclei in a molecule. Spin-spin coupling, or J-coupling, arises as a result of changes in the external magnetic field due to other nearby nuclei. When two nuclei with different chemical environments are near in space, the magnetic moments of each proton are either aligned or anti-aligned. Each arrangement appears in an NMR spectrum with equal intensity. The change in frequency because of mutual interaction with another nucleus is not dependent on the external magnetic field, and the value can be presented as a coupling constant ( $J$ ) in units of Hz. Although the change in the external field begins from other

nuclei, the coupling occurs through bonds and only appears when the nuclei are very close together (usually three bonds or less). Multiple NMR active nuclei in close proximity yield increasingly complex, but ultimately predictable splitting and intensity patterns.

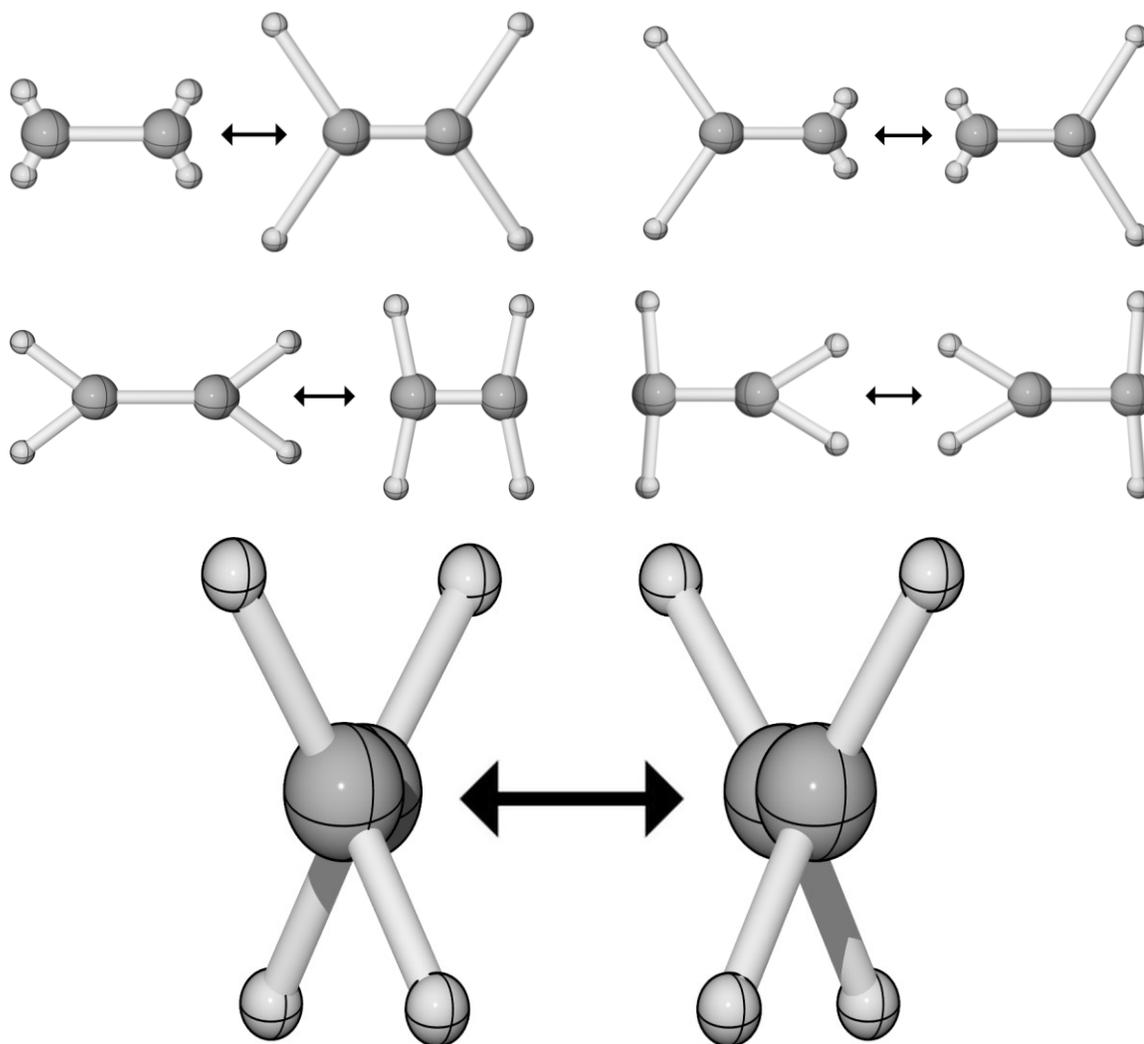
Often there are several protons in a molecule that are in equivalent chemical environments. This can be due to symmetry considerations or rotational freedom. Nuclei that are chemically equivalent will feel the same external magnetic field and will show the same splitting effects and chemical shift. Under these conditions nuclei will show additive intensity in an NMR spectrum.

Relaxation is the process by which the perturbed nuclear spins in an NMR experiment return to equilibrium. When an NMR experiment is conducted, nuclear spins are flipped to oppose the external magnetic field and the response is measured. Relaxation times for nuclear spin are large when compared to other chemical phenomena such as conformational isomerism, molecular vibrations, and protonation/deprotonation. Although short relaxation times may occur in specific experiments, the time it takes for a population of nuclei to return to equilibrium after an NMR experiment is typically in a range of tenths of a second to several minutes. A result of the slow nature of relaxation times is that changes in the NMR spectrum due to phenomena such as vibrations, rotations, and acid/base exchange are averaged into one single signal for a group of chemically equivalent protons.

## Molecular Vibrations and Polarizability

A molecular motion is said to be a vibration if the motion of the relative positions of its constituent atoms are changing in a periodic and synchronous fashion. This implies that all atoms reach their equilibrium and maximum displacements simultaneously and the center of mass does not change over the course of the vibration. Each normal vibrational mode has an associated quantum number beginning at zero and increasing in integer values. Vibrations have an associated quantum number that determines the energy level and amplitude of the vibrational state. A higher vibrational energy level corresponds to a greater vibrational amplitude and total energy.

Vibrations are typically described as a set of changing internal coordinates. There are three types of internal coordinates; bond lengths, bond angles, and dihedral angles. Visualizations of changes in these coordinates can be seen in Figure 10. Bond lengths are the distance between nuclei. Bond angles are the angle formed by any three nuclei. A dihedral angle is the angle between two intersecting planes. When defining a dihedral angle, the first plane is defined by atoms A, B, C and the second plane is defined by atoms B, C, D. To unambiguously define the relative location of all atoms in a non-linear molecule it is necessary to define  $N - 1$  bond lengths,  $N - 2$  bond angles, and  $N - 3$  dihedral angles, for a total of  $3N - 6$  degrees of freedom. The set of normal coordinates of a system is typically represented by the variable  $Q$ . The potential complexity of vibrations makes normal coordinates a useful tool for describing the changes in atomic positions in a vibration.<sup>31</sup>

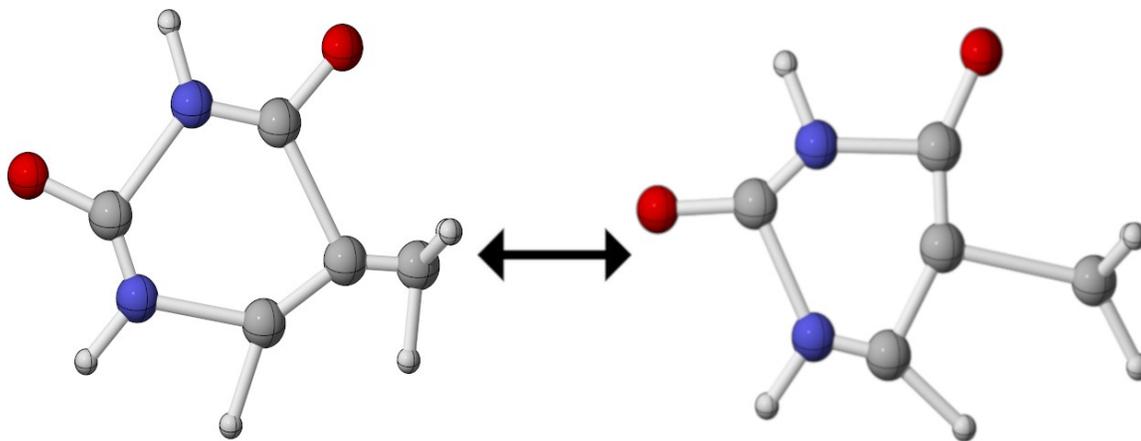


**Figure 10:** Stretching, bending, and torsions demonstrated through vibrations of the ethene molecule. Vibrational amplitudes are arbitrary. Stretches (top row) may occur at multiple sites on a molecule and may occur in a symmetric (left) or anti-symmetric (right) fashion. Similarly, bending (second row) may occur in a symmetric (left) or anti-symmetric (right) fashion. Torsions occur for all atoms adjacent to the torsional planes, so all hydrogens rotate during torsional motion in ethene.

Although it would be convenient, it is not necessarily the case that each normal mode corresponds to change along one internal coordinate. There are cases in which one normal coordinate change dominates the vibration, such as O-H stretches, but frequently a vibrational mode is a mix of several internal coordinate changes happening simultaneously. In even small molecules it is expected that certain normal modes will

span the entire molecule. It can be seen in Figure 11 that vibrations quickly become quite complex. The potential for complex vibrational motions necessitates experimental data to aid in vibrational analysis of a molecule.<sup>32</sup>

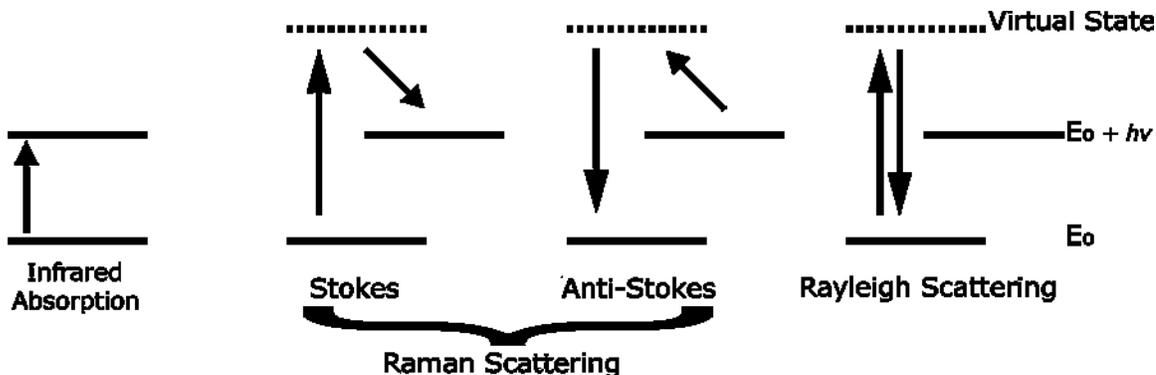
Information about vibrations can be gained through experimental methods such as Raman and Infrared (IR) spectroscopy. IR spectroscopy measures the amount of infrared light of a specific frequency is absorbed by a sample. The infrared region corresponds to the energy difference between vibrational levels and the absorption of an infrared photon can induce a transition between them. Raman spectroscopy measures scattering of photons. An IR spectrum is organized as a graph of photon frequency versus absorption. Raman spectroscopy involves scattering of photons and measuring the frequency changes of those scattered photons. When light is scattered by a molecule the molecule may absorb energy from the photon, or the photon may take energy from the molecule. The energy change of the photon results in an equal and opposite change in energy of the molecule. The energy change causes a transition in the molecule's vibrational state. By measuring the energy difference between the incident and scattered photons information about vibrational modes can be obtained. The path of excitement and summary of initial and final states can be found in Figure 12. A Raman spectrum is a graph of the frequency change of scattered photons versus the number of scattered photons at that frequency.



**Figure 11:** Complex vibration of Thymine molecule. It is typical for vibrations to include motion along several normal coordinates.

There are several types of vibrational transitions that can be present as peaks in a vibrational spectrum. The simplest to understand are fundamental modes. A fundamental mode is a change in vibrational quantum number from zero to one and is usually the most common feature of a Raman or infrared spectrum. The second type of peak is due to overtones. Overtones are changes in vibrational quantum number of 2 or more. The energy of an overtone is usually near an integer multiple of its corresponding fundamental mode. The third type of peaks are combination bands. A combination band occurs when multiple vibrational quantum numbers are changed simultaneously. The simplest combination band is when two fundamentals occur due to a single photon. The resulting peak frequency is then at the sum of the two fundamentals. More complex patterns can emerge when a combination band is due to an increase of one vibrational mode and a decrease of another. The result of this combination band is a peak at the difference of the two modes. Combination bands can also occur between allowed and forbidden modes if their combined vibration is Raman active. The final type of peak that can be present in a Raman spectrum is the result of Fermi Resonance. Fermi resonance

occurs when two vibrational modes have nearly equal frequencies. The result is a mixing of vibrational modes that separates the two frequencies and brings the two intensities closer together. Fermi resonance, combination bands, and overtones can also occur together to complicate interpretation of Raman spectra.<sup>31</sup>



**Figure 12:** Energy level diagram for infrared absorption, Raman scattering, and Rayleigh scattering.

### Calculating Spectroscopic Intensity

A vibration is said to be infrared active if, over the course of the vibration, the molecule undergoes a change in dipole moment. This can be represented as a transition moment integral of the form<sup>31</sup>:

$$\int \Psi_m^* \mu \Psi_n dQ \quad (12)$$

In which  $\Psi_m^*$  is the wave function associated with the initial vibrational state,  $\Psi_n$  is the wavefunction associated with the final vibrational state,  $Q$  is the set of normal coordinates, and  $\mu$  is the dipole moment operator. When the transition moment integral is zero the vibration is said to be forbidden.

A vibration is said to be Raman active if, over the course of the vibration, the molecule undergoes a change in polarizability. Polarizability is the tendency of a charge distribution to change because of an external electric field. An induced dipole and polarizability can be related by the following equation:<sup>33</sup>

$$\mu = \alpha E \quad (13)$$

In which  $\alpha$  is the polarizability,  $E$  is the external electric field, and  $\mu$  is the dipole induced by that electric field. The transition moment integral for a Raman transition is of the form<sup>32</sup>:

$$P_{n,m} = \int \Psi_m^* \alpha \Psi_n dQ \quad (14)$$

In which  $\alpha$  is the polarizability operator, and the other terms are equivalent to the terms in the infrared transition moment integral.

It is difficult to qualitatively imagine the degree of polarizability of a molecule. The induced dipole of a molecule in the X direction can be due to an electric field in any of the X, Y, or Z directions. The polarizability must then be represented as a tensor of the form:

$$\alpha = \begin{bmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{bmatrix} \quad (15)$$

In which the first subscript is the electric field in that direction, and the second subscript is the direction of the dipole in that direction. In general, the polarizability matrix is a symmetric matrix, in that off-diagonal terms of the form  $\alpha_{ab}$  are equal to the corresponding term  $\alpha_{ba}$ . This reduces the total number of unique polarizability terms to six. Still, the dependence of polarizability on all directions underlies why it is difficult to

qualitatively assess the polarizability of a system. A Raman transition moment integral should include a calculation for each of these polarizability terms.

Light scattered by the Raman effect is itself polarized. When a liquid sample is irradiated with natural (unpolarized) light traveling in the X direction, and the detector is oriented in the Y direction, then the ratio of light polarized in the Z or X direction (the depolarization ratio) will assume values  $0 \leq p < \frac{6}{7}$ .<sup>31,34</sup>

Polarizability can be expanded as a Taylor series with the form:

$$\alpha = \alpha_0 + \left(\frac{\partial\alpha}{\partial Q}\right)_0 Q \quad (16)$$

In which  $\alpha_0$  is the polarizability of the molecule at the equilibrium coordinates, and  $\left(\frac{\partial\alpha}{\partial Q}\right)_0$  is the rate of change of the polarizability along the vibrational coordinates at equilibrium. There are additional terms of the expansion, but they are only relevant when discussing overtones and other weakly-active Raman events. The transition moment integral using this expansion then becomes:

$$\alpha_0 \int \Psi_m^* \Psi_n dQ + \left(\frac{\partial\alpha}{\partial Q}\right)_0 \int \Psi_m^* Q \Psi_n dQ \quad (17)$$

When  $m = n \pm 1$ , corresponding to a fundamental transition, the left integral vanishes owing to the orthogonality of vibrational wave functions. The value for the right integral is solvable under the harmonic approximation yielding the following equation<sup>31</sup>:

$$P_{n,n\pm 1} = \left(\frac{\partial\alpha}{\partial Q}\right)_0 \sqrt{\frac{(n+1)h}{8\pi Mv}} \quad (18)$$

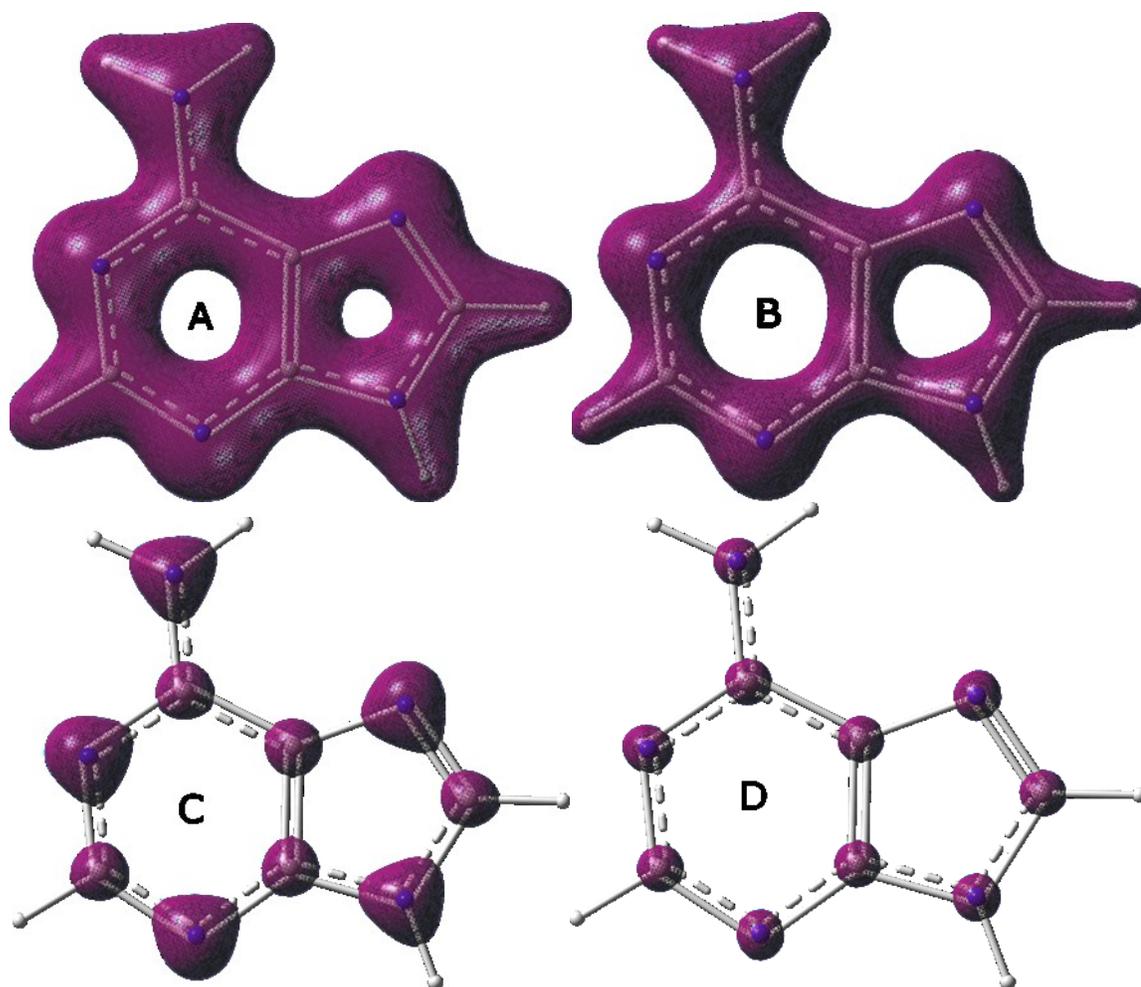
In which  $h$  is the Planck constant,  $M$  is the reduced mass, and  $v$  is the frequency of the mode transition.

