A Thermochemical Investigation of Gas Phase Anionic Nucleobase Complexes by Infrared Multiple Photon Dissociation Spectroscopy and Electronic Structure Calculations

by

Andres G. Escobar

A thesis presented to the University of Waterloo

in fulfillment of the thesis

requirements for the degree of

Master of Science

in

Chemistry

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this thesis, a computational and experimental investigation of gas phase cation-molecule complexes has been undertaken to analyse the gas phase clustering behaviour of biologically relevant cyclic anionic species. Evaluation of computationally determined thermochemical properties to predict the energetic favourability for the formation the cation-molecule clusters as well as the vibrational frequencies unique to each cluster has been carried out. The computational analyses performed in this thesis used the Becke 3-Parameter, Lee, Yang, and Parr (B3LYP) level of theory for geometric optimization and frequency calculations and second order Møller-Plesset perturbation theory (MP2) for single-point energy calculations. In both cases, a split basis set was used to evaluate the non-metallic atoms at 6-311++g(d,p) and the metallic sodium cation at aug-cc-pCVTZ to more accurately evaluate each of the electrons in the system. Complete characterization of each cluster includes comparison to experimentally determined vibrational spectra, which in turn allows for further geometric, electronic, and structural analysis.

High-powered lasers, such as a free-electron laser (FEL), produce experimental vibrational spectra through a process known as infrared multiple photon dissociation (IRMPD). IRMPD induces fragmentation of ions produced by electrospray ionization (ESI) through a phenomenon known as internal vibrational redistribution (IVR) where energy absorbed by the analyte ion is redistributed to other vibrational modes. The process of IRMPD can be coupled with electronic structure calculations to gain significant insight into the characterization of ions in the gas phase.

The energetic favourability for deprotonating nucleobase monomers is investigated in Chapter 3 to determine the isomer most likely to exist for each nucleobase. Once the most energetically favourable nucleobase isomer was validated experimentally, the clustering behaviour of the monomers with sodium cations was further explored. The clustering of a single deprotonated isomer with one cationic sodium yielded one minimum energy structure for each of the nucleobases, however, as the complexity of the clusters increased the number of possible clustering geometries increased as well. The pyrimidines exhibited unusually favourable complexation of three deprotonated nucleobase monomers in a near C_3 symmetry, as well as an extremely favourable linear arrangement. These two favourable geometrical structures were further explored in thymine and uracil complexes as they were the only high-complexity experimental data collected for the nucleobases. Unfortunately, for the purines and cytosine, there was no experimental vibrational spectra to draw conclusions from, however, general trends from computational analysis, as well as similarities between the pyrimidines, were possible to evaluate. This examination found that the clustering behaviour for all the deprotonated nucleobases (at each level of complexity) demonstrated extremely energetically favourable clustering behaviour the gas phase. In addition, the gas phase acidities (GPAs) of each nucleobase were determined to be of comparable strength to other well-documented gas phase acid strengths and found to have acid strengths non-linearly correlated to their respective solution phase pK_a's. Moreover, novel clustering geometries for multiple deprotonated nucleobase monomers with cationic sodium were revealed. Thus, the importance of evaluating deprotonated nucleobases, in the gas phase, was demonstrated by this study.

This investigation revealed strong evidence to support the favourable clustering behaviour of the deprotonated nucleobases with cationic sodium, as well as novel geometries that describe these interesting new complexes in the gas phase. In each case, novel theoretical clustering geometries and thermodynamic properties for each of the biological species evaluated were determined. The hope is to be able to experimentally corroborate these findings in the future.

Acknowledgements

I would like to extend my greatest thanks to my supervisor, Dr. Terry B. McMahon, for his support, insight, and guidance at every stage of my academic and personal development since my first coop with him in 2017. Terry's infectious enthusiasm for "playing with molecules" has become the foundation for my continued passion for theoretical chemistry and reaffirmed my love for scientific advancement. I was originally invited to conduct research in the Netherlands and France, prior to the travel restrictions due to COVID-19, and despite not being able to conduct this research, I cannot understate my appreciation for the wonderful opportunities I was given. I often found myself spending countless hours in Terry's office chatting about our travel experiences, cycling, various cultures and relationships and appreciated his wisdom and experience on each of these subjects. Through Terry, my experience in graduate school has been nothing short of amazing and inspirational. Because of Terry, I have developed into a student and person that can proudly carry on the legacy of his lab, as well as confidently aim to create my own legacy. I am sad my time in Terry's lab is coming to an end, but I am certain this will not be the end of our friendship. Thank you for everything Terry!

I would like to thank the members of my committee, Bill Power and Scott Hopkins, for agreeing to be on my committee and for taking the time to read my thesis. Bill Power has been nothing but supportive through the hardest times in both my academic and personal lives. The jokes, personal updates and frequent trips to the Grad House always gave me something to look forward to and reminded me to always look forward, as "most things work out in the end as long as you keep trying". In addition, I enjoyed my interpersonal

privileges with Bill as I was always welcomed into his office, even when I did not have an appointment, and being able to confide in him as a friend. Bill reaffirmed my love for teaching through our time spent on the CHEM 254 Thermodynamics course and is something I will carry with me through the rest of my future endeavours. Furthermore, my time with Scott Hopkins, although limited, was extremely helpful in understanding the intrinsic value to the details in one's scientific work as well as understanding the value in being proud of one's accomplishments. Scott allowed me to work closely with many of the students in his research group and led to many of the wonderful experiences I have had in graduate school. The barbeque held at his house in the summer of 2019 is one of the best experiences I have ever had and will be a memory I will cherish dearly. Thank you, Scott and Bill, for being some of the best people I have had the privilege to work with.

I would also like to thank Fred McCourt and the late Robert J. LeRoy for providing me with the opportunity that started my research career. Bob and Fred welcomed me into their labs as a member of their research family, and not just another coop student, providing me with guidance and the freedom to explore my research interests within the realm of quantum chemistry. As a reward for my passion and amount of work accomplished in the 4-months we initially spent together, I was given the opportunity to travel with Bob to the University of Illinois, Urbana-Champagne, to present my work in a 30-minute research talk at the International Symposium for Molecular Spectroscopy. This conference was my first step into the world of academia and I immediately fell in love with the vastness of research and the passion other academics from around the world had for science. More importantly, I must thank Bob and Fred for their unwavering confidence in my potential as a researcher and their effusive opinions of me. They were

the reason I was given a chance to prove myself academically, through times of great loss, something I could only hope to repay them for by continuing to make the most of their efforts and confidence in me. Bob and Fred, you treated me as "a son [you] never expected to have" and I could not have been more grateful to have met such impactful inspirations such as yourselves. I hope to continue "upsetting the apple cart" in the science world and make my mark as a researcher thanks to the dedication you have both shown me. Bob, I miss you dearly and I am sorry I could not come back in time to share the Blue Label we planned to have upon my return.

In addition, I would like to thank ex-McMahon group member Dr. Joshua Featherstone, for his help in obtaining the experimental spectra for deprotonated nucleobases at the Free Electron Laser for Infrared Experiments (FELIX) in the Netherlands. Moreover, since my appointment to the McMahon group, Josh has been one of the greatest and most patient mentors I could have ever hoped for. He is a bright and unbelievably intelligent academic that is now enjoying the fruits of his labour at Oxford University for this post-doctoral. The time I spent with Josh, in our office, was never dull and gave me more reasons to look forward to going to work every day. The laughs, lessons and tea shared in that office are memories I look forward to continuing in the future.

I would also like to acknowledge my parents, Yolanda and Gustavo, as well as my sister Emma. I know I have made the journey to this milestone full of ups and downs, but I cannot thank you three enough for always being there for me regardless of how late in the night it was and how many times I would complain about the same things every time I came home. Dad, your cooking made coming home for the weekend more enticing and

worthwhile. The time we spent going over my research and discussing literature, philosophy, and your takes on scientific topics were and will continue to be one of the best parts about our relationship. I cannot thank you enough for the weekends, growing up, that we almost religiously spent playing soccer and eating ceviche together. I learned how to be a better man because of you, and I cannot wait to show you more of what I can do as a person and a researcher in the future. Mom, I cannot repay you for instilling the curiosity and passion for the life sciences and math in me, from an early age. You were my teacher and my inspiration to become a scientist and it is clear to me that I will be able to enjoy this passion of mine because of the time and work you put into raising me. Emma, honestly words are never going to be enough to capture how much having such a patient and understanding sister such as yourself has helped at all points in my life. The time we spent together on our night drives, listening to music and searching for the best snacks in the city, was the best part about coming home. I am proud to be your brother.

Lastly, I would like to thank my family and friends who supported me through my less-flattering moments all the way to the beginning of my upcoming PhD, especially my aunt Ximena. Ximena, I hope I can repay your dedication and generosity by being able to share with you, my successes at every new and exciting milestone in my career. To my friends Amanda, Taelor, Norman, Wing, Vincent, Brandon, Tiffany, Natalie, Scholar, Tony, and Zainab, I could not have gotten here without all of you! The memories of the sleepless nights, celebrations, struggles and laughs we shared throughout my undergraduate and graduate studies I will never forget. I am quite aware of my great fortune in having all of you in my life, not just as my best friends, but as part of my family. I cannot wait to celebrate more of our future milestones together.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

FEL	Free Electron Laser
FELIX	Free Electron Laser for Infrared Experiments
FTICR	Fourier Transform Ion Cyclotron Resonance
IRMPD	Infrared Multiple Photon Dissociation
IVR	Intramolecular Vibrational Energy Redistribution
G16	Gaussian 16
GV6	GaussView 6.0
dA	Deprotonated Adenine
dG	Deprotonated Guanine
dC	Deprotonated Cytosine
dT	Deprotonated Thymine
dU	Deprotonated Uracil

Chapter 1

Introduction

Studies on the characterization of biologically relevant cyclic species provide us with the foundation to both develop and improve upon our understanding of fundamental biological processes and biochemical mechanisms intrinsic to all life.[56-57] The genetic material that is essential to all life, nucleic acids, is composed of sugars, a phosphate group and nitrogenous cyclic species called nucleobases.[1-3, 56-58] Adenine, cytosine, guanine, thymine and uracil are examples of these cyclic nucleobases, and they determine how nucleic acids such as DNA and RNA regulate important biological functions such as DNA replication, transcription, post-transcriptional modifications, and translation via reactions such as protonation and deprotonation.[56, 70]

$$AH(g) \rightarrow A^-(g) + H^+(g) \quad (1)$$

$$A(g) + H^+(g) \rightarrow AH^+(g) \quad (2)$$

Deprotonation and protonation reactions, as shown by (1) and (2) respectively, are dependent on the intrinsic acid-base chemistry of the cyclic species themselves, as well as the thermochemical interactions with the surrounding biological environment.[3-4, 12] Many biological functions, essential to all life, often exist in non-polar biological environments such as intracellular environments.[13-15] To determine the intrinsic thermochemical properties of these essential cyclic species, we must evaluate them within a non-polar environment, free from environmental interference; something that is difficult to

effectively mimic in the solution phase.[11, 27-28, 43] Therefore, it is necessary to characterize these cyclic species, and their respective thermochemical properties, through the determination of their gas-phase acidity, basicity, and other relevant thermodynamic metrics. In particular, the reactivity of the cyclic species can be determined by calculating the enthalpy of reaction (ΔH°_{rxn}) for the deprotonation (1) or the protonation (2) of the cyclic species. By examining biologically relevant cyclic species, in the gas phase, one can determine their intrinsic thermochemical properties, such as their acidity and basicity, free from the interferences of solvents.[16, 31, 42] However, determination of the intrinsic reactivity of these small cyclic species may not be sufficient. Interactions between these biologically relevant cyclic species and ions such as Na^+, K^+, and Cl^-, are of great significance to essential biological functions as well.[23, 25] These ions are necessary in the regulation of essential biological functions such as driving the Na^+/K^+ pumps and creating HCl to kill bacteria.[23, 25] Thus, the characterization of these gas-phase cyclic species must be further studied through investigations into their clustering behaviour with metal ion species to examine their complexing favourability. The goal of this investigation is to outline the thermochemical characterization of small biologically relevant anionic cyclic species and their complexes with metal ions, in the gas phase, to determine their potential as effective complexing agents. More specifically, this thesis will investigate the gas phase clustering potential of the deprotonated nucleobases with cationic sodium atoms. By determining the effectiveness of these cyclic species as clustering agents, the goal is to produce foundational data in the form of gas-phase clustering behaviour that can be utilized and translated to serve future medically focused endeavours such as drug-substrate design.

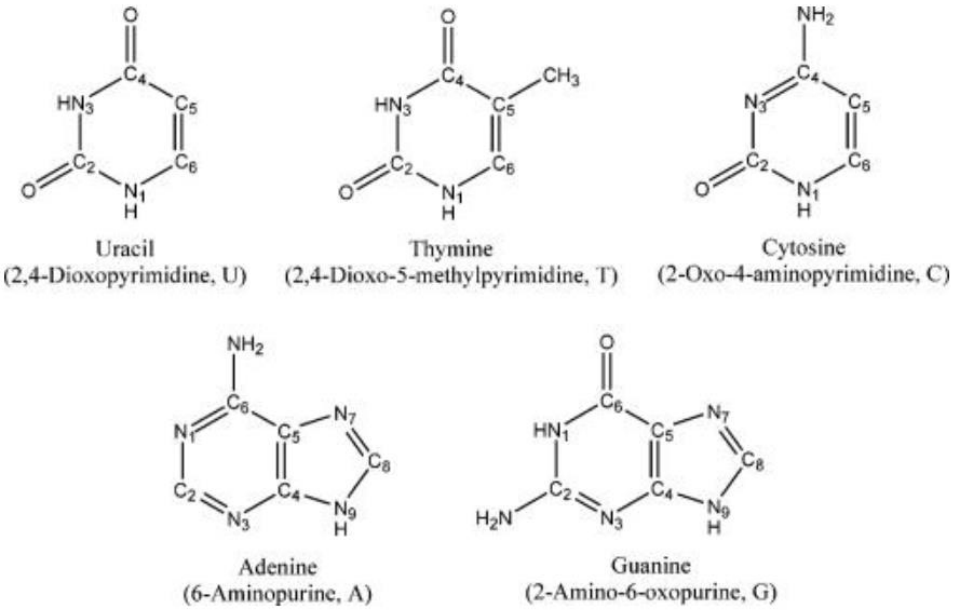


Figure 1: A summary of the neutral nucleobases used in this thesis and the labelling scheme used to evaluate geometries and reaction potentials.[7]