The transition moment integral is only one part of the intensity of Raman scattering. Observed Raman intensity at a right-angle to a non-polarized source is given by the following equation<sup>31</sup>:

$$I_n = \frac{64\pi^2 h}{3c^2} g_o I_o \frac{(v_o - v)^4}{Mv \left(1 - e^{-\frac{hv}{kT}}\right)} \left(\frac{\partial \alpha}{\partial Q}\right)_0^2 \left(1 + \frac{7}{6 - 7p}\right) \quad (19)$$

In which  $I_n$  is the observed Raman intensity,  $c$  is the speed of light,  $g_o$  is the degeneracy of the vibrational mode,  $I_o$  is the intensity of the incident light source,  $v_o$  is the frequency of the light source,  $v$  is the frequency change of the vibrational transition, the exponential term is a factor to account for molecules not in the ground state at a particular temperature, and  $p$  is the depolarization ratio. This equation is significant in that it relates the observed Raman intensity of vibrational modes to the polarizability derivative of a particular vibrational mode. By knowing or calculating the other terms we can then make quantitative assignments of polarizability derivatives and determine trends for various functional groups.<sup>31</sup>

An electronic volume can be defined as the volume encompassed by a surface of an electron density. Within molecules there is an inverse relationship between the specified electron density and volume, as can be seen in Figure 13, with the lower electron density occurring in regions further from atomic nuclei. As electrons move away from the molecule the force on the electron by the nuclei weakens, and external electric fields become more important to the electron's motion.



**Figure 13:** Electronic surface at 0.1 (A), 0.2 (B), 0.4 (C), and 0.8 (D)  $\frac{e^-}{\text{Bohr}^3}$ .

The polarizability of atoms making up a material is correlated with average molecular volume of its constituent molecules according to the Clausius-Mossotti equation:

$$\frac{\varepsilon^2 - 1}{\varepsilon^2 + 2} = \frac{N}{3\varepsilon_0 V} \alpha \quad (20)$$

In which  $\varepsilon$  is the dielectric constant of a material,  $\varepsilon_0$  is the permittivity of free space,  $N$  is the number of molecules present in the sample, and  $V$  is the volume of the sample. By rearranging we can find a linear relationship between the polarizability of a substance and its average molecular volume, assuming the function of the material's dielectric constant

varies slowly with the volume occupied by the sample. This relationship, although approximate, should hold for non-polar molecules.<sup>35</sup>

In addition to molecular volume, it is useful to include a measure of how tightly electrons are held in place. Inclusion of this term should improve correlation by accounting for the non-uniformity of molecular electron densities, and the quantum nature of molecular orbitals. Towards this end inclusion of a calculated average local ionization energy (ALIE) can be used<sup>36</sup>:

$$I_{av} = \frac{1}{n} \sum_{i=1}^n \frac{1}{p(r_i)} \sum_{j=1}^m \varepsilon_j p_j(r_j) \quad (21)$$

Electron density contributions from each molecular orbital at a certain point in space,  $p_j(r_j)$ , is evaluated and multiplied by the energy of the contributing orbital  $j$ . The sum of the density-energy products is then divided by the total electron density,  $p(r_i)$ , at that point. This yields an ionization energy at a single point within the molecule, or a local ionization energy. The local ionization energy is then evaluated over a number of points ( $n$ ) and averaged. Using this measure of electron tightness allows one to restrict the chosen points for an ALIE calculation to within the specified electron density surface.<sup>13,36</sup>

Raman intensity should be correlated to changes in volumes as the molecule progresses along its vibrational coordinates. By moving the molecule along its vibrational coordinates, we should see a volume change proportional to the polarizability change. If we calculate the relative intensity of Raman peaks we arrive at the following relationship:

$$I_{Rel} = \frac{\Delta V_a^2 (v_{ao} - v_a)^4 \left( \frac{13 - 7p_a}{6 - 7p_a} \right)}{\Delta V_b^2 (v_{bo} - v_b)^4 \left( \frac{13 - 7p_b}{6 - 7p_b} \right)} * \frac{M_b v_b \left( 1 - e^{\frac{-hv_b}{kT}} \right)}{M_a v_a \left( 1 - e^{\frac{-hv_a}{kT}} \right)} \quad (22)$$

In which the subscripts  $a$  and  $b$  refer to vibrational modes, and  $\Delta V$  a function of the volume change and local ionization energy of the molecule after a small amount of motion along the vibrational coordinates. By using this expression there should arise a correlation between experimental Raman intensities and theoretically predicted electronic volumes. Setting the intensities as relative values has the advantage of all constants and difficult-to-measure terms canceling out.<sup>31</sup>

## CHAPTER 3: METHODS

### Programs Utilized

All calculations were completed using the Gaussian16 program<sup>14</sup>, with the exception of volume and average local ionization energy calculations, which were completed using the WFA-SAS program.<sup>37</sup>

### Model Chemistries

To perform a quantum chemical calculation requires specification of a model chemistry. A model chemistry requires a specification of how to calculate energy of the system, and how to construct the atomic orbitals of the system. The former is typically done through wave function theory (WFT) methods, or density functional theory (DFT) methods. The latter is done through specifying a pre-built basis set, or by specifying a custom basis set of your own design.

**Wave Function Theory.** The Hartree-Fock method (HF) defines the energy of a molecular system as the sum of the one-electron Fock operator over each electron<sup>16</sup>:

$$f_i = -\nabla_1^2 - \sum_k \frac{1}{r_{ik}} + \sum_{j \neq i} \sum_j \frac{1}{r_{ij}} \quad (23)$$

This formulation contrasts the molecular Hamiltonian in that the electron-electron repulsion is now treated in terms of pairwise interactions, and kinetic energy is treated only for individual electrons. This has the effect of simplifying the expression from an unsolved many-body problem to a series of two-body problems. The primary limitation of this approximation is that it does not capture the influence electronic motion has on the

motion of other electrons, known as electron correlation energy. The greatest accuracy that can be obtained using HF is referred to as the Hartree-Fock limit.

When calculating electron-electron repulsion, the following two-electron integrals are used:

$$\int \Psi_i^2 \frac{1}{r_{ij}} \Psi_j^2 dr - \int \Psi_i \Psi_j \frac{1}{r_{ij}} \Psi_j \Psi_i dr \quad (24)$$

The left integral captures classical repulsion due to the like charges of electrons. The right integral captures electron exchange, a quantum mechanical effect that acts to reduce the coulomb repulsion energy of the system. As the molecular orbitals used are the sum of atomic orbitals these integrations must be performed for each atomic orbital used to create the molecular orbital.<sup>16</sup>

Semi-empirical methods typically use the Fock operator to define the energy, then make further approximations to save computational time. Typically, these are done by parameterizing integral values and reducing the size of the basis set. AM1, PM3, and PM6 are specific parameterizations of the modified neglect of differential overlap (MNDO) formalism. Under the MNDO formalism a minimum basis set is used, two-electron integrals originating from more than two atomic centers are set to zero, and a nuclear repulsion term is added. AM1 was specifically parametrized to improve hydrogen bond lengths, and PM3 was reparametrized from AM1 using more sophisticated statistical methods and the same goal.<sup>16</sup> The PM6 method was created with the addition of parameters for d-orbitals and an improved nuclear repulsion term. The parameterization of d orbitals within PM6 extends the applicability of this method beyond main group elements.<sup>38</sup>

Whereas semi-empirical methods were created with simplifications to HF, Moller-Plesset perturbation theory was created with the opposite philosophy. By taking the Fock operator as an unperturbed Hamiltonian and adding a correlation potential as a perturbation one can derive successive corrections to the Fock operator to account for electron correlation. The first non-vanishing correction is the second correction and is of the form:

$$E_0^2 = \sum_{k,l} \frac{|\langle \Psi_k | H' | \Psi_l \rangle|^2}{E_l - E_k} \quad (25)$$

In which  $H'$  is the difference between the true electronic Hamiltonian and the Fock operator,  $\Psi$  are molecular orbitals,  $k$  designates an orbital as occupied,  $l$  designates an orbital as unoccupied, and  $E_l - E_k$  is the difference between the corresponding orbital energies. Moller-Plesset perturbation methods are designated as MPn, with n being the highest order correction applied. When only the second order correction in equation 25 is used the method is called MP2. This method is generally enough to give significantly improved results over HF.<sup>16</sup>

**Density Functional Theory.** The Kohn-Sham one-electron operator is defined as:

$$h_i = -\frac{1}{2}\nabla_1^2 - \sum_k \frac{Z_k}{r_{ik}} + \int \frac{p(r_j)}{r_{ij}} dr_j + \frac{\partial E_{xc}}{\partial p(r_j)} \quad (26)$$

The first two terms are equivalent to the first two terms in the Fock operator. The third term is the operator for a classic charge density repelling itself. The final term can be thought of as an operator that returns the exchange-correlation energy,  $E_{xc}$ . Without an exactly defined exchange-correlation energy term the Kohn-Sham operator can only be said to define a system of orbitals in which electrons do not interact. These orbitals are

called Kohn-Sham orbitals. Several functionals have been proposed to fill in the undefined exchange-correlation energy function.<sup>16</sup>

Approximations used in the evaluation of exchange-correlation functions can be organized from least accurate to most accurate. This organization has been compared to Jacob's ladder, with the lowest rungs being closest to the imperfect Earth and the highest rung being closest to the heaven of perfect chemical accuracy. The first rung on the ladder is the local density approximation (LDA). Under LDA the evaluated energy due to a specific point of electron density depends only on the density at that single point and is evaluated as though it is a part of a uniform electron density. The second rung is the generalized gradient approximations (GGA). Under GGA the energy due to a point of electron density depends not only on the value of the electron density at that point, but also the rate of change of electron density at that point (the gradient). The third rung is for meta-GGA, which include modifications to the kinetic energy operator. The fourth rung are hybrid functionals. Hybrid functionals mix the energy term of GGA functionals with another source of electron exchange, typically from the HF energy calculated using Kohn-Sham orbitals. The hybrid functionals used in this thesis are the Minnesota functionals M06-2X<sup>39</sup> and M08-HX<sup>40</sup> by the Truhlar group, B3LYP<sup>41</sup>, BLYP<sup>42, 43</sup>, and B3P86<sup>42,44</sup> which include exchange operator derived by Becke with added correlation, as well as the mPW1PW91<sup>45,46</sup> and PBE1<sup>46</sup> functionals. The fifth, and current highest rung on Jacob's ladder is for nonlocal functionals. Nonlocal functionals are not only dependent on occupied orbitals but also unoccupied and incorporate electron correlation energy from WFT. This includes functionals such as B2PLYP<sup>47</sup>, which includes energy from an MP2 calculation using Kohn-Sham orbitals for the MP corrections.<sup>16</sup>

**Basis Sets.** Modern basis sets are created by utilizing Gaussian functions to approximate atomic orbitals. The general form of a Gaussian function is as follows:

$$f(x) = ae^{\frac{-(x-b)^2}{c}} \quad (27)$$

Traditional atomic orbitals (s, p, d, etc.) are atom-centered. However only including atom-centered basis functions does not provide enough flexibility to describe systems with highly diffuse electron densities. To more accurately describe these systems diffuse functions may be added, which are basis functions that are not atom-centered. This allows a greater flexibility to describe systems with negative charges or otherwise loosely held electrons.

Another limitation to using only traditional orbitals is that alone they do not result in accurate molecular geometries even for systems without diffuse electrons. To improve molecular geometries polarization functions are added, or functions of higher angular momentum quantum numbers than the valence orbitals of each atom. In practice this indicates inclusion of p-orbital functions for hydrogen, and d-orbital functions for Li through Ne.

The basis sets utilized in the present work are those designed by Pople and Dunning<sup>16</sup>, Sadlej's polarized triple zeta basis set (SPVTZ)<sup>48</sup>, and the def2-TZVP basis set.<sup>49</sup> The basis sets designed by Pople follow a specific nomenclature for specification of polarization and diffuse functions. For example the Pople basis set 6-311+G(d,p) includes diffuse functions (“+”) on non-hydrogen atoms (“++” would indicate diffuse functions also on hydrogen atoms) and polarization functions on all atoms (d-orbitals for non-hydrogen atoms and p-orbitals for hydrogen atoms). The Dunning basis sets used in this thesis always include polarization functions and follow the nomenclature “aug-cc-

pVNZ". The optional "aug-" portion indicates augmentation with diffuse functions and N indicates the number of Gaussian functions which are used to build each atomic orbital (D – double, T – Triple, etc.). SPVTZ and def2-TZVP both include diffuse and polarization functions in their specifications.

### Calculating an NMR Spectrum

To calculate a chemical shift a mathematical expression that is amenable to calculation must be derived. The energy of a nucleus in an external magnetic field can be determined by the following equation:

$$E = B\sigma h\gamma I \quad (28)$$

In which  $E$  is the energy of the nucleus due to its interaction with an external magnetic field,  $B$  is the magnitude of the external magnetic field at a nucleus,  $\sigma$  is the reduction in the external magnetic field due to nearby electrons (effectively a measure of chemical shift),  $\gamma$  is the gyromagnetic ratio of the nucleus,  $I$  is the angular momentum in the direction of the magnetic field, and  $h$  is the Planck constant. Differentiation of energy with respect to the external magnetic field, then by the magnetic moment of the nucleus of interest ( $h\gamma I$ ) yields the following equation from which shielding can be determined<sup>50</sup>.

$$\sigma = \frac{d^2E}{dBd(h\gamma I)} \quad (29)$$

This equation provides the basis for computation of chemical shifts through the local reduction in the magnetic field, and is accessible to calculation for model chemistries in which analytic second-derivatives are available.

For certain model chemistries chemical shifts can be calculated analytically. For others dependence of results on the coordinate system makes calculation unfeasible.