The nucleobases under investigation in this thesis are given in Figure 1. These nitrogenous cyclic species are essential components in regulating essential bodily functions such as ATP production and protein folding.[26, 56-57]

Clusters between the cyclic species and the metal ions are investigated in degrees of increasing complexity, by incrementally evaluating the effects of adding a monomer of the cyclic species and the metal ions. Previous computational and experimental investigations of compounds such as hydrochloric acid (HCl), have allowed for the quantification of HCl's strength as an acid in the gas-phase as well to evaluate its relative strength to other common haloacids.[66, 71] The gas phase acidity of these acids was then used to assist in characterizing their potential as proton donors in several inorganic and organic synthesis reactions in solution.[9, 48, 62-63] More specifically, this method of computational analysis has been employed in investigations of protonated cyclic species

to determine their relative gas phase acidities which allows comparison to that of other well-documented organic acids such as CH_3CO_2H.[17, 54, 65-66] As these biologically relevant cyclic species were able to act as strong acids in the gas-phase, they may have the potential to act as bases when deprotonated and demonstrate favourable complexation with cations, as well as anions. Therefore, the investigation on these species will focus on, but not be limited to, interactions between the deprotonated cyclic species and a sodium cation Na^+. Na^+ was chosen as the metal ion of most interest due to its biological abundance in the human body and its relevance in moderating essential functions such as Na^+/K^+ pumps.[24-25] With a combination of experimental and computational methods, as outlined in Chapter 2, the gas-phase acidities and thermodynamic properties of the cluster species studied are then experimentally investigated through electrospray ionization (ESI), mass spectrometry (MS), paired with infrared multiple photon dissociation (IRMPD) spectroscopy, and computationally through calculations involving Gaussian16 (G16) and GaussView06 (GV6).[6, 8, 18-20, 46]

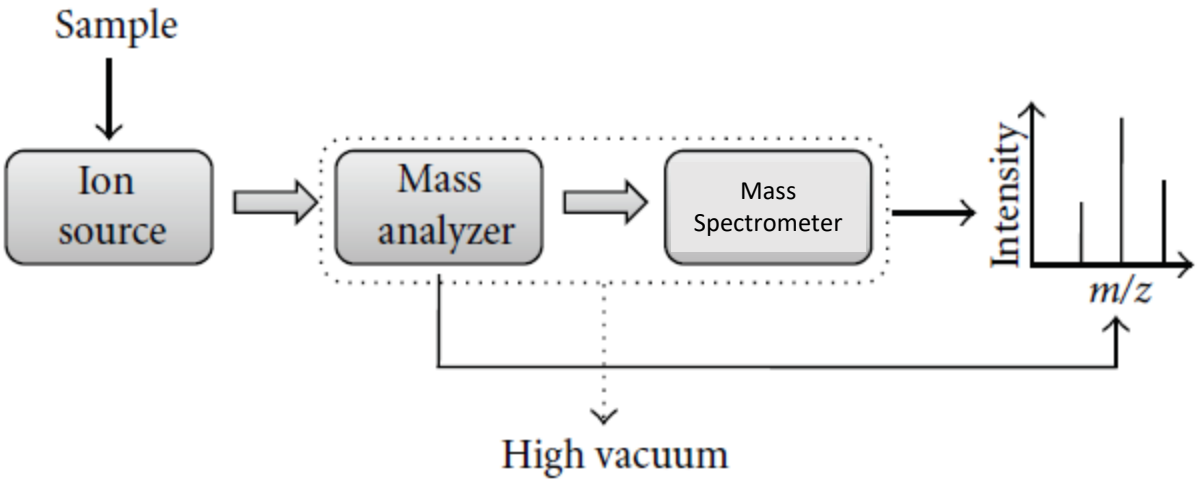


Figure 2: A simplified graphic representing the process of producing and analyzing gas phase ions.[46]

Figure 2 demonstrates the basic pipeline of processes involved in creating and analysing gas phase ions produced by ESI. Chapter 3 highlights the use of IRMPD and computational analysis to elucidate the structures of complexes for each of the deprotonated nucleobases with sodium cations. In addition, Chapter 3 will explore the gas-phase acidities of the nucleobases to evaluate their relative acid strength relative to other well documented acids. In doing so, Chapter 3 also addresses two very important questions: does the solution phase pK_a of well-documented acids reliably reflect the relative strength of acids in the gas phase, and are deprotonated nucleobases effective complexing agents? By determining the gas-phase properties of the nucleobases, it is possible to extrapolate their potential as reactive species carriers, and by extension, their potential as medical charged-species transporters.[22, 46, 48, 50, 55, 69]

Unfortunately, due to the Sars-COVID-19 (COVID19) pandemic, experimental data for these two groups of compounds was limited. However, an extensive computational examination of their complexing potential has been carried out.

Chapter 2

Methods

2.1 Computational

The thermochemical calculations performed in this thesis, namely geometry optimization and frequency calculations, have been effected using *GAUSSIAN16* (G16) and *GAUSVIEW06* (GV6) software. These software packages allow the investigator to create a three-dimensional (3D) chemical system, of the molecules of interest, to provide an appropriate initial candidate structure, prior to optimization. These initial candidate structures can often be intuited by evaluating the chemical system based on fundamental chemical principles.

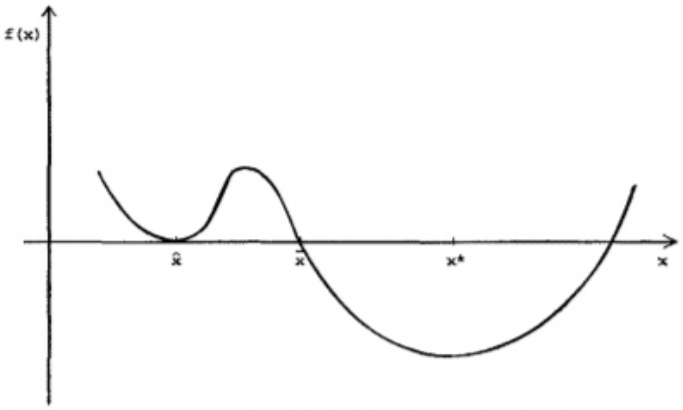


Figure 3: A graph highlighting the two possible types of minima in an energetic calculation; a local minimum and a global minimum.[60]

As analogized in Figure 3, by providing an initial educated guess structure to the software, the risk of optimizing to a transition state or a minimum energy that is not the global minimum (i.e. local minimum), is greatly reduced.[15, 17-22] Using fundamental chemical principles may allow us to intuit a structure that lies closer to the global minimum.

2.1.1 Electronic Structure Calculation Methods

Electronic structure calculations provide approximate solutions for the Schrodinger equation, an essential component in characterizing all quantum mechanical problems. In turn, the solution for the Schrodinger equation to a particular chemical system would describe the energy of that system.[29-30, 32-34]

$$\hat{H}\Psi = E\Psi \tag{3}$$

The Born-Oppenheimer (BO) approximation can be used to greatly simplify the system and allow for an approximate solution to the Schrodinger equation.[72-73] The BO approximation utilizes the concept which describes the difference between the speed in the motion of electrons compared to nuclei. It presumes that due to the difference in mass between an electron and nuclei, electrons can adjust their motion almost instantaneously in response to changes in the position of the nuclei such that the nuclei are effectively given a static nuclear position, on the electronic time scale.[72-73] Through this approximation, we can computationally determine the thermochemical properties of a system using both *ab initio* and semi-empirical methods.