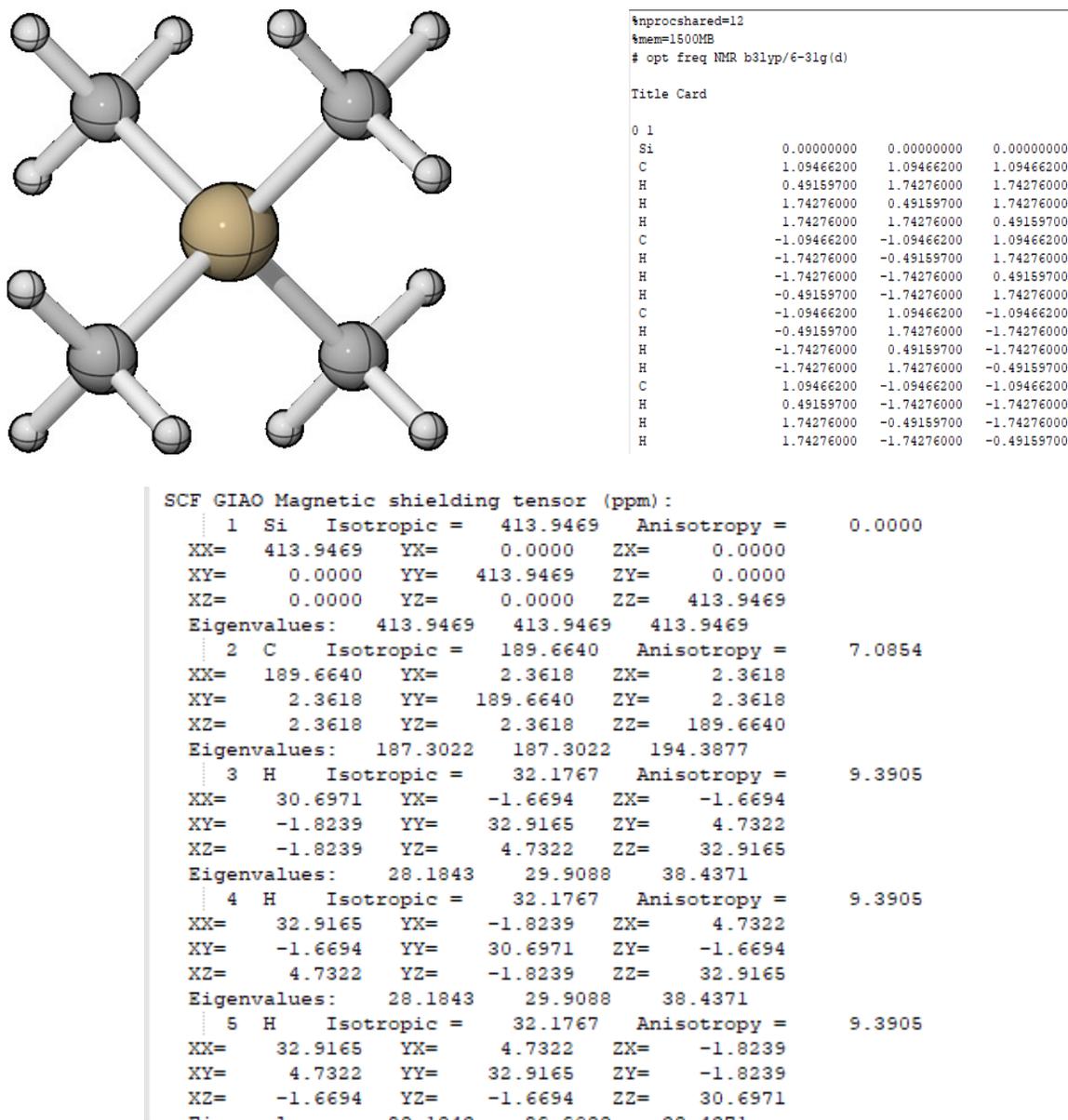
Model chemistries for which analytic calculation of chemical shifts are available include DFT, HF, and MPn methods. Some other commonly used methods, such as semi-empirical calculations, require numerical solutions. A numerical solution to the above equation requires defining a coordinate origin, or gauge origin, from which the magnetic moments can be calculated. While analytic results are invariant with respect to the gauge origin, numerical methods are not. The process of using analytic derivatives for the calculation of shielding tensors is referred to as a gauge-independent atomic orbital method (GIAO).<sup>50</sup>

The process of calculating NMR shifts in the Gaussian16<sup>14</sup> program involves three major steps: Optimization of geometries for the molecule of interest and reference molecule, calculation of magnetic shielding tensors for each, and comparison. The first step in calculating a chemical shift in Gaussian is to optimize the geometry of the molecule of interest. Optimizations are a routine procedure in Gaussian followed by a frequency calculation to verify the process has arrived at a minimum energy configuration. It is typically not necessary to use an expensive model chemistry for this step; accurate geometries are typically easy to arrive at. It is from this stationary geometry that an NMR calculation may be performed.

The output of an NMR calculation is the isotropic shielding, anisotropy, and, if specified, coupling constants and can be seen in Figure 14. Isotropic values are always positive and are a measure of the degree of reduction of the experienced magnetic field by a nucleus. The predicted chemical shift is related to the isotropic shielding value calculated in equation 27 by the following equation:

$$\sigma = \sigma_{Ref} - \sigma_M \quad (30)$$

In which  $\sigma$  is the chemical shift,  $\sigma_{Ref}$  is the calculated isotropic shielding value for the reference nucleus, and  $\sigma_M$  is the calculated isotropic shielding value for the nucleus of interest. The anisotropy is also reported and is a measure of the change in shielding as a result of changes in orientation of the molecule with respect to the magnetic field.

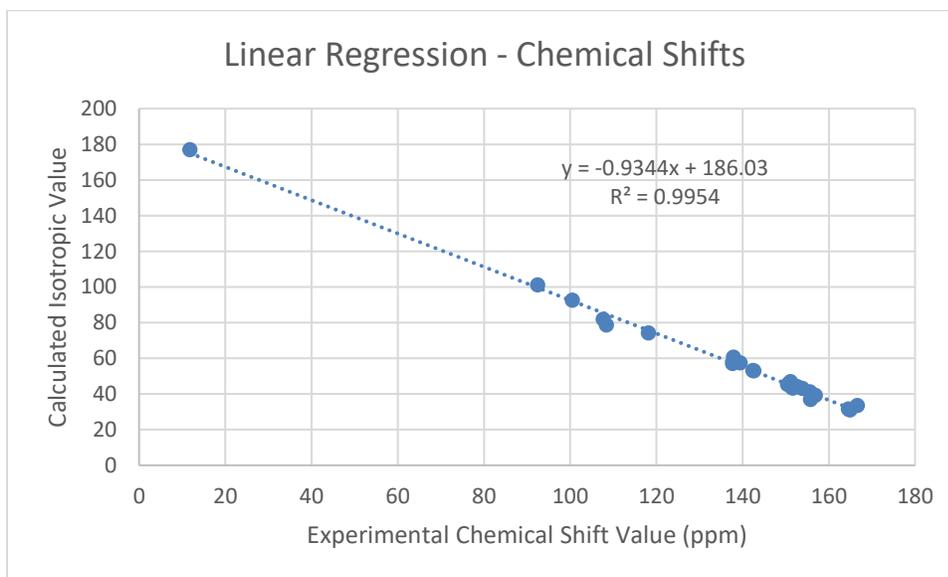


**Figure 14:** Top Left: Input structure for tetramethylsilane (TMS). Top Right: Input file text for geometry optimization followed by a frequency calculation, then an NMR calculation on the optimized structure. Geometry listed in the input file is the geometry for the molecule pictured in Top Left. Bottom: Output text from NMR calculation on optimized geometry.

An alternative method to calculating a chemical shift involves using several reference compounds, as opposed to only using TMS as a reference. This is accomplished by experimentally determining the chemical shifts for several reference compounds and building a plot of these values versus the calculated isotropic shift values. A linear regression can then be performed as in Figure 15 and chemical shifts calculated according to the following equation<sup>51</sup>:

$$\sigma = \frac{\text{Intercept} - \sigma_M}{-\text{slope}} \quad (31)$$

In which *Intercept* is the y-intercept calculated by linear regression and *-slope* is the negative of the slope calculated by linear regression. Using this method, systematic error for a particular model chemistry can be reduced, and potential errors that are specific to a single reference compound are minimized.



**Figure 15:** Example plot of linear regression method of calculating chemical shifts. These linear regression data can then be applied to calculation of other similar systems for which the structure is unknown or misassigned.

Due to the long relaxation times typical of an NMR experiment it is usually adequate to perform an NMR calculation only using the optimized geometry; however, when isomerization must be considered, results can be improved by inclusion of a Boltzmann weighting factor. To do this an NMR spectrum for each state is calculated and averaged using a Boltzmann function.

$$\sigma = \frac{\sum \sigma_n e^{\frac{-E_n}{kT}}}{\sum e^{\frac{-E_n}{kT}}} \quad (32)$$

The above equation is used to calculate the Boltzmann weighted chemical shift in which  $\sigma_n$  is the calculated chemical shift value of a nucleus for a particular isomer and  $E_n$  is the calculated energy of that isomer. The chemical shift for each nucleus in a molecule must be calculated this way and averaged.<sup>51</sup>

### Calculation and Interpretation of Vibrational Spectra

A frequency calculation is the process of determining the energy change and atomic displacements for a molecule's fundamental vibrational modes. This procedure is routine to most computational chemistry work as it is required to confirm the success of reaching a minimum energy structure during an optimization job.

A frequency calculation is performed by differentiating energy twice with respect to the atomic coordinates. This procedure can be understood when considering a simple harmonic oscillator. The potential energy of a simple harmonic oscillator is described by the following equations:

$$U = \frac{1}{2} kx^2 \quad (33)$$

$$\frac{d^2U}{dx^2} = k \quad (34)$$

In which  $U$  is the potential energy,  $k$  is the spring constant, and  $x$  is the displacement from equilibrium. Differentiating potential energy twice with respect to the displacement yields the force constant. This procedure can be generalized to larger systems with multiple degrees of freedom by the use of a Hessian matrix:

$$HE(x_1, x_2 \dots x_N) = \begin{bmatrix} \frac{\delta^2 E}{\delta x_1^2} & \dots & \frac{\delta^2 E}{\delta x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\delta^2 E}{\delta x_n x_1} & \dots & \frac{\delta^2 E}{\delta x_n^2} \end{bmatrix} \quad (35)$$

In which  $H$  specifies the hessian of the following function,  $x_1$  to  $x_n$  are the degrees of freedom for the system and  $E$  is any scalar function. In the case of a frequency calculation there are  $3N$  degrees of freedom representing displacement of each atom along each X, Y, or Z coordinate, and  $E$  is the total potential energy of the system. This expression can be used to derive information about the fundamental vibrational modes of a molecule.

The output of a frequency calculation using the Gaussian 16 program are the frequencies, reduced masses, force constants, and relative atomic displacements for each of the  $3N - 6$  fundamental modes of the molecule. The format of these outputs is summarized in Figure 16. The frequency value, given in wavenumbers, can be converted to the energy value of the fundamental mode. Each of the atomic displacements are given as unitless relative values. These values can then be used in further calculations as needed.<sup>52</sup>

There are two methods of interpreting calculated vibrational modes. The first is visual inspection using a program to display the atomic motions. The second is using potential energy density (PED) analysis. The former method is useful for a qualitative

understanding of atomic motions but can be misleading. Small mass atoms such as hydrogen will appear to contribute significantly to certain modes due to their large amplitude even when their energy contribution is negligible. For this reason, a quantitative analysis is necessary to gain a full understanding of molecular vibrations. A PED analysis provides a quantitative analysis of vibrational modes by determining the energy change due to each changing normal coordinate. The normal coordinate with the highest energy change, regardless of amplitude, are considered to contribute more to the overall vibrational characteristic.<sup>53</sup>

1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										
36										
37										
38										
39										
40										
41										
42										
43										
44										
45										
46										

**Figure 16:** Output from Gaussian following a frequency calculation reported to high precision and including Raman activity results. Specified by the “freq=(raman, hpmodes)” command in Gaussian<sup>14</sup>. Frequency values are in units of wavenumbers, reduced masses in units of amu, and force constants in units of mDyne/Angstrom.

The Vibrational Energy Density Analysis (VEDA) program can be used to perform a PED analysis on results obtained from a frequency calculation in Gaussian. Use of the VEDA program is fairly straightforward. During the frequency calculation the checkpoint (.chk) file should be retained. The checkpoint file should then be formatted by the Gaussian program's formchk function and output as a file with the extension ".fmu". The VEDA program can then open the ".fmu" file to obtain the atomic coordinates and force constants. After opening the ".fmu" file "Create .DD2" will create a normal coordinate set that can be used to describe vibrational motion. A PED analysis can be done using the internal coordinates or using internal coordinates that have been mixed into complex coordinates. Complex coordinates are created from mixing multiple normal coordinates into a single motion and can improve the overall fit of the PED analysis, although with the cost of making the interpretation of results a bit more difficult. After mixing or not mixing the normal coordinates the energy change due to each complex coordinate can be determined. Whether the mixing of normal coordinates into complex coordinates improves the PED analysis or not can be measured by an EPM parameter, which is printed in the output files after optimization. The best results are those which maximize this EPM parameter.<sup>53</sup> The difference between complex and normal coordinates can be seen in Figure 17, in which each coordinate (s 1, s 2, etc.) is described. The type of coordinate change (stretch, bend, tors) is listed, followed by how the mode changes (ex. 1.00 for lengthening, -1.00 for contracting), then the atom numbers and types involved in the mode.

Calculation of infrared and Raman intensities can be performed analytically for HF, DFT, and MPn methods. As with other spectroscopic methods both of these

intensities are expressed as a derivative of molecular electronic energy, with vibrational activities specifically being energy derivatives with respect to electric field directions and nuclear coordinates.

Average max. Potential Energy <EPm> = 33.258							Average max. Potential Energy <EPm> = 26.457						
TED Above 100 Factor TAF=0.000							TED Above 100 Factor TAF=0.000						
Average coordinate population 1.700							Average coordinate population 2.067						
Most complex coordinate No. 9 , population = 4							Most complex coordinate No. 8 , population = 4						
s 1	1.00	STRE	11	6	NH	1.013891	s 1	1.00	STRE	11	6	NH	1.013891
s 2	1.00	STRE	12	5	NH	1.010505	s 2	-1.00		12	5	NH	1.010505
s 3	1.00	STRE	3	7	CH	1.081379	s 3	1.00	STRE	11	6	NH	1.013891
s 4	1.00	STRE	4	8	CH	1.084816	s 4	1.00		12	5	NH	1.010505
s 5	1.00	STRE	9	1	OC	1.216531	s 5	1.00	STRE	3	7	CH	1.081379
	-1.00		10	2	OC	1.219350	s 6	-1.00		4	8	CH	1.084816
s 6	1.00	STRE	10	2	OC	1.219350	s 7	1.00	STRE	3	7	CH	1.081379
	1.00		9	1	OC	1.216531	s 8	1.00		4	8	CH	1.084816
s 7	1.00	STRE	12	1	NC	1.395540	s 9	1.00	STRE	9	1	OC	1.216531
	-1.00		12	4	NC	1.375798	s 10	-1.00		10	2	OC	1.219350
s 8	1.00	STRE	12	1	NC	1.395540	s 11	1.00	STRE	9	1	OC	1.216531
	-1.00		3	4	CC	1.350160	s 12	1.00		10	2	OC	1.219350
s 9	-1.00	STRE	3	4	CC	1.350160	s 13	1.00	STRE	3	4	CC	1.350160
	1.00		11	1	NC	1.385004	s 14	1.00	STRE	11	1	NC	1.385004
	1.00		11	2	NC	1.413789	s 15	-1.00		11	2	NC	1.413789
	-1.00		12	4	NC	1.375798	s 16	-1.00		12	1	NC	1.395540
s 10	-1.00	STRE	3	4	CC	1.350160	s 17	1.00		12	4	NC	1.375798
	-1.00		11	2	NC	1.413789	s 18	1.00	STRE	11	1	NC	1.385004
	1.00		12	1	NC	1.395540	s 19	-1.00		11	2	NC	1.413789
	-1.00		12	4	NC	1.375798	s 20	1.00		12	1	NC	1.395540
s 11	1.00	STRE	11	1	NC	1.385004	s 21	-1.00		12	4	NC	1.375798
	-1.00		11	2	NC	1.413789	s 22	-1.00	STRE	11	1	NC	1.385004
	1.00		12	1	NC	1.395540	s 23	-1.00		11	2	NC	1.413789

**Figure 17:** Sample output from VEDA program using both mixed (right) and unmixed (left) modes. Coordinates listed are combined to describe normal modes of a molecule.

The dipole of a molecule is expressed in terms of energy by considering a molecule in a weak electric field. The energy of a dipole in an electric field can be calculated using the following equation:

$$E = \mu\epsilon_f \quad (36)$$

In which  $E$  is the energy of the dipole,  $\mu$  is the magnitude of dipole, and  $\epsilon_f$  is an electric field in the  $f$  direction. Differentiation with respect to an electric field yields the following:

$$\mu = \frac{dE}{d\epsilon_f} \quad (37)$$

Calculation of an infrared intensity for a particular vibration then requires finding the change in molecular dipole as the nuclear coordinates ( $Q$ ) change:

$$I_{IR} \propto \frac{d\mu}{dQ} = \frac{d^2E}{d\epsilon_f dQ} \quad (38)$$

It is known that the relationship between molecular polarizability and the dipole change in an external electric field is as follows:

$$\mu = \alpha\epsilon_g \quad (39)$$

Differentiation allows the equation to be rearranged to find the molecular polarizability as a function of the induced dipole and electric field:

$$\alpha_{f,g} = \frac{d\mu}{d\epsilon_g d\epsilon_f} \quad (40)$$

The calculation of Raman intensities is dependent on the change in molecular polarizability with respect to the change in nuclear coordinates:

$$I_{Raman} \propto \left( \frac{d\alpha}{dQ} \right)^2 \quad (41)$$

Using the above equations, a relationship between Raman activity and energy derivatives can be found.

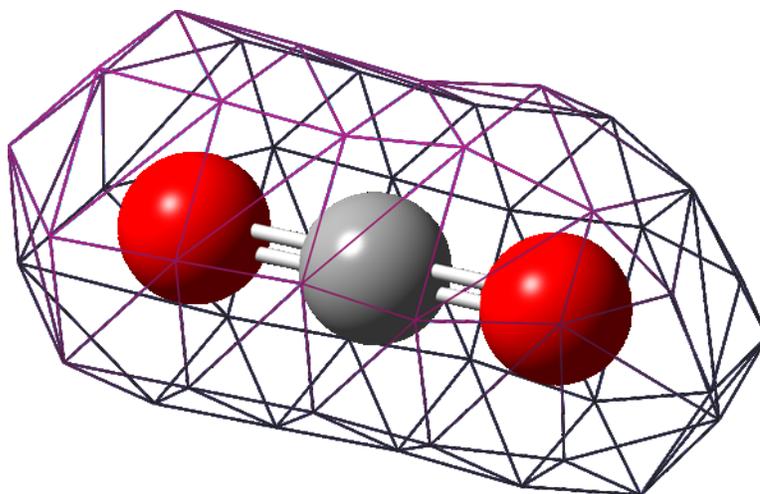
$$I_{Raman} \propto \left( \frac{d^3E}{dQ d\epsilon_f d\epsilon_g} \right)^2 \quad (42)$$

Each of the nine tensor components of polarizability can be calculated by varying the direction of the electric field.<sup>54, 55</sup>

### **Calculating Electronic Volume Changes**

Calculation of electronic volume has been implemented in the Wave Function Analysis – Surface Analysis Suite (WFA-SAS) program.<sup>37</sup> First, a square grid is placed over the molecule and the electron density is evaluated at each point. The cubes along the grid that intersect the isodensity surface are kept and the rest are discarded. The isodensity surface contained within each cube is then modeled as a number of triangles with touching edges. After the entire surface has been rendered by triangular sections the enclosed volume can be calculated as the sum of all internal cubes plus the portion of the cubes along the surface that fall inside the specified value. A surface rendered with a low number of points in the grid for visualization can be found in Figure 18. More details on the algorithm can be found in the paper describing the WFA-SAS program's capabilities.<sup>37</sup>

The WFA-SAS program is also capable of evaluating average local ionization energy (ALIE). Using the grid of points previously established to define an isodensity surface the ALIE of a molecule can be calculated using points that fall within that surface.