2.1.1.1 Density Functional Theory

The computational quantum mechanical modelling method of density functional theory (DFT) provides researchers with a useful tool to determine the electronic structure and the thermodynamic properties of molecular systems.[31, 59] At a fraction of the computational cost of other higher levels of theory, such as the second order Møller-

Plesset perturbation theory (MP2), continuous updates allow DFT methods to remain comparable in accuracy.[29-30] DFT methods can achieve such accuracy by calculating the electron correlation effects with a general, spatially dependent, functional of the electron density as opposed to treating individual electrons independently. In its simplest form, the electron density functionals can be determined through the aggregation of the kinetic energy, electron-nuclear interaction energy, Coulomb energy, and the exchange-correlation energy.[32-33] There are several variations of the basic DFT methods which incorporate additional terms that can uniquely address different systems of interest, each with their own shortcomings and strengths.[37-39, 44-45] In some cases, however, to achieve a greater accuracy, at little to no additional computational cost, components of different functionals can be combined to create a hybrid functional.[59, 61]

These hybrid functionals incorporate Hartree-Fock (HF) theory with the electronic density approximation and have been used extensively for their impressive speed and accuracy in characterizing a wide range of chemical systems. One hybrid functional used extensively throughout this investigation is the B3LYP (Becke 3-Parameter, Lee, Yang, and Parr) method.[53, 74] B3LYP mixes the HF theory with local density functional theory and has been proven to show adequate performance with sufficient accuracy across a variety of chemical systems and properties.[53] In addition, B3LYP has remained one of the most popular DFT methods over the years for its ability to be more efficient and accurate than higher-order systems such as Second Order Møller-Plesset Perturbation Theory (MP2) when determining infrared vibrational spectra.[32-33] These computationally determined spectra rely on the sufficient accuracy of the method used, and without proper initial-guess candidate structures to optimize with the hybrid DFT functions, the

likelihood of optimizing to a transition state or a non-globally minimum structure. Therefore, it was often necessary to determine initial structures for evaluation with DFT, by chemical intuition.

2.1.1.2 Second Order Møller-Plesset Perturbation Theory

Modern DFT methods and their continuously improving accuracy are a luxury that was preceded by higher-level theory methods, such as the Møller-Plesset perturbation theory (MPn), which required more computational power.[29-30, 32-33] These *ab initio* methods were among the most reliable means to increase accuracy at a proportional increase in computational cost.[29-30, 32-33] The MP2 method, more specifically, is a particularly popular method among researchers interested in increasing accuracy without sacrificing much computational power. MP2 is a sufficiently high level of theory, capable of characterizing the electronic and thermochemical properties across a variety of chemical systems.[31, 53, 59] Due to its widely accepted robustness in accuracy, and the marginally small cost of computational time, this method will be used extensively throughout this thesis to determine the single point electronic energies of various systems. In addition, the MP2 calculations will include all the electrons for each system, both the core and valence electrons.

2.1.2 Evaluating the Accuracy of Electronic Structure Calculations to Experimental Data

IRMPD experiments can be used in combination with computational means to determine detailed structural information about a particular species in the gas-phase, including the

thermochemical properties and single-point energies of the investigated species.[18-22] The combination of IRMPD spectroscopy and electronic structure calculations greatly aids in elucidating gas phase ion structures and energetics relevant to this investigation.[31-59]

To characterize structures of interest, computational analyses determining the tautomer with the most energetically favourable structure will be essential. Through a process known as optimization, a system of interest is mathematically optimized towards minimizing the Gibbs energy of that system.[34, 42, 59, 60] A lower Gibbs energy directly correlates to a more stable structure with respect to dissociation; therefore, computationally derived minimal energy structures provide suggestions for the most energetically favourable (stable) structures. Geometric optimization will therefore computationally determine, based on the initial candidate structure, the chosen level of theory, and the basis sets, a minimum energy structure for the system under investigation.[31, 35-36, 59] In conjunction with the optimization calculations, frequency calculations can be performed to determine the thermochemical properties for the optimized chemical system. These optimization and frequency calculations yield absolute values for the thermochemical properties, as well as vibrational frequencies, for the optimized system. The vibrational frequencies generated can then be used to create a theoretical infrared (IR) spectrum by applying a Gaussian smooth function over the individual vibrational frequencies. These computationally generated IR spectra can then be used to evaluate the identity of the spectra produced from experimental procedures by qualitatively matching the theoretical spectra to the experimental spectra.[5, 10, 41, 58, 69] Producing experimental spectra, creates a fingerprint of the chemical system that uniquely identifies one, or a group of, isomers that may have been experimentally

produced.[60, 63 67, 69] Determining theoretical spectra that match the experimental spectra would then facilitate the identification of the molecules observed experimentally. Moreover, by comparing the theoretical and experimental spectra, the dominant isomer of a particular chemical system can be identified. The identification of a particular chemical system can aid in determining intermolecular and intramolecular characteristics which can provide explanations to the relative isomeric abundances in nature or experimentally.

In a mixture of isomers, the relative abundances of these isomers can be reliably predicted by the differences in their Gibbs energies; assuming the ions maintain a Boltzmann distribution of internal energies.[31, 35, 36, 59] Molecules with higher energy are less likely to form compared to their isomers, or related complexes, with lower energy. To compare isomers, the relative Gibbs energy of each isomer is determined, with respect to the lowest energy isomer. Isomers with a relative Gibbs energy less than 15 kJ/mol, with respect to the lowest energy isomer, could statistically exist in a ratio of 1:424 at 298 K, based on equation 4 shown below.[31, 35, 36, 59]

$$K = \frac{[B]}{[A]} = e^{\frac{-\Delta G^{o}_{298}}{RT}} \qquad (4)$$

As a result, for this thesis, only isomers less than 15 kJ/mol higher in energy relative to the lowest energy isomer will be considered as plausible experimental structures. Therefore, there exists several isomers of the cyclic species investigated through this thesis, but only the lowest energy structure will serve as the starting point for each monomer of the cyclic species. To maintain this relationship, all Gaussian calculations are defaulted to 298 K, unless otherwise specified. For this thesis, all the calculations were

kept at 298 K, thus, maintaining the relationship between the literature-accepted standard of experimentally plausible structures having Gibbs energies that differ by less than 15 kJ/mol, with respect to the lowest energy.[31, 59] The relative Gibbs energies are calculated by determining the thermochemical properties of a particular system, providing information on the correction to enthalpy (H_{corr}) in Hartrees, the entropy (S^o) in calories per Kelvin-mol, and the single point energies (E_{MP2}) in Hartrees.[44, 65, 70] These thermochemical properties can then be combined to determine other thermochemical properties for a particular reaction, such as the loss or addition of a cyclic species or sodium cation. Enthalpies (H^o) and Gibbs energies (G^o) are thermochemical properties that can be determined using equations 5-10 below.[44, 66]