**Figure 18:** Electronic density surface rendered with a low number of triangles to easy visualization.

Calculating an electronic volume change during a vibration requires defining a meaningful vibrational amplitude. It seems logical, when considering a vibration as harmonic motion, to use the classical turning point of a spring as a starting point and vary the displacement in multiples. The turning point of a classical harmonic oscillator can be derived as the following:

$$100hcv = \frac{1}{2}kx^2 \quad (43)$$

$$x = \sqrt{\frac{200hcv}{k}} \quad (44)$$

In which  $\nu$  is the vibration frequency in wavenumbers,  $c$  is the speed of light,  $h$  is the Planck constant, and  $k$  is the spring constant. To move each atom in the molecule the turning point value should be multiplied by each atom's Cartesian displacements for each axis returned by Gaussian after a frequency calculation.

## CHAPTER 4: RESULTS AND DISCUSSION

### NMR of Nucleobases

**Published NMR Data.** The ultimate goal of the NMR prediction portion of this thesis is to identify a model chemistry allows for accurate prediction of 2-dimensional spectra for the Drew-Dickerson dodecamer.

Towards this end it seemed appropriate to begin by modeling the individual canonical nucleotides adenine, guanine, cytosine, and thymine. Uracil was included in addition to the DNA specific nucleotides previously listed in hopes the data can be generalized to RNA molecules in the future after our selected DNA system has been adequately modeled.

There were several sources available for NMR spectra in dimethyl sulfoxide (DMSO) for all of the nucleotides, with the exception of guanine. Only one  $^{13}\text{C}$  experimental spectrum for guanine in DMSO has been reported, and no  $^1\text{H}$  spectrum. Guanine spectra in  $\text{D}_2\text{O}$  generally agree with each other to the same extent of other reported chemical shift values.

There is some deviation expected within reported experimental results. It can be seen from Tables 1-5 that the variation among experimental sources for these compounds are within 2 ppm for carbon chemical shifts, and within 0.1 ppm, but occasionally as high as 0.2 ppm, for proton chemical shifts.

There are, in some cases, large deviations in reported chemical shifts. Most notably H8 of adenine in Table 2. It can be seen the three reported values are 6.67 ppm, 7.09 ppm, and 7.298 ppm. Additionally, the chemical shifts for H1 and  $\text{NH}_2$  protons of

Table 1: Experimental  $^{13}\text{C}$  chemical shifts (ppm) obtained from chemical literature. Solvents are DMSO unless denoted with an asterisk, in which case the solvent is D<sub>2</sub>O.

Reference	Molecule	C2	C4	C5	C6	C8/CH <sub>3</sub>
56	G	153.80	150.45	108.42	155.74	137.87
57	G*	160	162.2	119.6	168.8	150.1
58	G*	161.905	163.681	120.987	170.388	151.317
59	A	152.4	151.3	117.5	155.3	139.3
60	A	153.41	151.71	119.07	156.36	140.3
61	A	152.2	151.1	117.3	155.1	139.2
62	A	152.5	150.3	118.5	156.0	138.9
63	A	152.4	151.2	118.3	155.1	139.3
64	C	157.77	167.49	93.35	92.47	X
65	C	156.89	166.62	92.47	142.52	X
66	C	156.63	166.51	92.32	142.46	X
67	C	156.0	165.7	91.6	141.9	X
68	T	151.46	164.87	107.33	137.63	11.72
69	T	151.49	164.93	107.68	167.72	11.79
70	T	151.5	165.0	107.7	Not Rep.	11.8
71	U	152.27	165.09	101.01	142.89	X
69	U	151.39	164.2	100.11	142.07	X
70	U	151.5	164.4	100.3	142.2	X

cytosine have been reported as deviating from the other reported values by 3.1 ppm, and 3.5 ppm, respectively. For the purpose of comparison to calculated chemical shifts, the

two closest values were averaged and used for comparison. In the case of H8 and NH2 of cytosine, the deviant values were discarded in favor of the two sources that agree to within 0.02 ppm, and 0.03 ppm, respectively.

Table 2: Adenine <sup>1</sup>H chemical shifts (ppm) obtained from chemical literature. Solvents are DMSO unless otherwise noted.

Reference	H2	H8	H9
74 – CDCl <sub>3</sub>	8.138	7.09	8.113
75	8.168	7.298	8.179
76	8.12	6.67	8.03

Table 3: Cytosine <sup>1</sup>H chemical shifts (ppm) obtained from chemical literature. Solvents are DMSO.

Reference	H1	NH2	H5	H6
65	7.15	3.4	5.6	7.35
66	10.29	6.94	5.53	7.28
67	10.27	6.91	5.6	7.35

Table 4: Thymine <sup>1</sup>H chemical shifts (ppm) obtained from chemical literature. Solvents are DMSO.

Reference	H1	H3	CH <sub>3</sub>	H6
69	10.98	10.57	1.72	7.24
70	11.02	10.60	1.71	7.24
77	11.0	10.6	1.752	7.276

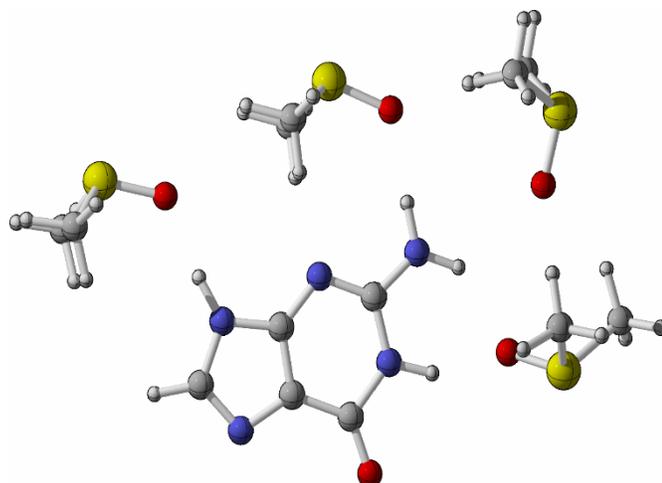
Table 5: Uracil  $^1\text{H}$  chemical shifts (ppm) obtained from chemical literature. Solvents are DMSO.

Reference	H1	H3	H5	H6
70	10.83	11.03	5.45	7.38
78	10.82	11.02	5.473	7.406
69	10.8	11.0	5.45	7.39

### Variable Optimization Model Chemistry, Constant NMR Calculation Model

**Chemistry.** It is generally accepted that highly expensive model chemistries are not required to achieve accurate geometry optimizations.<sup>35</sup> However, the ultimate goal of this project is to optimize and calculate an NMR spectrum for the Drew-Dickerson dodecamer, a system consisting of 494 non-hydrogen atoms without including any explicit solvent molecules or ions. Utilizing the 6-31G(d) basis set, which can reasonably be considered to be a minimal basis set to obtain an accurate result, this system ends up having over eight thousand basis functions! For comparison the guanine molecule with explicit DMSO in Figure 19, a system of only 479 basis functions, required nearly two and a half days of processor time to optimize. Assuming this method scales as  $N^3$  in which  $N$  is the number of basis functions (a low estimate) this would imply our target system may take as long as 28 years of processor time to fully optimize! It is clear that no shortcut should go unconsidered when attempting calculations on this scale.

To this end, each of the nucleotides considered in this study (adenine, thymine, cytosine, guanine, and uracil) were optimized using various levels of theory, then an NMR calculation was performed at the B3LYP/6-31G(d) level of theory. The optimization model chemistries tested were the semi-empirical methods AM1, PM3, and



**Figure 19:** Optimized structure of guanine with four DMSO molecules.

PM6, and each combination of the functionals MP2, B2PLYP, B3LYP, PBE1, mPW1PW91, and HF with each of the basis sets 6-31G(d), 6-311+G(2d,p), def2-TZVP, SPVTZ, and aug-cc-pVDZ. The calculated isotropic values for each nucleus were then averaged. The difference between the maximum calculated value for each nucleus and minimum calculated value for each nucleus was then calculated. The model chemistry that deviated from the average the most was eliminated and the maximum, minimum, and average recalculated. This procedure was repeated until the spread of results (maximum value for each nucleus minus minimum value) was within five percent of the average. The model chemistries removed during this process can be found in Table 6.

The most notable model chemistries eliminated are all of the semi-empirical methods. This is expected, as they are the least accurate model chemistry used in this test. Another interesting result is that even while using comprehensive basis sets such as aug-cc-pVDZ, HF failed for every base. Another interesting trend is that Sadlej's PVTZ basis set was inconsistent for three bases using MP2 and B2PLYP, the latter of which

Table 6: Model chemistries excluded as candidates for optimization of larger nucleic acid systems.

Functional	Basis Set	Bases Failed
AM1	Included	A G C T U
PM3	Included	A G C T U
PM6	Included	A G C T U
MP2	SPVTZ	A G C
MP2	6-31G(d)	C
MP2	aug-cc-pVTZ	C
B2PLYP	SPVTZ	A G C
B3LYP	def2-TZVP	A
PBE1	SPVTZ	G
mPW1PW91	SPVTZ	G
Hartree-Fock	6-31G(d)	A G C T U
Hartree-Fock	SPVTZ	A G C T U
Hartree-Fock	6-311+G(2d,p)	A G C T U
Hartree-Fock	def2-TZVP	A G C T U
Hartree-Fock	aug-cc-pVTZ	A G C T U

utilizes an MP2 calculation to include correlation. Sadlej’s PVTZ was also inconsistent for guanine under the mPW1PW91 and PBE1 functionals. These results clearly indicate that semi-empirical methods and HF will likely not be sufficient to optimize the Drew-Dickerson dodecamer, and neither will Sadlej’s PVTZ basis set

An interesting result is that 6-31G(d) was only found to be inconsistent with the other methods under MP2 calculations, and even then, only for cytosine. This makes this small basis set an attractive option for optimizing the Drew-Dickerson dodecamer as it will take the least computational resources among all of the tested basis sets. The model chemistry B3LYP / 6-31G(d), being a very cheap method that was shown to be consistent with higher levels of theory, was selected for optimizations going forward.

**Evaluating NMR Model Chemistries for Accuracy.** Having selected a model chemistry for optimizations, the next step is determining which model chemistries are appropriate for NMR calculations. Each of the five nucleotides were optimized at the B3LYP/6-31G(d) level of theory. Using this geometry an NMR calculation was performed utilizing each combination of the B3LYP, B3P86, BLYP, M08HX, M06-2X, mPW1PW91, and PBE1PBE functionals with the 6-31G(d), 6-31+G(d,p), 6-311+G(2d,p), cc-pVDZ and aug-cc-pVDZ basis sets. A linear regression was then performed using the calculated isotropic shielding values and experimental NMR chemical shifts, and chemical shifts calculated using this linear regression. The calculated chemical shift values were then compared to the published literature values to determine which nuclei were

NMR calculations were initially performed using only implicit solvation in DMSO utilizing the PCM<sup>79</sup> solvation model in Gaussian16.<sup>14</sup> Table 7 shows the linear regression data for all proton chemical shifts using the B3LYP and B3P86 functionals with no explicit solvent molecules. These results are representative of all functionals, with the highest R<sup>2</sup> value among all model chemistries being 0.66. However, after removing exchangeable protons, which are bonded to nitrogen and can then be

exchanged with protic solvents, the fit drastically improves as can be seen in Table 8. The lowest  $R^2$  value for any model chemistry when analyzing only non-exchangeable protons is 0.988. This demonstrates that there is a clear problem in describing the exchangeable protons using only implicit solvation.

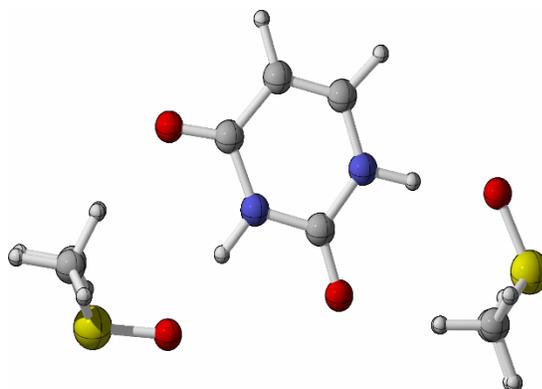
Table 9 demonstrates that model chemistries that are good at predicting  $^1\text{H}$  chemical shifts also tend to be good at predicting  $^{13}\text{C}$  chemical shifts. The errors for each  $^{13}\text{C}$  and  $^1\text{H}$  nucleus, excluding exchangeable protons, is summarized in Table 10. It can be seen that most nuclei are predicted to within experimental error, defined as 2.0 ppm for  $^{13}\text{C}$  and 0.20 ppm for  $^1\text{H}$ . Where there are significant errors, such as C5 and C8 of guanine, C5 of thymine, and H2 of adenine, the errors are always in the same direction and of roughly equal magnitude for all model chemistries. This is fortunate as if these errors are known to occur for all model chemistries they can be easily corrected after the calculation.

In an attempt to more accurately describe the system, calculations were performed using a minimum number of explicit solvent molecules. Since DMSO has an available hydrogen bond acceptor, the minimum number of explicit solvent was considered to be one DMSO molecule to accept each available hydrogen bond. The explicit DMSO molecules were added to the previously optimized structures and an optimization was performed followed by a frequency calculation to ensure a minimum energy had been achieved. The B3LYP/6-31G(d) optimized structures are displayed in Figures 19 through 23.

Table 7: Linear regression data for nucleobase  $^1\text{H}$  chemical shifts predicted using B3LYP and B3P86 functionals.

Functional	B3LYP				
Basis Set	6-31+G(d,p)	6-31G(d)	6-311+G(2d,p)	aug-cc-pVDZ	cc-pVDZ
$R^2$	0.64	0.46	0.59	0.62	0.53
Y-Intercept	29.32	29.77	29.44	29.23	29.26
Slope	-0.52	-0.43	-0.51	-0.53	-0.47
Functional	B3P86				
Basis Set	6-31+G(d,p)	6-31G(d)	6-311+G(2d,p)	aug-cc-pVDZ	cc-pVDZ
$R^2$	0.64	0.47	0.59	0.63	0.53
Y-Intercept	29.21	29.62	29.32	29.14	29.15
Slope	-0.52	-0.44	-0.51	-0.53	-0.47

The  $R^2$  values for all proton chemical shifts using explicit solvation is summarized in Table 11, and errors in Table 12. By including explicit solvent molecules, the correlation drastically increases. This makes sense as inclusion of a hydrogen bond acceptor would increase the electron density around the hydrogen nucleus and significantly alter its NMR signal. It is also interesting that the smallest basis set tested was the best performer among all basis sets after including explicit solvation. As with the results for only implicit solvation, most nuclei were predicted within experimental error, and the results that deviated significantly from experimental results did so in a consistent way for all model chemistries with a few exceptions, including C2 and C5 of uracil,  $\text{CH}_3$  and C2 of thymine, and C5 of cytosine. These exceptions have shown excellent prediction using some model chemistries but are outside of experimental error with



**Figure 20:** Optimized structure of uracil with two DMSO molecules.

Table 8: Linear regression data for nucleobase non-exchangeable  $^1\text{H}$  chemical shifts predicted using B3LYP and B3P86 functionals and implicit solvation.

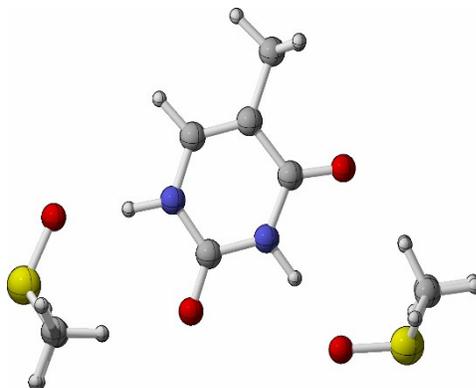
Functional	B3LYP				
Basis Set	6-31+G(d,p)	6-31G(d)	6-311+G(2d,p)	aug-cc-pVDZ	cc-pVDZ
$R^2$	0.996	0.990	0.995	0.995	0.989
Y-Intercept	31.39	32.15	31.60	31.39	31.54
Slope	-0.97	-0.96	-0.99	-1.00	-0.97
Functional	B3P86				
Basis Set	6-31+G(d,p)	6-31G(d)	6-311+G(2d,p)	aug-cc-pVDZ	cc-pVDZ
$R^2$	0.996	0.991	0.995	0.995	0.990
Y-Intercept	31.29	32.03	31.48	31.30	31.46
Slope	-0.98	-0.97	-0.99	-1.00	-0.98

others. However, they are not so far out of experimental error that predictions for these nuclei using all model chemistries cannot be brought to within experimental error by adding or subtracting a constant specific to each nucleus and each nucleotide. This would, however, be less ideal than finding a method without this systematic error.

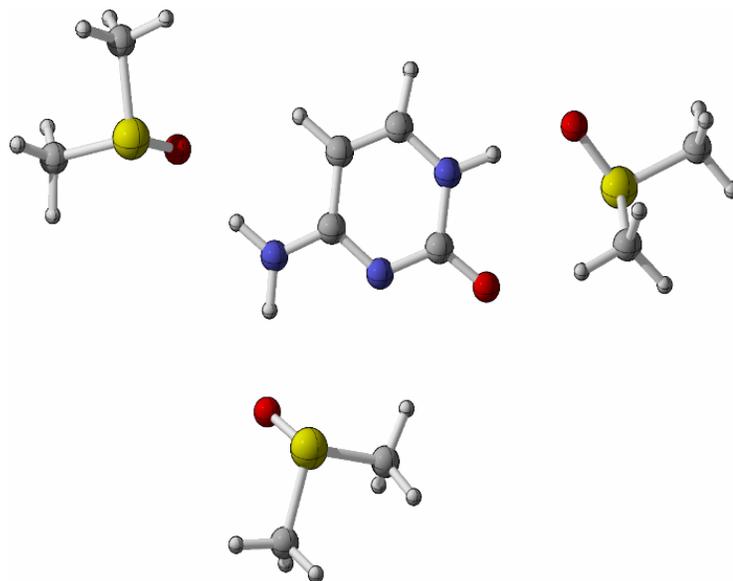
Table 9: Top and Bottom performing model chemistries using implicit solvation as judged by non-exchangeable  $^1\text{H}$  chemical shift linear regression.

	Functional	Basis Set	$R^2$ $^1\text{H}$	$R^2$ $^{13}\text{C}$
Top-Performing	B3LYP	6-31+G(d,p)	0.996	0.993
	mPW1PW91	6-31+G(d,p)	0.996	0.994
	PBE1PBE	6-31+G(d,p)	0.996	0.994
	B3P86	6-31+G(d,p)	0.996	0.993
	B3LYP	6-311+G(2d,p)	0.995	0.995
Bottom-Performing	M062X	cc-pVDZ	0.989	0.986
	M08HX	cc-pVDZ	0.989	0.982
	B3LYP	cc-pVDZ	0.989	0.991
	BLYP	6-31G(d)	0.987	0.989
	BLYP	cc-pVDZ	0.986	0.987

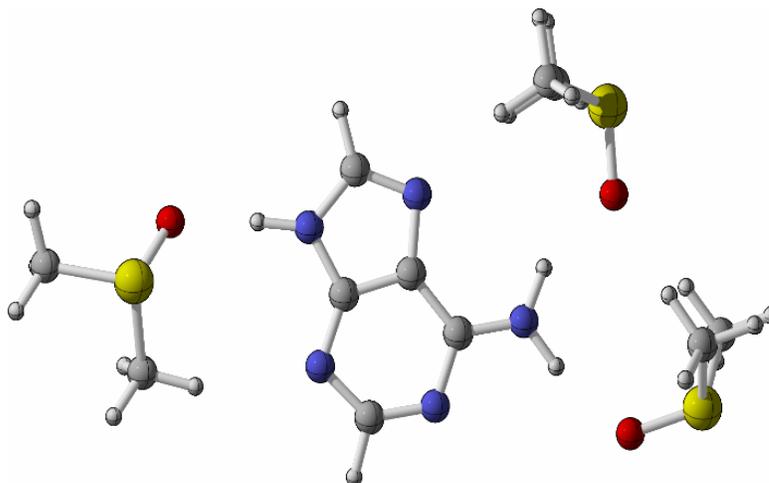
Ultimately, which method is used to model larger systems will depend on which nuclei need to be accurately predicted. If the exchangeable nuclei are of concern then it



**Figure 21:** Optimized structure of thymine with two DMSO molecules.



**Figure 22:** Optimized structure of cytosine with three DMSO molecules.



**Figure 23:** Optimized structure of adenine with three DMSO molecules.

is clearly a requirement that some degree of explicit solvation is included to satisfy all hydrogen bonding requirements. Otherwise, a better fit is achieved by excluding these nuclei from consideration all together.

Table 10: Error for non-exchangeable  $^1\text{H}$  and  $^{13}\text{C}$  nuclei using the five best model chemistries utilizing implicit solvation as judged by  $^1\text{H}$   $R^2$ .

Base	Nucleus	B3LYP 6-31+G(d,p)	mPW1PW91 6-31+G(d,p)	PBE1PBE 6-31+G(d,p)	B3P86 6-31+G(d,p)	B3LYP 6-311+G(2d,p)
Guanine	C8	-3.9	-3.1	-3.2	-3.5	-3.3
	C5	10.1	9.4	9.4	10.2	9.6
	C6	0.8	0.5	0.5	0.5	0.4
	C2	-0.6	-0.7	-0.8	-0.8	-0.8
	C4	1.3	0.7	0.7	1.0	0.8
Adenine	C8	-2.2	-1.2	-1.7	-1.7	-1.7
	C5	0.8	0.2	0.7	0.8	0.7
	C6	0.3	0.2	-0.2	0.1	-0.2
	C2	2.9	3.2	1.4	3.0	1.4
	C4	-0.4	-0.8	-0.8	-0.6	-0.8
	H2	-0.24	-0.24	-0.25	-0.24	-0.25
	H8	0.19	0.17	0.17	0.18	0.23
Cytosine	C5	-3.1	-2.8	-2.7	-2.7	-2.7
	C4	-0.2	-0.4	-0.7	-0.5	-0.7
	C2	-0.8	-1.2	-0.5	-1.2	-0.5
	C6	0.8	1.8	1.0	1.3	1.0
	H5	0.04	0.04	0.04	0.03	0.06
	H6	0.07	0.09	0.09	0.09	0.08
Thymine	C5	3.5	2.9	3.4	3.6	3.4
	C4	-0.5	-0.8	-0.3	-0.7	-0.3
	C2	-1.9	-2.1	-1.1	-2.2	-1.1
	C6	0.0	0.6	0.1	0.3	0.1
	CH <sub>3</sub>	-1.1	-1.0	-0.8	-0.6	-0.8
	H6	-0.07	-0.09	-0.08	-0.08	-0.09
	CH <sub>3</sub>	-0.04	-0.04	-0.04	-0.05	-0.07
Uracil	C5	-0.9	-1.1	-0.8	-1.1	-0.8
	C4	-2.2	-2.4	-1.3	-2.5	-1.3
	C2	-0.3	0.5	0.0	0.2	0.0
	C6	-2.3	-2.3	-2.6	-3.0	-2.6
	H6	0.12	0.15	0.15	0.16	0.12
	H5	-0.07	-0.09	-0.09	-0.09	-0.07

## Volume Changes as a Replacement for Analytic Polarizabilities

**Introduction.** The goal of this experiment was to determine if there is a correlation between electronic volume changes and experimental Raman intensities. To

this end, a set of fourteen vibrations was selected, seven molecules contributing two vibrations each, and the volume changes were calculated at various parameters utilizing various model chemistries. Raman spectra were recorded using the Raman spectrometer at Missouri State University with a laser wavelength of 1064 nm. The ratio of electronic volume changes for each vibration set were then compared with experimentally obtained Raman intensities and a linear regression was performed.

Table 11: Top performing model chemistries using minimum explicit solvation as judged by  $^1\text{H}$  chemical shift linear regression.

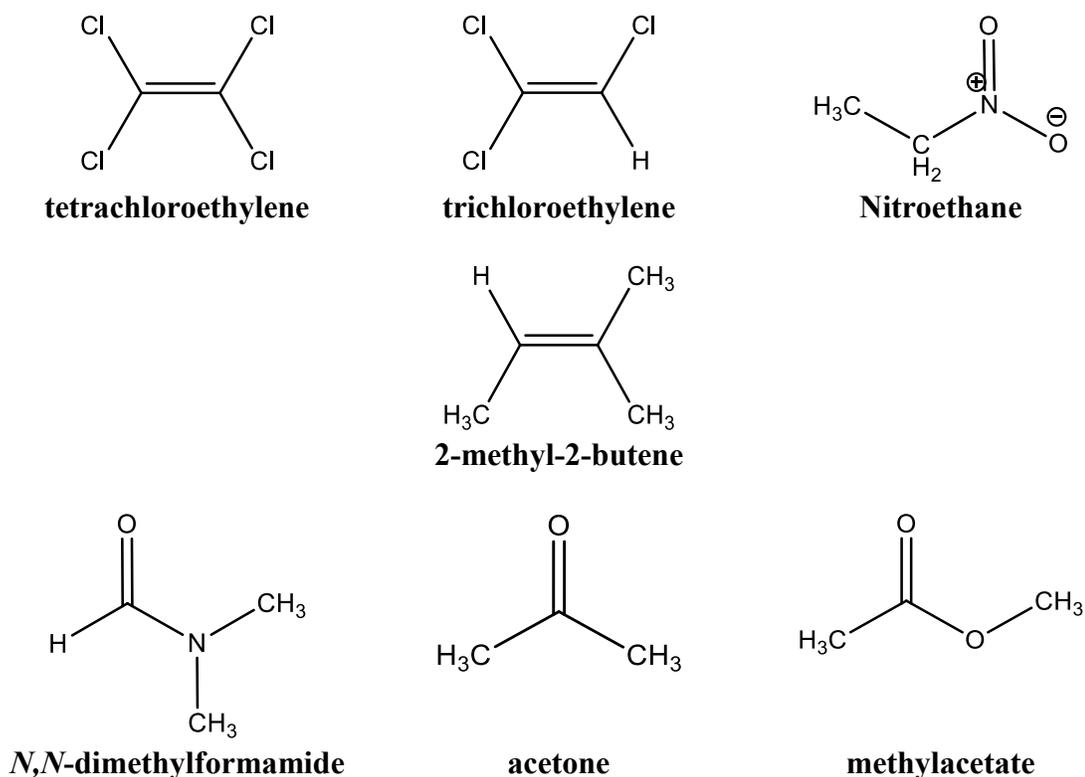
Functional	Basis Set	R <sup>2</sup> Proton	R <sup>2</sup> Carbon
M08HX	6-31G(d)	0.989	0.991
mPW1PW91	6-31G(d)	0.989	0.995
PBE1PBE	6-31G(d)	0.989	0.995
M062X	6-31G(d)	0.988	0.992
B3P86	6-31G(d)	0.988	0.994

**Selecting a Set of Test Molecules.** A set of vibrations was identified that could be used for this experiment. To ensure electronic volumes could be simply described only molecules that do not hydrogen bond were considered. Additionally experimental restrictions necessitated molecules be liquid or solid at 25 °C, and so only liquid samples were used. Additionally, each molecule should contribute at least two vibrations. This is so relative intensities are obtained from the same spectrum. To simplify the process of

Table 12: Error for all  $^1\text{H}$  and  $^{13}\text{C}$  nuclei using the five best model chemistries and explicit solvation, as judged by the proton  $R^2$  value.

Base	Nucleus	M08HX 6-31G(d)	mPW1PW91 6-31G(d)	PBE1PBE 6-31G(d)	M062X 6-31G(d)	B3P86 6-31G(d)
Guanine	C8	-6.4	-5.7	-5.8	-6.6	-6.0
	C5	8.7	8.1	8.1	9.2	8.8
	C6	0.0	0.6	0.6	0.3	0.5
	C2	-2.9	-0.8	-0.8	-1.9	-0.9
	C4	-0.3	0.9	0.8	0.3	1.0
Adenine	C8	-3.0	-2.7	-2.8	-3.4	-3.1
	C5	-0.5	-1.1	-1.1	-0.2	-0.5
	C6	0.1	-0.9	-0.9	0.2	-1.1
	C2	3.7	2.1	2.2	2.4	2.0
	C4	-1.6	-1.1	-1.2	-1.2	-1.0
	H2	-0.31	-0.21	-0.20	-0.21	-0.23
	NH2	0.27	0.30	0.28	0.22	0.30
	H8	0.39	0.45	0.47	0.48	0.42
	H9	-0.60	-0.63	-0.63	-0.67	-0.64
	Cytosine	C5	0.1	-2.4	-2.3	0.0
C4		-0.1	-1.0	-1.0	0.3	-1.1
C2		-2.1	0.2	0.1	-1.5	0.2
C6		2.4	-0.4	-0.4	0.9	-0.8
H5		-0.18	-0.18	-0.17	-0.10	-0.21
H1		-0.29	-0.37	-0.38	-0.32	-0.38
H6		-0.03	-0.11	-0.11	0.00	-0.11
NH2		0.07	0.03	0.00	0.01	0.03
Thymine	C5	2.2	0.9	0.9	2.4	1.6
	C4	0.8	1.3	1.3	0.9	1.3
	C2	-3.4	-0.1	-0.1	-2.4	-0.1
	C6	4.3	2.0	2.1	3.0	1.7
	CH <sub>3</sub>	-3.6	-0.3	-0.4	-3.1	-0.9
	H6	-0.07	-0.05	-0.04	-0.01	-0.07
	H1	-0.01	0.03	0.03	-0.06	0.05
	H3	0.45	0.46	0.46	0.48	0.47
	CH <sub>3</sub>	-0.24	-0.15	-0.16	-0.33	-0.13
Uracil	C5	1.0	-1.5	-1.4	0.4	-1.2
	C4	0.3	1.0	1.0	0.4	1.0
	C2	-3.4	-0.1	-0.1	-2.4	-0.1
	C6	3.7	1.0	1.1	2.0	0.7
	H6	0.04	0.01	0.02	0.07	-0.01
	H1	0.38	0.42	0.41	0.33	0.44
	H3	0.04	0.06	0.05	0.07	0.07
H5	0.09	-0.04	-0.03	0.06	-0.03	

correlating experimental Raman peaks to normal modes each vibration should occur at a characteristic location (e.g. carbonyl stretch  $1700\text{ cm}^{-1}$ ). However, because these are a small fraction of the total peaks in nearly any Raman spectrum, peaks that were well resolved ( $>100\text{ cm}^{-1}$  away from any other peaks) were also used. The set of molecules from which the vibrations were selected can be found in Figure 24, and the selected vibrations in Table 13. The relevant portions of the experimental Raman spectra are available in Appendix A.



**Figure 24:** Seven molecule test set to calculate volume changes of vibrational modes.

**Optimizable Parameters.** This experiment has several parameters that must be defined. These include the model chemistry used for the optimization and frequency calculations, the electron density at which the surface is defined, and the degree of

displacement along the vibrational coordinates. It is important to understand how each will affect volume measurements so that a reasonable scope for this approach can be defined.

The first factor, model chemistry, was relatively easy to choose. In the paper by Politzer<sup>13</sup> it was found that good correlations between equilibrium volumes and material polarizabilities were achieved using the model chemistry HF/6-31G(d). This is a very inexpensive model chemistry, so it provides a very low floor at which to start. To potentially improve the fit, the B3LYP functional was also used. The 6-31G(d) basis set

Table 13: Description of vibrations used for experimental comparison. Relative Intensity is calculated as the experimental Raman intensity of the smaller wavenumber peak divided by the experimental Raman intensity of the larger wavenumber peak.

Molecule	Vibrational Mode	Frequency (cm <sup>-1</sup> )	Vibrational Mode	Frequency (cm <sup>-1</sup> )	Relative Intensity
acetone	C-C Str	787	C=O Str	1710	2.4
methyl acetate	C-C Str	641	C=O Str	1740	5.2
2-methyl-2-butene	C-C Str	529	C=C Str	1678	0.38
nitroethane	C-C-N Bend C-N-O Bend	493	C-N-O Bend	1555	9.7
<i>N,N</i> -dimethylformamide	O-C-N Bend C-N-C Bend	659	C=O Str	1660	3.6
tetrachloroethylene	C-Cl Str	449	C=C Str	1574	0.35
trichloroethylene	Cl-C=C Bend	631	C=C Str	1590	1.6

can be considered a minimum basis set for all but the largest systems. For this reason the size of the basis set was increased by utilizing the 6-31+G(2d,p) basis set, and also by utilizing the cc-pvdz, aug-cc-pVDZ, and aug-cc-pVTZ basis sets. This should span a large range between a minimum acceptable and largest practicable basis set.

The next variable to consider is the electron density contour at which the electronic volume is evaluated. In the same paper by Politzer good correlations were found between  $10^{-2}$  and  $10^{-3}$  electrons per cubic Bohr. To get a wide sampling the  $10^{-1}$ ,  $10^{-2}$ ,  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-5}$  contours were tested. This is likely to span from only encompassing the core electron density to highly diffuse electrons far away from all nuclei.

The final variable that must be optimized is the degree of displacement along the vibrational coordinates. The starting point for this search was the classical turning point. This was accomplished by converting the calculated vibrational frequency to energy, calculating the classical turning point for a spring with the same force constant, and displacing each atom in the molecule by the turning point distance multiplied by the relative atomic displacements for the vibration. Volumes were then calculated at whole number multiples of the turning point displacement using the code in Appendix B to generate structures.

The parameters described above merely represent a search for a reasonable correlation from which further optimization can take place. The work presented in this thesis represent 70 frequency calculations, 350 single-point calculations, and 5600 volume calculations. These calculations are intended to cast a wide net from which

trends can be identified. In the future a larger test set will be implemented and tested with the parameters found to yield good results in the present work.

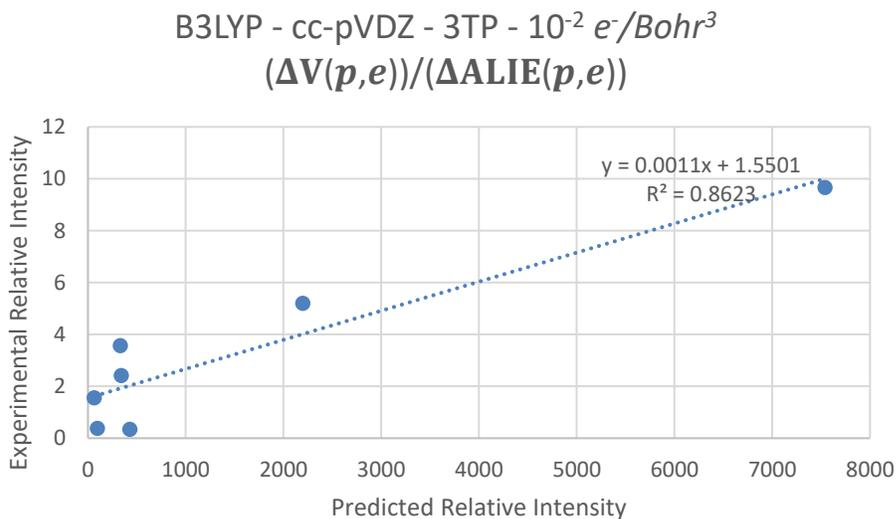
Using Equation 20 the theoretical Raman intensities were calculated using each of the parameters above. Politzer<sup>13</sup> demonstrated previously that bulk polarizabilities could be adequately modeled by both the equilibrium volume and the ratio of volume to average local ionization energy, so the corresponding term was used as applied to different points along the selected vibrations. The volume change term ( $\Delta V_{a,b}$ ) represents the molecular volume change from point  $a$  to  $b$  along the vibration, whereas the  $\Delta\left(\frac{V}{ALIE}\right)_{a,b}$  term represents the change in ratio of volume and average local ionization energy from point  $a$  to  $b$ . A third expression was included as well, the ratio of the volume difference and difference in average local ionization energy from point  $a$  to  $b$  ( $\frac{\Delta V_{a,b}}{\Delta ALIE_{a,b}}$ ). This expression was not taken from literature, but was included to encourage thoroughness. Each of these were calculated from the most positive amplitude to the most negative amplitude, from the most positive amplitude to equilibrium, and from equilibrium to the most negative amplitude. The last two were included in hopes of expanding the utility of this approach to symmetric vibrational modes. A linear regression for each of these was performed comparing the predicted intensities to the obtained experimental Raman intensities.