$$H^o_{MP2} = [H_{corr}{}^0(\text{Hartree}) + E_{MP2}\,(\text{Hartree})] * 2625.5\ (\text{kJ/mol})/\text{Hartree} \quad \textbf{(5)}$$

$$G^o = H^o_{MP2} - [T * S^o(\text{kJ/K*mol})] \quad \textbf{(6)}$$

$$\Delta G^o = G_{(Na)dA}{}^o - (G_{Na^+}{}^o + G_{dA^-}{}^o) \quad \textbf{(7)}$$

$$\Delta H^o = H_{(Na)dA}{}^o - (H_{Na+}{}^o + H_{dA-}{}^o) \quad \textbf{(8)}$$

$$\Delta S^o = S_{(Na)dA}{}^o - (S_{Na+}{}^o + S_{dA-}{}^o) \quad \textbf{(9)}$$

The factor used in equation (5), 2625.5 (kJ/mol)/Hartree, is the necessary conversion factor in translating the theoretical data into a more widely employed units. By using these equations to determine the ΔG^o of a particular reaction, such as the deprotonation of a nucleobase, we can then compare the energies of the isomers of these species and assess the most stable system. The lowest relative Gibbs energy implies the most likely structure to exist.

2.1.3 Defining Electronic Structure Calculations

The accuracy of the electronic structure calculations relies on an appropriately stringent level of theory to be used, as well as the combination of the basis set and method used to characterize the desired system.[44] An optimal mix of accuracy and reduced computational cost must be met, at the discretion of the investigator, based on the type of system under investigation. For the systems under investigation in this thesis a large majority of the calculations were performed with a split basis set, a mix of different basis sets to account for the inclusion of sodium cations, and were otherwise evaluated consistently with one theory-basis-set combination as discussed further in the following sections.[44, 53, 59]

The theoretical spectra used in this thesis are calculated from vibrational frequencies that utilize the harmonic oscillator approximation to quantify the vibrational movement of the bonds in each chemical system. This approximation characterizes the potential energy curve of the bonds in a system as a parabola, however, the true potential energy curve of a bond is not completely represented by a parabola.[31, 37-38, 59] Although the potential energy curve near the equilibrium distance of a bond is well represented by this parabola, limiting the flexibility of the harmonic oscillator approximation, it neglects the bond dissociation threshold.[29, 44, 59] As a result, calculations performed exclusively with a harmonic oscillator approximation have been shown to over-estimate experimental values. To account for this potential overestimation, there is a literature-accepted scaling factor of 0.968 applied to the vibrational frequencies when producing the computational spectra based on the use of B3LYP theory to perform the calculations.[76] Scaling factors vary depending on the level of theory used to perform the calculations and can be used to

correct discrepancies from systematic errors and to better approximate for anharmonicity.

A harmonic oscillator approximation characterizes the entirety of the potential energy function by a parabola and fails to incorporate the anharmonicity, coupling of modes, that exists.[44, 76] The anharmonic oscillator approximation better incorporates the characteristics from the dissociation threshold and allows for greater flexibility in its calculations. As a result, sufficiently small chemical systems may be able to be characterized by higher level theories and methods to better represent the anharmonic frequencies of the system.

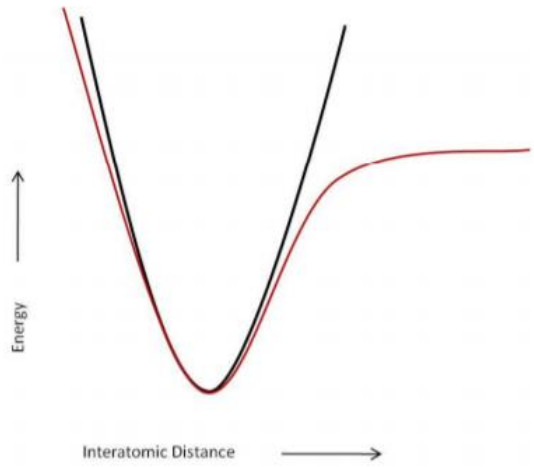


Figure 4: A simplified example describing the differences between the harmonic oscillator approximations (black) and the anharmonic approximations (red).[59]

Optimization and frequency calculations are used to provide the necessary data to determine the thermochemical properties of a system, such as the correction to enthalpy and the electronic energy. It has been demonstrated that the electronic energies determined by optimization and frequency calculations are not sufficient for accurate calculations of gas phase energetics.[29, 32-33] Therefore, a separate calculation at a higher

level of theory and accuracy must be performed to approximate the electronic energy of a species. By using the successfully optimized geometries of the species in question as a candidate structure for the higher order single point calculation, often using MP2, a more accurate representation of the single point energy calculation can be achieved.

2.1.4 Defining Basis Sets for Electronic Structure Calculations

The wave behaviour of an electron can be described by the electronic wavefunction, and alone, has no physical meaning. However, this wavefunction can be squared and used to define the probability of finding an electron, at a specified set of values.

$$\Psi(r, \phi, \theta, t) = R(r)Y_l^m(\phi, \theta) \tag{11}$$

The electronic wavefunction can be expressed by equation (11) and can be separated into the radial wavefunction (R(r)) and the angular wavefunction (Y_l^m).[44] The square modulus of the angular wavefunction defines the probability of finding an electron at any point on a ray defined by (Φ, θ).[44] In other words, the square modulus of the angular wavefunction will determine the direction and shape of the orbital. The Schrodinger equation defines 3 quantum numbers (*n, l,* and m_l) which, when combined, characterize a particular orbital (location of electron) in a system.[42, 44] When performing electronic structure calculations on a chemical system with more than one electron, the basis set used in these calculations represents a mathematical interpretation of all the combined orbitals in a system. Larger basis sets can better approximate the orbitals in a system due to the addition of

parameters which provide electrons with more spatial freedom. The basic forms of the basis sets utilize linear combinations of Gaussian functions to determine the orbitals of a system and are usually named by the number and type of basis functions used.[53, 76]

There are various types of basis sets, each with its own advantages and disadvantages, that are distinguished based on the size of the basis set and the degrees of freedom imposed on the electrons' location in space. For example, a split-valence basis set, such as 6-311G, denotes a basis set where the amount of basis functions per atom increases and can account for orbitals of different sizes.[53, 76] John Pople established the notation for split-valence basis sets and, as a result, are sometimes referred to as People-style basis sets.[32] The notation of the 6-311G basis set describes the six sets of basis functions used for the core electrons and three sets for the valence electrons.[74] More specifically, the valence electron basis sets include three basis functions for the first set, and one basis function for each of the two remaining sets.

These basis sets can be further expanded with the addition of polarization parameters that provide the orbitals more degrees of spatial freedom and allow for the orbitals to change shape.[53, 74, 76] This is achieved through the addition of orbitals with more angular momentum than needed for the ground state, thus, ridding the orbitals of their limitations in shape. Polarization parameters such as d, p, and f, permit the addition of d functions on carbon atoms, p functions on hydrogen atoms and f functions on transition metals, respectively.[42, 53, 76] In addition to the polarization parameters there are other parameters, such as the diffuse functions, that can be added to basis sets to expand the range in which the orbitals can exist in space.[42, 53, 76] These diffuse functions are especially

important for systems that include electrons relatively far from the nucleus, such as with ions and lone pairs. The addition of a single + denotes diffuse functions being added to heavy atoms and the addition of the ++ denotes diffuse functions being added to both heavy atoms and hydrogen atoms. These parameters can be combined to increase the size of the orbitals, the shape of the orbitals, as well as the range of electrons in which they can reside to still be included in the electronic structure calculations. The basis set 6-311++G(d, p) is an example of a split-valence basis set which incorporates many of the aforementioned parameters, and is the foundational basis set used in the majority of this thesis.