**Results.** The strongest correlation was found using the  $\frac{\Delta V_{p,e}}{\Delta ALIE_{p,e}}$  function and the B3LYP/cc-pVDZ model chemistry at the  $0.01 e/Bohr^3$  contour at three turning points of vibrational displacement. The  $R^2$  value found from the linear regression was 0.86, the graph of which can be found in Figure 25, and the data of which can be found in Table

14. Although this was the strongest correlation the predicted relative intensities were far higher than experimentally obtained.

Upon investigation it was found the reason for this massive discrepancy is that the volume changes and ALIE changes are very small. Volume changes tend to span from  $0.01 \text{ Bohr}^3$  to several tens of  $\text{Bohr}^3$ . ALIE values tend to be on the order of  $10^{-3} \text{ eV}$  or lower. Small errors in the calculation of ALIE can then greatly change the calculated Raman intensity and can potentially explain the large error.

Considering the volume-only functions retains the same problem. The best correlation using the volume only function was  $R^2 = 0.74$  and was found by utilizing the HF/6-31G(d) model chemistry at the  $10^{-5} \text{ e/Bohr}^3$  contour and fourth turning point. The linear regression can be seen in Figure 26. Although there was a correlation indicated we again have the issue of the predicted relative intensities being far higher than the experimental relative intensities.

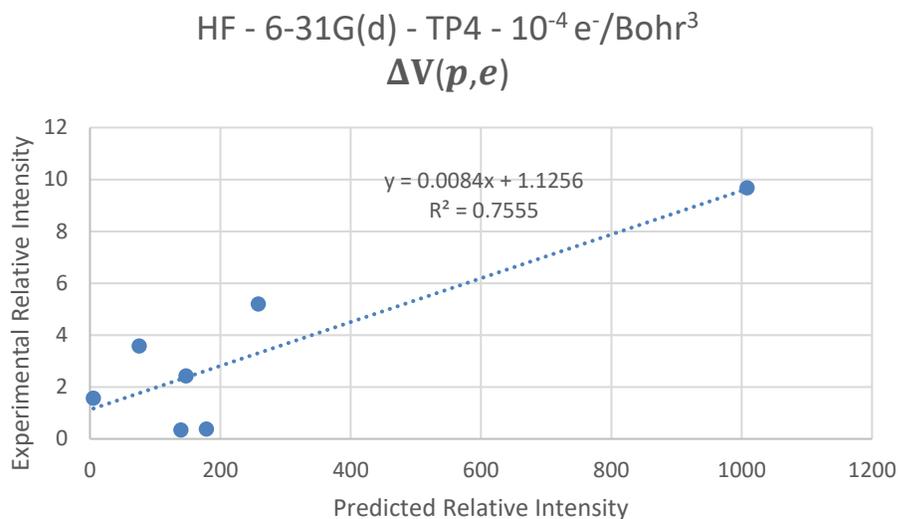


**Figure 25:** Linear regression for the strongest correlation found using the  $\frac{\Delta V_{a,b}}{\Delta ALIE_{a,b}}$  function.

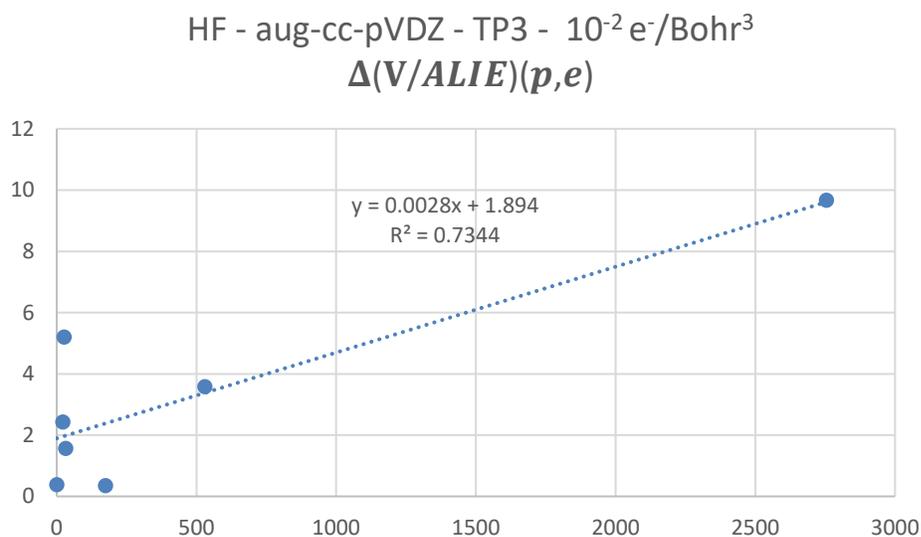
Table 14: Linear regression data for Figure 25.

Molecule	Predicted Relative Intensity	Experimental Relative Intensity
acetone	340	2.42
methyl acetate	2199	5.21
methylbutene	98	0.38
nitroethane	7539	9.67
n,n-dimethylformamide	330	3.58
tetrachloroethylene	428	0.35
trichloroethylene	64	1.57

The strongest correlation using the  $\Delta\left(\frac{V}{ALIE}\right)_{a,b}$  function was  $R^2 = 0.73$ , the linear regression of which can be seen in Figure 27.



**Figure 26:** Linear regression for the strongest correlation found using the  $\Delta V_{a,b}$  function.



**Figure 27:** Linear regression for the strongest correlation found using the  $\Delta\left(\frac{V}{ALIE}\right)_{a,b}$  function.

## CHAPTER 5: CONCLUSIONS & FUTURE WORK

### NMR of Nucleobases

It has been shown that NMR chemical shifts of nucleotides can be reasonably well predicted computationally. Carbon chemical shifts can typically be predicted to within a few percent of the experimental values, often within experimental error of 1-2%. In some cases, chemical shifts of specific nuclei are predicted less accurately, however the error is consistent across all model chemistries, and may be correctable.

The next step will be expanding the current work to include nucleosides. By comparing the method discussed herein against experimentally measured spectra of nucleosides it can be determined if the chemical shifts predicted are accurate for not only the nitrogenous base, but also for proton and carbon nuclei that are part of the ribose ring. Another important factor will be predicting J-coupling values of the ribose protons to determine if these can be adequately modeled. If they can be reasonably predicted then the prediction of large multi-dimensional spectra of DNA and RNA molecules will ensue.

### Volume Changes as Polarizability Changes

Relative Raman intensities were calculated using a function of electronic volume changes for a set of test molecules using several vibrational displacements, electron density contours, and model chemistries. Although a positive correlation exists between the predicted relative Raman intensities, the calculated values are very far away from the experimentally obtained values. Further work will need to be done to correct this error. This can be accomplished by tightening the convergence requirements for geometric

optimization, tightening the convergence requirements for determining the electron density, and tightening the grid used to compute the volume and ALIE changes.

## REFERENCES

- (1) Abhijit, M.; Seaton, P. J.; Assarpour, R. A.; Williamson, T., Unprecedented concentration dependent chemical shift variation in  $^1\text{H}$ -NMR studies: A caveat in the investigations of molecular recognition and structure elucidation. *Tetrahedron* **1998**, 54 (51), 15489-15498.
- (2) Campagne, S.; Gervais, V.; Milon A.; Nuclear magnetic resonance analysis of protein-DNA interactions. *J. Royal. Soc. Interface* **2011**, 8 (61), 1065-1078
- (3) Drew, H. R.; Wing R. M.; Takano T.; Broka, C.; Tanaka, S.; Itakura K.; Dickerson, R. E., Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.*, **1981**, 78 (4), 2179-2183
- (4) Bruice, P. Y., *Organic Chemistry*, Seventh Edition, Pearson, London England, 2013.
- (5) Kong, K.; Kendall, C.; Stone, N.; Notingher, I, Raman spectroscopy for medical diagnostics--From in-vitro biofluid assays to inv-vivo cancer detection. *Adv. Drug Deliv. Rev.* **2015**, 89, 121-134.
- (6) San Feliciano, A.; Medarde, M.; Miguel del Corral, J. M.; Aramburu, A.; Gordaliza M.; Barrero, A. F., Aquatolide. A new type of humulane-related sesquiterpene lactone. *Tetrahedron Lett.* **1989**, 30 (21), 2851-2854.
- (7) Lodweyk, M. W.; Soldit, C.; Jones, P. B.; Olmstead, M. M.; Rita, J.; Shaw, J. T.; Tantillo, D. J., The correct structure of aquatolide—Experimental validation of a theoretically-predicted structural revision. *J. Am. Chem. Soc.*, **2012**, 134 (45), 18550-18553.
- (8) Evidente, A.; Abou-Donia, A. H.; Darwish, F. A.; Amer, M. E.; Kassem, F. F.; Hammada, H. A.; Motta, A., Nobilisitine A and B, two masanane-type alkaloids from *Clivia nobilis* fn1. *Phytochemistry*. **1999**, 51 (8), 1151-1155.
- (9) Schwartz, B. D.; Jones, M. T.; Banwell, M. G.; Cade, I. A., Synthesis of the Enantiomer of the Structure Assigned to the Natural Product Nobilisitine A. *Org. Lett.* **2010**, 12 (22), 5210-5213
- (10)Zanardi, M. M.; Sarotti, A., M., GIAO C-H COSY Simulations Merged with Artificial Neural Networks Pattern Recognition Analysis. Pushing the Structural Validation a Step Forward. *J. Org. Chem.* **2015**, 80 (19) 9371-9378.
- (11)Zvereva, E. E.; Shagidullin, A. R.; Katsyuba, S. A., Ab Initio and DFT Predictions of Infrared Intensities and Raman Activities. *J. Phys. Chem. A*, **2011**, 115 (1), 63-69.

- (12) Politzer, P.; Jin, P.; Murray, J. S., Atomic polarizability, volume and ionization energy. *J. Chem. Phys.*, **2002**, 117, 8197-8202.
- (13) Brinck, T.; Murrays, J. S.; Politzer, P., Polarizability and volume. *J. Chem. Phys.*, **1992**, 98 (5), 4305-4306
- (14) Gaussian16 Revision A.03 Frisch, M. J.; et al, Gaussian, Inc., Wallingford CT, 2016
- (15) McQuarrie, D. A., Simon, J. D.; *Physical Chemistry: A Molecular Approach*, University Science Books: Sausalito, CA, 1997
- (16) Cramer, C. J., *Essentials of Computational Chemistry Theories and Models*, Second Edition, John Wiley and Sons, Inc., Hoboken NJ, 2004.
- (17) Voet, D.; Voet, J. G.; Pratt, C. W.; *Fundamentals of Biochemistry Life at the Molecular Level*, Fourth edition; John Wiley & Sons, Inc., Hoboken NJ, 2013
- (18) II-Cooperativity effect on the base stacking interactions in DNA: is there a novel stabilization factor coupled with base pairing H-bonds?. *Phys. Chem. Chem. Phys.* **2014**, 16 (29), 15527-15538\
- (19) Drew, H. R.; Wing R. M.; Takano T.; Broka, C.; Tanaka, S.; Itakura K.; Dickerson, R. E., Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.*, **1981**, 78 (4), 2179-2183
- (20) Dans, P. D.; Danilane, L; Ivani, I.; Drsata, T.; Lankas, F.; Hospital, A.; Walther, J.; Pujagut, R. I.; Battistini, F.; Lavery, J. L. G. R.; Orozco, M., Long-timescale dynamics of the Drew-Dickerson dodecamer. *Nucleic Acids Res.*, **2016**, 44 (9), 4052-4066.
- (21) Aguiar, P.H.; Furtado D.; Repoles, B. M.; Ribeiro, G. A.; Mendes, I. C.; Peloso, E. F.; Gadelha, F. R.; Macedo, A. M.; Franco, G. R.; Pena, S. D.; Teixeira, S. M.; Vieira, L. Q.; Guarneri, A. A.; Andrade, L. O.; Machado, C. R., Oxidative stress and DNA lesions: the role of 8-oxoguanine lesions in Trypanosoma cruzi cell viability. *PLoS. Negl. Trop. Dis.*, **2013**, 7 (6), e2279
- (22) Frances-Monerros, A.; Merchan, M.; Roca-Sanjuan, D., Mechanism of the OH Radical Addition to Adenine from Quantum-Chemistry Determinations of Reaction Paths and Spectroscopic Tracking of Intermediates. *J. Org. Chem.*, **2017**, 82 (1), 276-288.
- (23) Levine, R. L.; Yang, I. Y.; Hossain, M.; Pandya, G. A.; Grollman, A. P.; Moriya M., Mutagenesis induced by a single 1,N6-ethenodeoxyadenosine adduct in human cells. *Cancer Res.*, **2000**, 60 (15), 4098-4104.
- (24) Saporbaev, M.; Laval, J., 3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for Escherichia coli double-stranded uracil-DNA glycosylase and human

- mismatch-specific thymine-DNA glycosylase. *Proc. Natl. Acad. Sci. U. S. A.*, **1998**, 95 (15), 8508-8513
- (25) Shrivastav, N.; Li, D.; Essigmann, J. M., Chemical biology of mutagenesis and DNA repair: cellular responses to alkylation. *Carcinogenesis.*, 2010, 31 (1), 59-70.
- (26) Lewis, C. A.; Caryle, J.; Zhou, S.; Swanstrom, R.; Wolfenden, R., Cytosine deamination and the precipitous decline of spontaneous mutation during Earth's history. *Proc. Natl. Acad. Sci. U S A.*, 2016, 113 (29), 8194-8199.
- (27) Freisinger, E.; Fernandes, A.; Grollman, A. P.; Kisker, C., Crystallographic characterization of an exocyclic DNA adduct: 3,N4-etheno-2'-deoxycytidine in the dodecamer 5'-CGCGAATTepsilonCGCG-3'. *J. Mol. Biol.*, **2003**, 239 (4), 685-697
- (28) Leonard, G. A.; McAuley-Hecht, K. E.; Gibson, N. J.; Brown, T.; Watson, W. P.; Hunter, W. N., Guanine-1,N6-Ethenoadenine Base Pairs in the Crystal Structure of d(CGCGAATT(.epsilon)dA)GCG). **1994**, 33 (16) 4755-4761
- (29) Hoppins, J. J.; Gruber, D. R.; Miers, H. L.; Kiryutin, A. S.; Kasymov, R. D.; Petrova, D. V.; Endutkin, A. V.; Popov, A. V.; Yurkovskaya, A. V.; Fedechkin, S. O.; Brockerman, J. A.; Zharkov, D. O.; Smirnov, S. L., 8-Oxoguanine Affects DNA Backbone Conformation in the EcoRI Recognition Site and Inhibits Its Cleavage by the Enzyme. *PLoS One*, **2016**, 11 (10).
- (30) Jacobsen, N. E., *NMR Spectroscopy Explained Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*, John Wiley and Sons, Inc., Hoboken NJ, 2007
- (31) Wilson, E. B.; Decius, J. C.; Cross, P. C., *Molecular Vibrations The Theory of Infrared and Raman Vibrational Spectra*, Dover Publication, Mineola NY.
- (32) Szymanski, H. A., *Raman Spectroscopy Theory and Practice*, Plenum Press, New York, 1967.
- (33) Halliday, D.; Resnick, R.; Walker, J., *Fundamentals of Physics*, John Wiley and Sons, Inc., Hoboken NJ, 2010.
- (34) D. N. Sathyanarayana., *Vibrational Spectroscopy: Theory and Applications.*, New Age International (P) Limited Publishers, New Delhi India
- (35) Laidig, K. E.; Bader, R. F. W., Properties of atoms in molecules: Atomic polarizabilities. *J. Chem. Phys.*, **1990**, 93 (10), 7213-7224.
- (36) Politzer, P.; Murray, J. S.; Bulat, F. A., Average local ionization energy: A review. *J. Mol. Model.*, **2010**, 16 (11), 1731-1742

- (37)Toro-Labbe, A.; Brinck, T.; Murray, J. S.; Politzer, P., Quantative analysis of molecular surfaces: areas, volumes, electrostatic potentials and average local ionization energies. *J. Mol. Model.*, **2010**, 16 (11), 1679-1691.
- (38)Stewart, J. J., Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.*, **2007**, 13 (12), 1173-1213.
- (39)Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionls., *Theor. Chem. Acc.*, **2008**, 120, (1-3), 215-241.
- (40)Zhao, Y.; Truhlar, D. G., Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions., *J. Chem. Theory Comput.* **2008**, 4 (11), 1849-1868.
- (41)Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange., *J. Chem. Phys.* **1993**, 98 (7), 5648-5653.
- (42)Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior., *Phys. Rev. A.* **1988**, 38 (6), 3098-3101.
- (43)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 (2), 785-790.
- (44)Perdew, J. P., Density-functional approximation for the correlation energy of the inhomogeneous electron gas., *Phys. Rev. B.* **1986**, 33 (12), 8822-8825
- (45)Adamo, C.; Barone, V., Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models., *J. Chem. Phys.* **1998**, 108 (2), 664-676.
- (46)Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation made Simple., *Phys. Rev. Lett.* **1996**, 77 (18), 1396
- (47)Grimme, S., Semiempirical hybrid density functional with perturbative second-order correlation., *J. Chem. Phys.* **2006**, 124 (3), 34108-34112.
- (48)Sadlej, A. J.; Urban, M., Medium-size polarized basis sets for high-level-correlated calculations of molecular electric properties: III. Alkali (Li, Na, K, Rb) and alkaline-earth (Be, Mg, Ca, Sr) atoms., *J. Mol. Struct. (Theochem).* **1991**, 234 (20), 147-171.
- (49)Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, 18, 3297-3305.