2.1.4.1 The Efficacy of Split Basis Sets in Determining Electronic Structure Calculations

There are a wide variety of singular basis sets and levels of theory that can be used to evaluate the entirety of a given chemical system, however, some of these chemical systems can be better represented by a combination of these singular basis sets. These combinations of basis sets, used to characterize one chemical system, are often referred to as split basis sets. In a split basis set, the valence atomic orbitals of each atom can be evaluated at specific levels of theory. There are trade-offs in terms of computational cost, accuracy and this thesis generally utilizes the combination of the B3LYP level of theory with a split basis set to evaluate the non-metallic atoms at 6-311++G(d,p) and the metallic sodium cation at an augmented basis set aug-cc-pCVTZ.[31, 59, 74, 76] The single point energy calculations were performed at MP2 level of theory, with a split basis of 6-311++G(d,p) and aug-cc-pCVTZ for the cyclic systems involving sodium cations, evaluated in Chapter 3. The theoretical spectra produced throughout this thesis had a scaling factor of 0.968

applied to the frequency calculations as this scaling factor has demonstrated its effectiveness at balancing computational cost and accuracy.[76] Although diffuse functions for the hydrogen atoms have been shown to barely influence the accuracy of the calculations rarely significantly, they were included to keep a standard of accuracy consistent throughout this thesis.

The augmented basis set, aug-cc-pCVTZ, uses diffuse functions on both the heavy and hydrogen atoms similarly to that of the 6-311++G(d,p) basis set.[31, 59] This augmentation of diffuse functions allows for the orbitals to change in size depending on the computational requirements of the calculation. The correlation-consistent polarized core/valence triple zeta (cc-pCVTZ) denotes the increase in the zeta factor, which correlates directly to an increase of accuracy at an increased computational cost, compared to basis sets that only use a single theta or a double theta factor in their calculations.[45, 75] These triple zeta calculations create each molecular orbital by linearly combining three different basis functions in each orbital. The correlation consistent component of this basis set will add more layers of functions to a central set of HF functions, each contributing a similar amount of energy, to increase the accuracy of the calculations. In addition, the core/valence (CV) portion of the basis set defines the inclusion of both the core and valence electrons in the correlation energy, however, this can be similarly achieved by adding the keyword “full” as a subset of the MP2 level of theory.[45, 53, 59, 75]

2.2 Experimental

2.2.1 Electrospray Ionization

Producing ions in the gas phase has been greatly facilitated by the introduction of a process called electrospray ionization (ESI). This process involves the aerosolization of a volatile solution, containing the molecules of interest, by a charged capillary. Volatile solvents are required for this process as it helps to reduce the surface tension and to facilitate evaporation of the solution prior to aerosolization.[40, 46] The charged capillary carries a voltage between 2-5 kV that promotes the formation of a Taylor cone in the solution once a certain threshold voltage is achieved.[47-49] This Taylor cone is the result of the solution surface tension being inverted and releasing a stream of liquid containing droplets of extremely small diameters, usually 1 μm to 10μm, containing roughly 150,000 analyte molecules per 1 μm in diameter.[47]

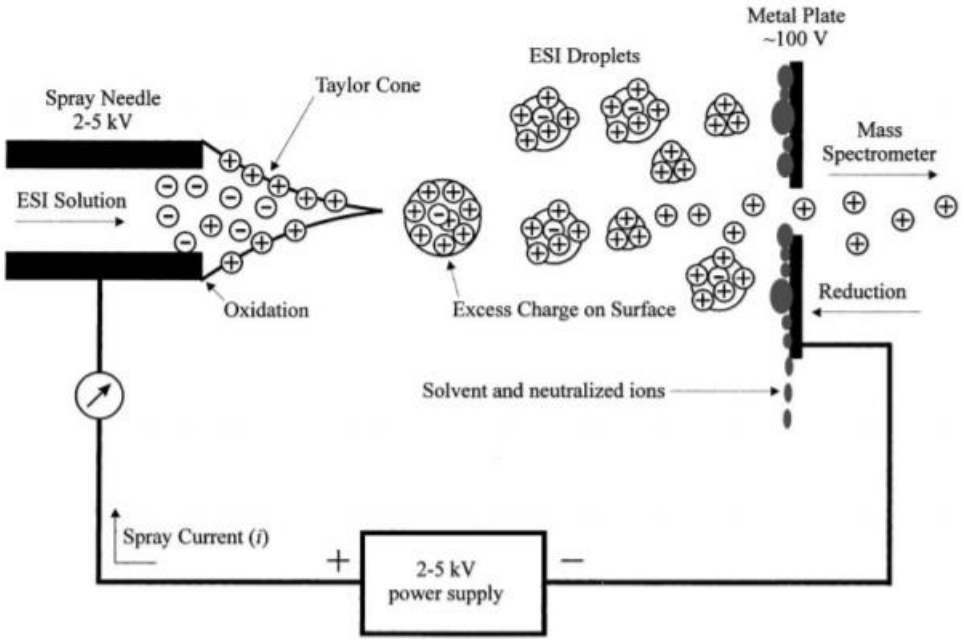


Figure 5: A graphical representation of the process of ESI to aerosolize the analyte molecules. It makes use of the Taylor Cone to focus analyte droplets into a fine stream.[48]

These ESI droplets, as seen in Figure 5, are then passed through a high-temperature chamber filled with nebulizing gas, often N2, in temperatures between 350 K and 500K, to evaporate the solvent by thermal evaporation.[48] As the ESI droplets pass through the chamber they continue to evaporate and shrink until they reach the Rayleigh limit. The Rayleigh limit, in the context of ESI, describes the phenomenon where droplets containing a certain amount of analyte shrink by evaporation until it reaches a physical limit where the surface tension of the droplet can no longer support the amount of charge in that droplet.[46-49] Once this limit is reached, the droplets then undergo Coulombic explosions and break up into smaller droplets with a new Rayleigh limit.[49] The ESI droplets are directed to the next part of the instrument via counter electrode that surrounds the opening to the ladder part of the instrument.

The majority of experimental work presented in this thesis involves clusters formed by association of anionic molecules with cationic sodium. To observe the formation of a positively charged cluster, as shown in Figure 5, a positive charge is passed through the capillary to attract undesired anionic species away from the spout of the needle.[24, 46] Moreover, the cations will be repelled by the positive current and subsequently ejected from the needle, forming a Taylor cone, and drive the positively charged ions through towards the evaporating chamber.[46-49] These positively charged ions are then driven towards the outlet of the evaporating chamber by applying a negative charge at the counter-electrode, near the end of the evaporating chamber. By combining ESI with a mass spectrometer, somewhere further downstream in the analyzing instrument, samples that are normally difficult to vaporize for analysis such as non-volatile and polar samples, can be effectively observed.[46-49]

2.2.2 IRMPD

Infrared multiple photon dissociation (IRMPD) spectroscopy is a powerful technique that has demonstrated its value in facilitating elucidation of molecular structures in the gas phase by allowing researchers to determine the vibrational modes of molecules.[10] The aggregate of any molecule's vibrational frequencies across a range of wavelengths produces a vibrational spectrum. Any vibrational spectrum that covers the range between 600 and 2000 cm^{-1} is often called the vibrational fingerprint of a molecule.[5, 10] These vibrational modes that are unique to a specific molecule or chemical system act as "fingerprints" that can be used to identify experimental spectra that would be difficult to analyze without other useful information.[69] The experimental work presented in this thesis was performed using IRMPD and was helpful in determining the vibrational spectra of the gas-phase ions under investigation. IRMPD is therefore a very useful method of observing the chemical signatures of ionic species, when combined with computational studies, and is the main technique used in this thesis for elucidating ionic structures.