- (50)Gauss, J., Accurate Calculation of NMR Chemical Shifts. *Ber. Bunsenge. Ph. Chem.*, **1995**, 99 (8), 1001-1008
- (51)Lodewyk, M. W.; Siebert, M. W.; Tantillo, D. J., Computational Prediction of  $^1\text{H}$  and  $^{13}\text{C}$  Chemical Shifts: A Useful Tool for Natural Product, mechanistic, and Synthetic Organic Chemistry
- (52)Vibrational Analysis in Gaussian <http://www.gaussian.com/vib> (accessed 5/5/2018)
- (53)Jamroz, M. H., Vibrational energy distribution analysis (VEDA): Scopes and limitations. *Spectrochim. Acta A*, **2013**, 114, 220-230
- (54)Frisch, M. J.; Yamaguchi, Y.; Gaw, J. F.; Schaefer III, H. J., Analytic Raman intensities from molecular electronic wave functions. *J. Chem. Phys.*, **1986**, **84**, 531-532.
- (55)Yamaguchi, Y.; Frisch, M. J.; Gaw, J.; Schaefer III, H. F.; Binkley, J. S., Analytic evaluation and basis set dependence of intensities of infrared spectra., *J. Chem. Phys.*, **1986**, 84, 2262-2278
- (56)*Spectral Database for Organic Compounds (SDBS)*: 1088CDS-05-207; <http://sdb.sdb.aist.go.jp> (National Institute of Advanced Industrial Science and Technology, accessed May 15<sup>th</sup> 2018)
- (57)Purnell, L. G.; Hodgson, D. J., Carbon-13 n.m.r. studies of purines and 8-azapurines in basic aqueous medium. *Magn. Reson. Chem.*, **1977**, 10 (1), 1-4.
- (58)*Human Metabolome Database* Accessed May 15<sup>th</sup> 2018 <http://www.hmdb.ca/metabolites/HMDB0000132>
- (59)Chenon, M. T.; Pugmire, R. J.; Grant, D. M.; Panzica, R. P.; Townsend, L. B., Carbon-13 magnetic resonance. XXVI. A quantitative determination of the tautomeric populations of certain purines. *J. Am. Chem. Soc.*, **1975**, 97 (16), 4636-4642.
- (60)*Spectral Database for Organic Compounds (SDBS)*: 1085CDS-03-114; <http://sdb.sdb.aist.go.jp> (National Institute of Advanced Industrial Science and Technology, accessed May 15<sup>th</sup> 2018)
- (61)Yamada, H.; Hirobe, M.; Higashiyama, K.; Takahashi, H.; Suzuki, K. T., Detection of carbon-13-nitrogen-15 coupled units of adenine derived from doubly labeled hydrogen cyanide or formamide., *J. Am. Chem. Soc.* **1978**, 100 (14), 4617-4618.
- (62)Botlan, D. J. L.,  $^1\text{H}$  and  $^{13}\text{C}$  NMR study of the interaction of formaldehyde on adenine and its derivatives. *Magn. Reson. Chem.*, **1989**, 27 (3), 295-298.

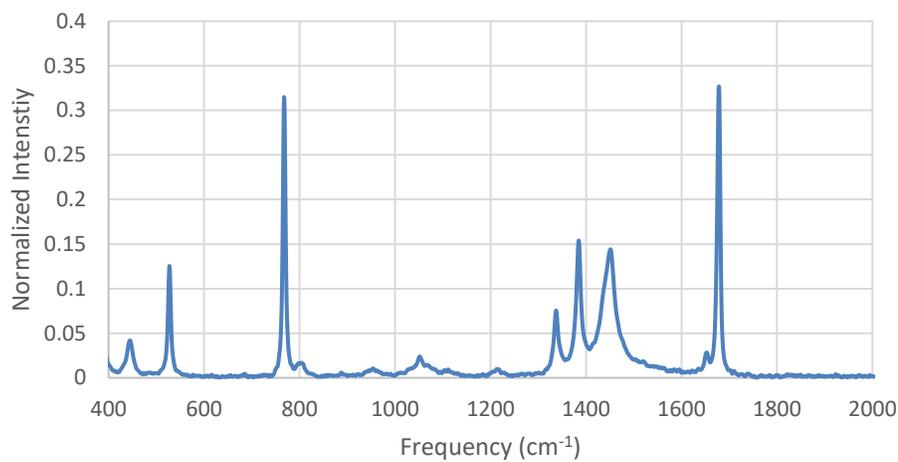
- (63)Laxer, A.; Major, D. T.; Gottlieb, H. E.; Fischer, B., (15N5)-Labeled Adenine Derivatives: Synthesis and Studies of Tautomerism by 15N NMR Spectroscopy and Theoretical Calculations. *J. Org. Chem.*, **2001**, 66 (16), 5463-5481.
- (64)*Spectral Database for Organic Compounds (SDBS)*: 1066CDS-05-240; <http://sdb.db.aist.go.jp> (National Institute of Advanced Industrial Science and Technology, accessed May 15<sup>th</sup> 2018)
- (65)Saladino, R.; Crestini, C.; Neri, V; Circirello, F.; Costanzo, G.; Di Mauro, E., Origin of informational polymers: The concurrent roles of formamide and phosphates. *Chembiochem.* **2006**, 7 (11), 1707-1714.
- (66)Kondratyuk, I. V.; Kolomiets, I. N.; Somoilenko, S. A.; Zheltovsky, N. V., A study of complexes between cytosine bases and amino acid carboxylic group by NMR spectroscopy. *Biopolym. Cell*, **1989**, 5 (6), 21-25
- (67)Coletta, F.; Etorre, R.; Gambaro, A., 1H and 13C NMR studies of palladium (II) and platinum (II) complexation of cytosine. *J. Magn. Reson.*, **1969**, 22 (3), 453-457.
- (68)*Spectral Database for Organic Compounds (SDBS)*: 1089HSP-49-450; <http://sdb.db.aist.go.jp> (National Institute of Advanced Industrial Science and Technology, accessed May 15<sup>th</sup> 2018)
- (69)Kubica, D.; Molchanov, S.; Gryff-Keller, A., Solvation of Uracil and its Derivatives by DMSO: A DFT\_Supported 1H NMR and 13C NMR Study. *J. Phys. Chem. A.*, **2017**, 121 (8), 1841-1848.
- (70)Ding, Z.; Zhao, J.; Yang, P.; Li, M.; Huang, R.; Cui, X.; Wen, M., 1H and 13C NMR assignments of eight nitrogen containing compounds from *Nocardia alba* sp.nov (YIM 30243T). *Magn. Reson. Chem.*, **2009**, 47 (4), 366-370.
- (71)*Spectral Database for Organic Compounds (SDBS)*: 1062CDS-05-214; <http://sdb.db.aist.go.jp> (National Institute of Advanced Industrial Science and Technology, accessed May 15<sup>th</sup> 2018)
- (72)Wevers, R. A.; Engelke, U. F.; Moolenaar, S. H.; Brautigam, C.; de Jong, J. G.; Duran, R.; de Abreu, R. A.; van Gennip, A. H., 1H-NMR spectroscopy of body fluids: inborn errors of purine and pyrimidine metabolism. *Clin. Chem.*, **1999**, 45 (4), 539-48.
- (73)*Spectral Database for Organic Compounds (SDBS)*: 1088HPM-00-748; <http://sdb.db.aist.go.jp> (National Institute of Advanced Industrial Science and Technology, accessed May 15<sup>th</sup> 2018)
- (74)*Spectral Database for Organic Compounds (SDBS)*: 1085HSP-06-316; <http://sdb.db.aist.go.jp> (National Institute of Advanced Industrial Science and Technology, accessed May 15<sup>th</sup> 2018)

- (75) Shen W.; Kim, J. S.; Mitchell, S.; Kish, P.; Kijek, P.; Hilfinger, J., 5'-O-D-valyl ara A, a potential prodrug for improving oral bioavailability of the antiviral agent vidarabine. *Nucleosides Nucleotides Nucleic Acids*, **2009**, 28 (1), 43-55.
- (76) Laxer, A.; Major, D. T.; Gottlieb, H. E.; Fischer, B., (15N5)-Labeled Adenine Derivatives: Synthesis and Studies of Tautomerism by 15N NMR Spectroscopy and Theoretical Calculations. *J. Org. Chem.*, **2001**, 66 (16), 5463-5481.
- (77) *Spectral Database for Organic Compounds (SDBS)*: 1089CDS-13-026;  
<http://sdb.sdb.aist.go.jp> (National Institute of Advanced Industrial Science and Technology, accessed May 15<sup>th</sup> 2018)
- (78) *Spectral Database for Organic Compounds (SDBS)*: 1062HSP-42-229;  
<http://sdb.sdb.aist.go.jp> (National Institute of Advanced Industrial Science and Technology, accessed May 15<sup>th</sup> 2018)
- (79) Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.*, **2005**, 105 (8), 2999-3094