2.2.2.1 Free Electron Lasers

Fragmentation of a species can only occur once a dissociation threshold of energy is achieved.[18-22] To meet this threshold, enough energy must be absorbed by the species through the absorption of a sufficient number of high-energy photons. In order to generate the required number of photons to induce fragmentation via the IRMPD process, extremely powerful lasers are required. Several types of lasers have been previously used in tandem with the IRMPD process, however, certain physical limitations to these lasers

restrict the feasibility of their use.[68] For example, gas discharge lasers lack the tunability to be effective lasers to use with IRMPD. There exist other similar types of tunable lasers capable of generating the number of photons required to induce fragmentation via IRMPD, such as the optical parametric oscillator (OPO), lasers which are tunable, however power and pulse duration become points of concern for them.[68] Therefore, the various types of lasers that can be used in tandem with IRMPD are relatively limited, thus, the free electron laser (FEL) has been our option of choice.[55, 59, 68]

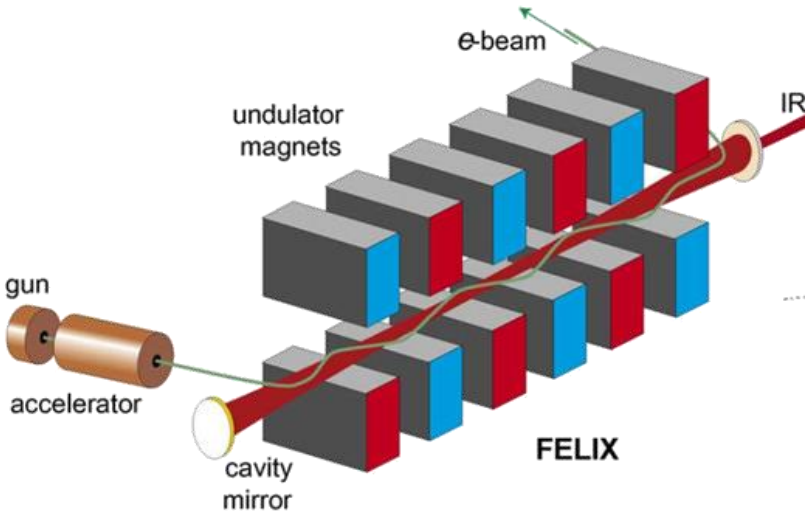


Figure 6 – A schematic representation of the process of IRMPD in a FEL.[10]

FEL's produce high-energy photons through coherent synchrotron radiation achieved through the oscillation of free electrons produced by an electron beam, as shown in Figure 6. The electrons in the beam are initially treated by an accelerator of roughly 10-50 MeV before reaching the undulator, supplying high-energy radiation like that of a linear-accelerator-based light source while maintaining optical characteristics of a typical laser such as: coherence, directionality, monochromatic and high-intensity. Undulators are a series of magnets, with periodic poles, that promote the emission of radiation by forcing

the electron beam to undergo an oscillatory motion. The initial emitted radiation can then be magnified by cavity mirrors to form more coherent radiation, resulting in the formation of "electron bunches", otherwise known as photons.[10, 50]

The tunability of the FEL is the result of modifications in the distance between the magnets in the undulator. FEL's are effective at inducing fragmentation of ionic species through IRMPD and can modulate several parameters to allow for the reliable production of spectra within a certain range of wavelengths.[10, 77] Once the photons have been generated, they travel through focusing mirrors until they reach the mass analyzer to identify both the parent and fragment ions for each of the species.

2.2.2.2 IVR

Ions that are held in the ion trap are exposed to the photons produced by the tunable FEL to induce ion fragmentation. Ions exposed to high-energy photons of a particular wavelength can be excited from their ground state to a higher energy state, from v = 0 to v = 1.[51-52, 59] The transition from v = 0 to v = 1 can allow for the accumulation of energy in the ion, making the ion more unstable. This instability can often be corrected by releasing the energy as light, but in the case of the ions under the process of IRMPD, the energy gained by being excited to a higher state can be transferred to other vibrational modes of the species. This phenomenon of redistributing the gained energy from the excited state to other available vibrational modes defines the IVR process.[51-52] After one instance of IVR has been performed, the ion returns from its excited state of v = 1 back to its ground state of v = 0 in the absorbing mode, allowing the ion the possibility to absorb more energy

from another incoming photon. This cycle of energy gain and redistribution can continue in each ion, until a thermodynamic threshold is achieved, and the original parent ion begins to fragment as seen in Figure 7.[51-52]

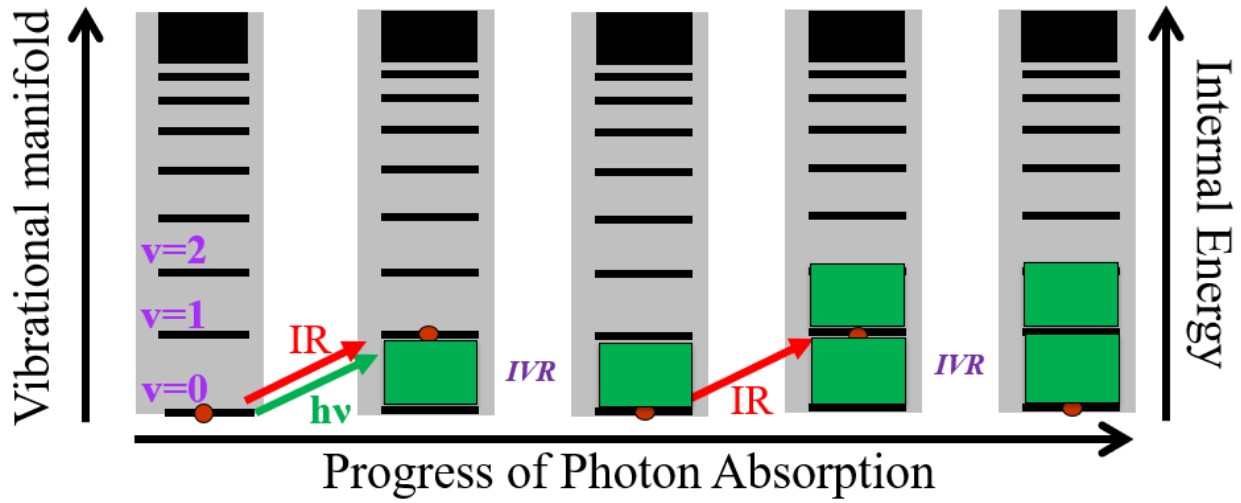


Figure 7 – The process of intramolecular vibrational energy redistribution (IVR).[78]