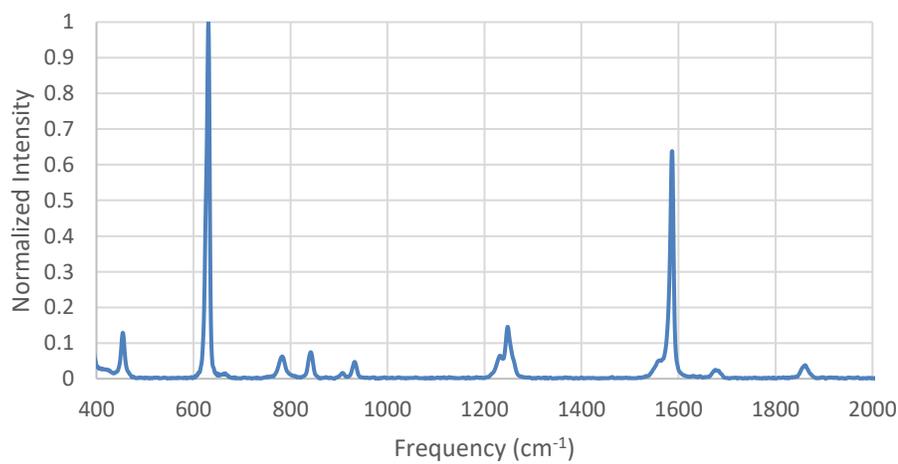
## APPENDICES

### Appendix A: Experimental Raman Spectra for Test Set Molecules

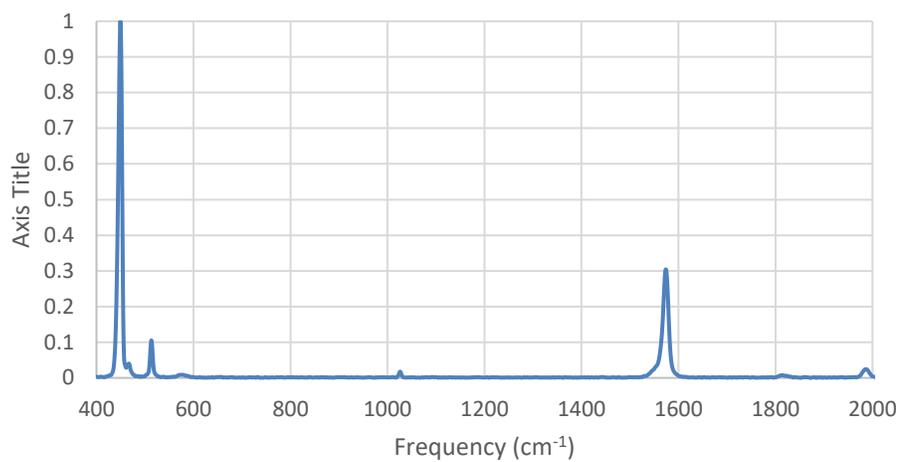
2-methyl-2-butene Raman Spectrum



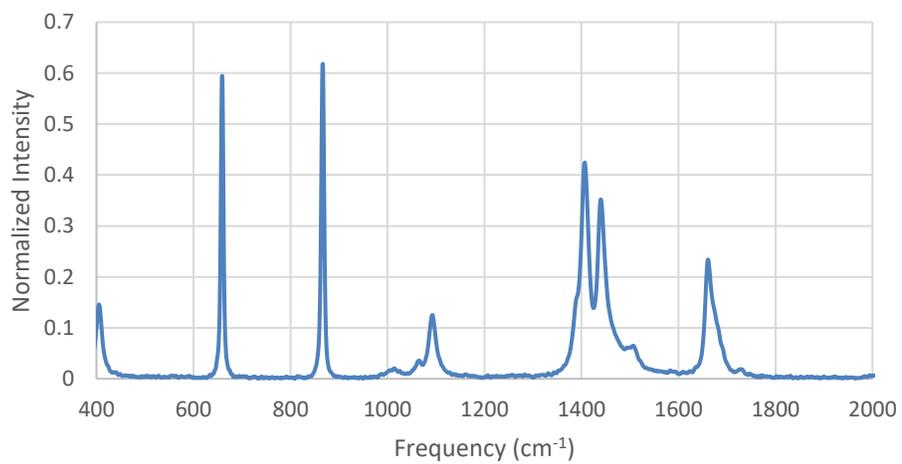
trichloroethylene Raman Spectrum



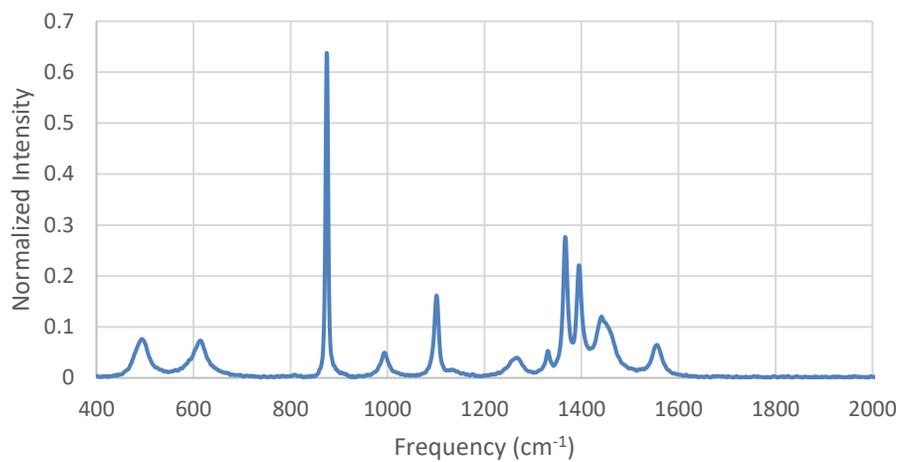
tetrachlorethylene Raman Spectrum



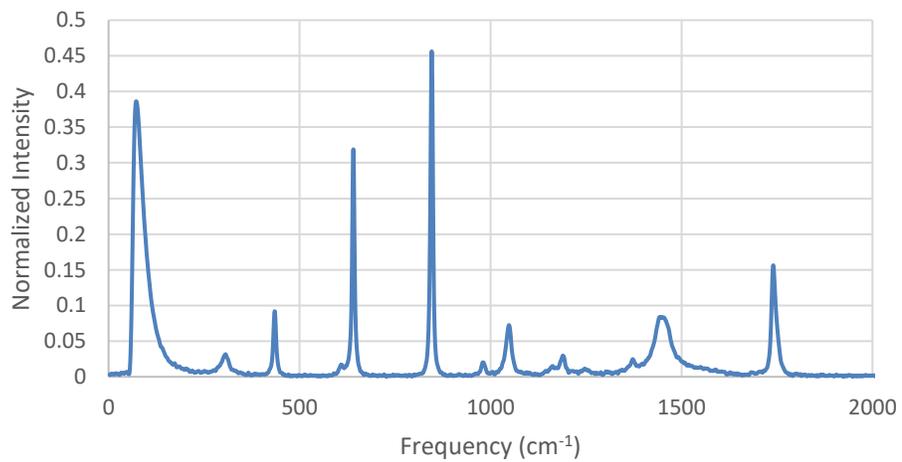
*N,N*-dimethylformamide Raman Spectrum



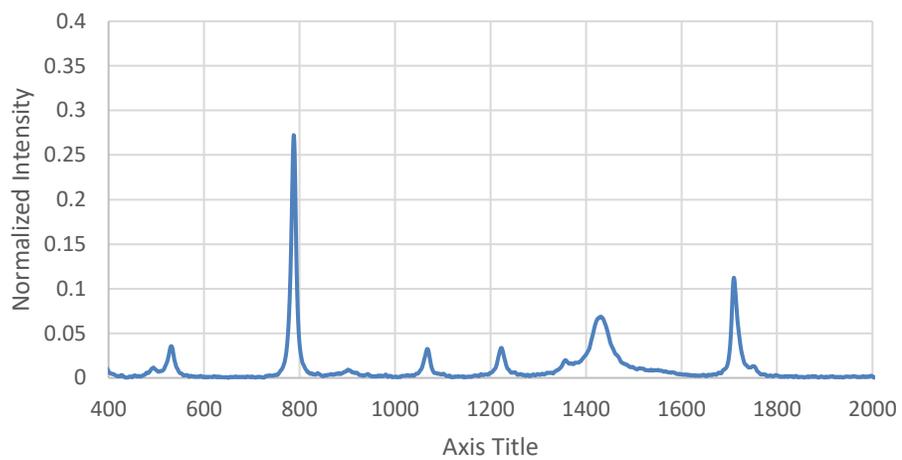
nitroethane Raman Spectrum



methylacetate Raman Spectrum



### acetone Raman Spectrum



## Appendix B: Code to Generate Vibrationally Displaced Structures

```
import math
import inspect
import sys
import os

#####
###
# THIS PROGRAM WAS WRITTEN TO EXTRACT THE NECESSARY INFORMATION FROM
A GAUSSIAN 16 LOG FILE
# AND GENERATE A CLASSICAL TURNING-POINT STRUCTURE FOR EACH
VIBRATION
# This script only works with opt+freq jobs, specifically with "opt
freq=(raman, hpmodes)" in the route line exactly and with no other
jobs.
#The .log file from the opt+freq job should be in the same directory
as this script.
#####
###

def GetRoute(FileName):
    File = open(FileName+".gjf", 'r')
    Temp = []
    for x in File:
        if x.count("#") > 0:
            return x.replace('opt freq=(raman, hpmodes)', '')

def GetFreqJobs(FileName): #separate out frequency section for each
opt+freq. Makes the rest easier. Assumes run with opt freq=(raman,
hpmodes)
    File = open(FileName+".log", 'r')
    ListOfFreqs = []
    Count = 1
    TempList = []
    for x in File: #For every line in the file
        if x.count("Normal termination") > 0: #If you hit a
termination line
            Count+=1 #then you are proceeding to a new job link
            if Count != 2: #For every time except when you hit the
first one (because the TempList will be empty)
                ListOfFreqs.append(TempList) #append the TempList
containing the freq job lines
                TempList = [] #and reset the original list so a new
freq job can be stored
            if Count%2 == 0: #Only grab freq lines every other
termination step
                TempList.append(x)
    for x in range(ListOfFreqs.count([])):
        ListOfFreqs.remove([])
    return ListOfFreqs

def GetGeometry(FreqJobs):
    Orientation = []
```

```

Trigger = 0
for x in FreqJobs:
    TempList = []
    for y in x: #Separate out relevant parts of the log file
        if Trigger == 0:
            pass
        else:
            TempList.append(y)
            if y.count("Standard orientation") == 1:
                Trigger = 1
            if y.count("Rotational constants") == 1:
                Trigger = 0
            Orientation.append(TempList)
            TempList = []
    Orientation = RemoveDuplicates(Orientation)
    Orientation2 = []
    for x in Orientation: #Separate out irrelevant information from
standard orientation, leaving only atom descriptions.
        TempList = []
        for y in range(len(x)):
            if x[y] in [x[0], x[1], x[2], x[3], x[-1], x[-2]]:
                pass
            else:
                TempList.append(x[y].replace('\n', ' '))
        if TempList != []:
            Orientation2.append(TempList)
    Orientation3 = []
    for x in Orientation2: #Finally organize the final output into a
list containing each job. Each job's list contains lists of format
[AtomicNumber, X, Y, Z]
        List = []
        for y in x:
            OneAtom = []
            for z in y.split(' '):
                if z != '':
                    OneAtom.append(z)
            List.append(OneAtom)
        Orientation3.append(List)
    for x in Orientation3:
        for y in x:
            del y[0]
            del y[1]
    return Orientation3

def RemoveDuplicates(List): #Removes duplicate items from a list
NewList = []
for x in List:
    if x not in NewList:
        NewList.append(x)
return(NewList)

def CheckHPModes(FileName): #Checks to make sure all freq jobs were
run with freq=hpmodes. Else this program won't work
Count = 0

```

```

for x in open(FileName+'.log', 'r'):
    if x.count("HPModes") > 0 or x.count("hpmodes") > 0:
        return True
return False

def GetLogInfo(LogFile, AtomNumber):
    #Returns Frequencies, Force Constants, then an array for
    relative displacements.
    Lines = []
    Trigger = 0
    for x in LogFile:
        Lines.append(x)
    FreqLines = []
    for x in Lines: #Separate out relevant parts of the log file
        if Trigger == 0:
            pass
        else:
            FreqLines.append(x)
            if x.count("normal coordinates") == 1:
                Trigger = 1
            if x.count("Thermochemistry") == 1:
                Trigger = 0
    del FreqLines[-1]
    del FreqLines[-1]
    del FreqLines[-1]
    for x in range((7+AtomNumber)*(AtomNumber-2)): #Remove second
    data set that is lower sig figs
        del FreqLines[-1]
    del FreqLines[-1]
    del FreqLines[-1] #Killing stuff we don't need
    del FreqLines[-1]
    del FreqLines[-1]
    OldFreqLines = FreqLines
    FreqLines = []
    for x in OldFreqLines:
        if x.count("Depol. (Plane)") > 0:
            pass
        else:
            FreqLines.append(x)
    #Vibrations are broken up into sections of 5
    # floor((3N-5)/5) is the number of sections.
    #7+3n lines per section
    #Important data for each calculation is Frequency, Reduced Mass,
    Force Constants, Raman Activities, Depolarization Ratios and atom
    coordinates.
    SectionList = []
    for x in range(math.ceil((3*AtomNumber-6) / 5)):
        TempList = []
        for y in range((9+3*AtomNumber)):
            TempList.append(FreqLines.pop(0))
        SectionList.append(TempList)
    ForceConstantLines = []
    FrequencyLines = []
    RamanActivityLines = []

```

```

DepRatioLines = []
RedMassLines = []
for x in SectionList:
    DepRatioLines.append(x.pop(7).replace('\n', ''))
    RamanActivityLines.append(x.pop(6).replace('\n', ''))
    ForceConstantLines.append(x.pop(4).replace('\n', ''))
    RedMassLines.append(x.pop(3).replace('\n', ''))
    FrequencyLines.append(x.pop(2).replace('\n', ''))
    for y in range(4):
        del x[0]
SectionCount = len(SectionList)
ForceConstants = GetData(ForceConstantLines)
Frequencies = GetData(FrequencyLines)
RamanActivity = GetData(RamanActivityLines)
DepRatios = GetData(DepRatioLines)
RedMasses = GetData(RedMassLines)
print(DepRatios)
Temp = []
for x in SectionList:
    for y in x:
        y = y.replace('\n', '').split(' ')
        y = DeleteFromList(y, '')
        Temp.append(y)
SectionList = Temp
Temp = []
for x in SectionList:
    del x[0]
    del x[0]
    del x[0]
    Temp.append(x)
SectionList = Temp
Temp = []
Sections = []
OriginalSectionCount = len(SectionList)
for y in range(SectionCount):
    List = []
    for x in range(int(OriginalSectionCount/SectionCount)):
        List.append(SectionList.pop(0))
    Sections.append(List)

Temp = Sections[0]
for x in range(len(Sections)):
    for y in range(len(Sections[x])):
        if x == 0:
            pass
        else:
            Temp[y] += Sections[x][y]
SectionList = Temp
XList = []
YList = []
ZList = []
Vibration = 0
for x in SectionList:
    Vibration += 1

```

```

    for y in x:
        if Vibration%3 == 1:
            XList.append(y)
        if Vibration%3 == 2:
            YList.append(y)
        if Vibration%3 == 0:
            ZList.append(y)
    AtomList = []
    VibrationNumber = 0
    TempList = []
    for x in range(len(XList)): #Iterate over XList, YList, and
ZList, popping off first coordinates
        VibrationNumber+=1
        Coordinates = [XList.pop(0), YList.pop(0), ZList.pop(0)]
#Add the coordinates to an [X, Y, Z] List until you've run out of
vibrations. There are 3N-6 Vibrations.
        AtomList.append(Coordinates)
        if VibrationNumber == (3*AtomNumber-6): #If you've gone
through it the number of times as there are vibrations
            TempList.append(AtomList) #Then add it to the temporary
list. This is a list of one atom's normalized vibrational
cooridantes
            VibrationNumber = 0 #and reset the vibration number
            AtomList = [] #and start a new atom's list
        VibrationList = TempList #Rename it so it looks nicer.
#VibrationList is now a list. Each element is an atom in the
molecule. Each one of the atom's list is it's normalized
vibrational displacement, in order from Vib1 to Vib(3n-6).
        return [Frequencies, ForceConstants, VibrationList,
RamanActivity, DepRatios, RedMasses]
#Returns the above. Organized as follows. List of Atoms ->
List of XYZ Coordinates in vibrational order -> XYZ Coordinates

def GetData(Lines): #Gets data from lines like freq and force
constants. Can be used to get data from any part of that section if
needed.
    Data = []
    for x in Lines:
        for y in x.replace('\n', '').split(' '):
            Data.append(y)
    ItemsToDelete = ['Frequencies', '---', '', 'Force', 'constants',
'Depol.', '(Unpol)', 'Raman', 'Activities', "Reduced", "masses"]
    for x in ItemsToDelete:
        Data = DeleteFromList(Data, x)
    return Data

def DeleteFromList(List, ItemToDelete): #Got tired of typing it.
NewList = []
for x in List:
    if x != ItemToDelete:
        NewList.append(x)
return NewList

```

```

def WriteInfo(Stuff):
    for x in Stuff:
        NumberOfAtoms = x[0]
        OriginalGeometry = x[1]
        FrequencyJob = x[3]

def GetTurningPoints(Frequencies, ForceConstants):
    SoL = 3*10**8 #meters/second
    Planck = 6.626*10**-34 #J s
    TurningFactors = []
    for x in range(len(Frequencies)):
        try:
            TurningFactors.append(float(math.sqrt(float(Frequencies[x])*200*SoL*
            Planck/(float(ForceConstants[x])*100)))*(10**10))
        except ZeroDivisionError:
            TurningFactors.append(float(0))
    return TurningFactors

def GenerateJobFiles(FileName, WriteInfo, Route, Vibrations):
    FilePath = os.path.basename(sys.argv[0])
    Directory =
inspect.getfile(inspect.currentframe()).replace(FilePath, '')
    print("Writing Files To: \n")
    print(Directory)
    Count = 1
    JobFile = open(FileName+"-"+str(Count)+".gjf",'w')
    if Vibrations == "All" or Vibrations == "all":
        Vibrations = []
        for x in range(len(WriteInfo[0])):
            Vibrations.append(x+1)
        PeriodicTableDictionary = {1: 'H', 2: 'He', 3: 'Li', 4: 'Be', 5:
'B', 6: 'C', 7: 'N', 8: 'O', 9: 'F', 10: 'Ne', 11: 'Na', 12: 'Mg',
13: 'Al', 14: 'Si', 15: 'P', 16: 'S', 17: 'Cl', 18: 'Ar', 19: 'K',
20: 'Ca', 21: 'Sc', 22: 'Ti', 23: 'V', 24: 'Cr', 25: 'Mn', 26: 'Fe',
27: 'Co', 28: 'Ni', 29: 'Cu', 30: 'Zn', 31: 'Ga', 32: 'Ge', 33:
'As', 34: 'Se', 35: 'Br', 36: 'Kr', 37: 'Rb', 38: 'Sr', 39: 'Y', 40:
'Zr', 41: 'Nb', 42: 'Mo', 43: 'Tc', 44: 'Ru', 45: 'Rh', 46: 'Pd',
47: 'Ag', 48: 'Cd', 49: 'In', 50: 'Sn', 51: 'Sb', 52: 'Te', 53: 'I',
54: 'Xe', 55: 'Cs', 56: 'Ba', 57: 'La', 58: 'Ce', 59: 'Pr', 60:
'Nd', 61: 'Pm', 62: 'Sm', 63: 'Eu', 64: 'Gd', 65: 'Tb', 66: 'Dy',
67: 'Ho', 68: 'Er', 69: 'Tm', 70: 'Yb', 71: 'Lu', 72: 'Hf', 73:
'Ta', 74: 'W', 75: 'Re', 76: 'Os', 77: 'Ir', 78: 'Pt', 79: 'Au', 80:
'Hg', 81: 'Tl', 82: 'Pb', 83: 'Bi', 84: 'Po', 85: 'At', 86: 'Rn',
87: 'Fr', 88: 'Ra', 89: 'Ac', 90: 'Th', 91: 'Pa', 92: 'U', 93: 'Np',
94: 'Pu', 95: 'Am', 96: 'Cm', 97: 'Bk', 98: 'Cf', 99: 'Es', 100:
'Fm', 101: 'Md', 102: 'No', 103: 'Lr', 104: 'Rf', 105: 'Db', 106:
'Sg', 107: 'Bh', 108: 'Hs', 109: 'Mt', 110: 'Ds', 111: 'Rg', 112:
'Cn'}
        NumberOfAtoms = WriteInfo[0][0]
        OriginalGeometry = WriteInfo[0][1]
        Displacements = WriteInfo[0][2]
        TurningFactors = WriteInfo[0][3]
        RamanActivities = WriteInfo[0][4]

```

```

DepRatios = WriteInfo[0][5]
Frequencies = WriteInfo[0][6]
RedMasses = WriteInfo[0][7]
NumberOfVibrations = len(TurningFactors)
Vib = 0
JobFile.write("%nprocshared=12\n%mem=1500MB\n")
JobFile.write(Route+'\n\n')
JobFile.write("Equilibrium Geometry:  No Extra Info"+'\n\n0
1\n')
    for z in range(NumberOfAtoms):
        AtomType =
PeriodicTableDictionary[int(OriginalGeometry[z][0])]

JobFile.write(AtomType+"\t"+str(OriginalGeometry[z][1])+"\t"+str(OriginalGeometry[z][2])+"\t"+str(OriginalGeometry[z][3])+'\n')
    Directory =
inspect.getfile(inspect.currentframe()).replace(FilePath,
').replace("Geometry Calcs", "WFN Files")
    if not os.path.exists(Directory):
        os.makedirs(Directory)
    JobFile.write('\n'+Directory.replace("NonPolarSet",
"WFN_Files")+FileName+"-"+str(Count)+".wfn")
    JobFile.write("\n\n\n")
    JobFile.close()
    Pair=0
    for z in range(1, NumberOfVibrations+1):
        Vib+=1
        for a in [1, -1]:
            TPConstant = 1
            if Vib in Vibrations:
                Count+=1
                JobFile = open(FileName+"-"+str(Count)+".gjf", 'w')
                JobFile.write("%nprocshared=12\n%mem=1500MB\n")
                JobFile.write(Route+'\n\n')
                JobFile.write("VibrationNumber: "+str(a*Vib)+"
RamanActivity: "+RamanActivities[Vib-1] + " DepRatio:
"+DepRatios[Vib-1]+" Frequency: "+Frequencies[Vib-1]+" RedMass:
"+RedMasses[Vib-1]+'\n\n0 1\n')
                ThisVibrationDisplacements = []
                for y in Displacements:
                    ThisVibrationDisplacements.append(y[z-1])
                for y in range(NumberOfAtoms):
                    AtomType =
PeriodicTableDictionary[int(OriginalGeometry[y][0])]
                    DX = float(OriginalGeometry[y][1]) -
(float(a)*float(TPConstant)*float(ThisVibrationDisplacements[y][0]))*
TurningFactors[z-1])
                    DY = float(OriginalGeometry[y][2]) -
(float(a)*float(TPConstant)*float(ThisVibrationDisplacements[y][1]))*
TurningFactors[z-1])
                    DZ = float(OriginalGeometry[y][3]) -
(float(a)*float(TPConstant)*float(ThisVibrationDisplacements[y][2]))*
TurningFactors[z-1])

```

```

JobFile.write(AtomType+"\t"+str('{:.8f}'.format(DX))+"\t"+str('{:.8f}
}'.format(DY))+"\t"+str('{:.8f}'.format(DZ))+'\n') #write each
number to 8 digits past decimal.
    JobFile.write('\n'+Directory.replace("NonPolarSet",
"WFN_Files")+FileName+"-"+str(Count)+".wfn")
    JobFile.write('\n\n\n')
    JobFile.close()

```

```

def MakeVibFiles(FileName,Vibrations): #If you want to just make all
vibrations, use "all" or "All" as the Vibrations variable.

```

```

    FreqJobs = GetFreqJobs(FileName)
    OriginalGeometry = GetGeometry(FreqJobs)
    Route = GetRoute(FileName) + " 6d 10f output=wfn nosymm"
    if CheckHPModes(FileName): #Kill the generation if all jobs
weren't run with freq=hpmodes, since this program won't work with
the log file.

```

```

        pass

```

```

    else:

```

```

        print("Make sure all jobs are run with Freq=HPModes\nNo
Files Written")

```

```

        return False

```

```

    NumberOfAtoms = len(OriginalGeometry[0])

```

```

    LogInfo = GetLogInfo(FreqJobs[0], NumberOfAtoms)

```

```

    Frequencies = LogInfo[0]

```

```

    Terminate = 0

```

```

    for y in Frequencies:

```

```

        if float(y) <= 0:

```

```

            print("One or more imaginary frequencies. Program
terminated. No Files Written")

```

```

            Terminate = 1

```

```

    if Terminate == 1:

```

```

        return False

```

```

    WriteInfo = []

```

```

    OriginalGeometry = OriginalGeometry[0]

```

```

    Displacements = LogInfo[2]

```

```

    Frequencies = LogInfo[0]

```

```

    ForceConstants = LogInfo[1]

```

```

    TurningPoints = GetTurningPoints(Frequencies, ForceConstants)

```

```

    RamanActivities = LogInfo[3]

```

```

    DepRatios = LogInfo[4]

```

```

    ReducedMasses = LogInfo[5]

```

```

    WriteInfo.append([NumberOfAtoms, OriginalGeometry,
Displacements, TurningPoints, RamanActivities, DepRatios,
Frequencies, ReducedMasses])

```

```

    print("Writing Files")

```

```

    GenerateJobFiles(FileName, WriteInfo, Route, Vibrations)

```

```

def main():

```

```

    FileName = "Type File Name Here"

```

```

    Vibrations = ["Vibration Numbers Go Here"]

```

```

    # Example: FileName = "Methane"

```

```

    # Example: Vibrations = [1, 2, 3]

```

```
    MakeVibFiles (FileName, Vibrations)
main ()
```