This graphic, in Figure 7, is useful in visualizing the process of successive IVR energy redistribution to other accessible vibrational modes.[31, 51-52, 59] FELs will fire single pulses of photons for approximately 1 ps, once every 16 ns within a 10 μs timeframe, for a total of 625 pulses per 10 μs.[10] The timing between the firing of pulses is a result of the technical limitations of the laser, but provide enough time for the ion to redistribute the gained energy to other vibrational modes before starting another cycle of pulses. Each of these pulse cycles occur at a specified wavelength and repeat on that same wavelength roughly ten times before changing wavelengths.[78] Each completed scan should span roughly 40 ms and should be scanned multiple times before changing wavelengths. [10]

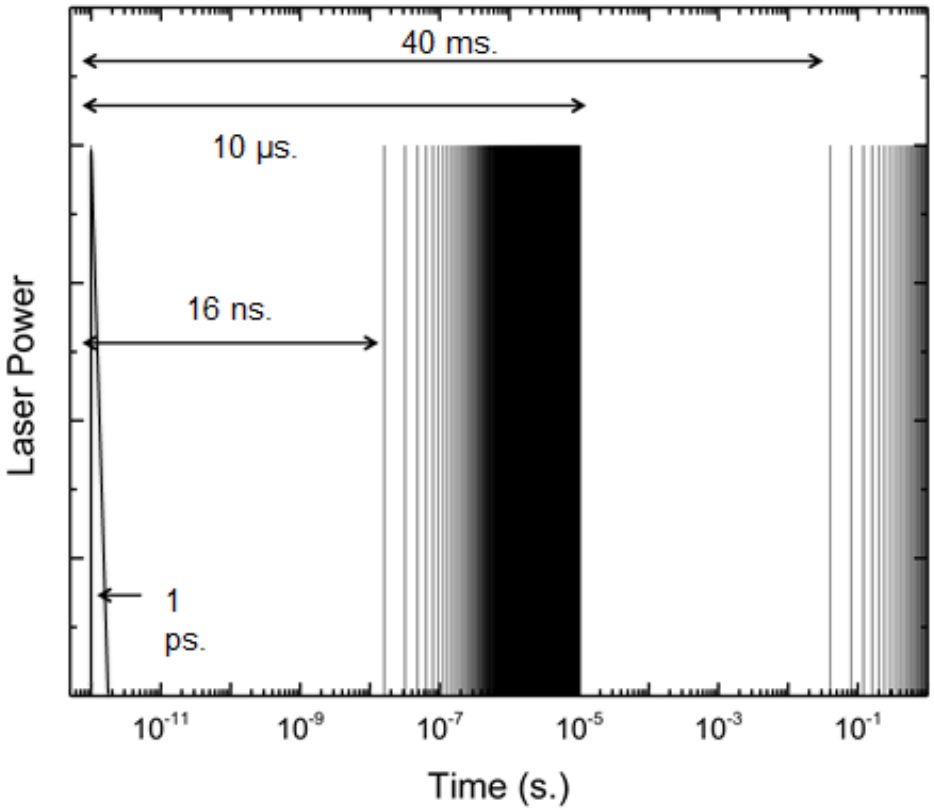


Figure 8 – Pulse train of FEL at CLIO inducing IVR. Pulse trains allow for dissociation of the parent species to be measured by a depletion in the parent peak, or an increase in the detection of the fragment peaks.[78]

In summary, IRMPD interrogates ions formed by the ESI process. IRMPD achieves this by depositing multiple high-energy photons into the ions generated by ESI, in an ion trap, until a thermodynamic threshold is achieved and induces fragmentation of those ions via intramolecular vibrational energy redistribution (IVR).[51-52] Therefore, to create a vibrational spectrum, a sufficient amount of energy must be deposited into the chemical system to induce fragmentation.[18-22]

Chapter 3: Gas-phase Acidities and Complexation of Nucleobases

3.1 Introduction

Nucleobases have been studied numerous times in the gas phase to assist in characterizing their significance to essential biological functions in all living things.[1, 4, 7] For example, nucleobases serve as the foundation for the genetic code which governs the functionality and variety of living things. Nucleobases are of specific scientific interest as they play integral roles in the execution of biological functions and biochemical reactions in the human body. Additionally, there have been studies exploring nucleobase-mediated reactions that have been used to facilitate the synthesis of a variety of chemical compounds with medical applications, such as 5-fluorouracil and 5-fluorocytosine.[2] Protonated nucleobases have been shown to act as effective complexing agents and as strong acids in both the gas phase and solution phase.[54, 58] However, the neutral nucleobase monomers themselves have been shown to be relatively strong acids in the gas phase.[7, 11-16]

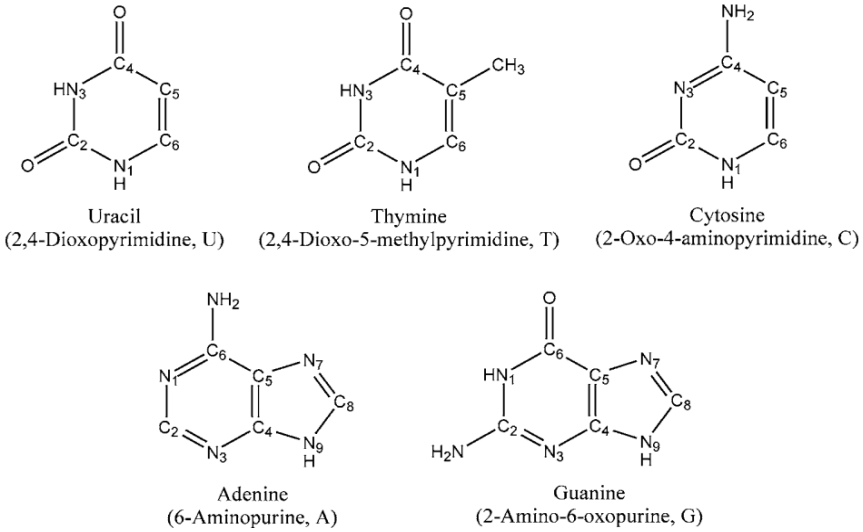


Figure 9: Neutral nucleobase monomers and their atomic positions for reference.[7]

The atomic labelling used in this thesis, given in Figure 9, facilitates descriptions and analysis used in later sections. The reported gas-phase acidities (GPAs) for the nucleobases, shown in Table 1, represent the Gibbs energy change for the loss of the N9 proton, in both adenine and guanine, and the N1 proton in cytosine, uracil and thymine. Table 1 demonstrates how the reported nucleobase GPAs are comparable to those of the GPAs for other well-documented acids in the gas phase, such as HCl and acetic acid. [11-15, 62-65] However, it is of interest to note the disparity in the relative strengths of these acids in the solution phase, as demonstrated by the pK_a values. HCl and acetic acid are particularly strong acids in the solution phase, however, the nucleobases are not nearly as strong in solution. Interestingly, the acidity of each of the nucleobases, both in the gas phase and in solution, is stronger than that of water.

Table 1: Gas-phase acidities and solution-phase pKa's.[11-15, 62-65]

Nucleobase	pK_a	Reported Gas Phase Acidity (kJ mol^{-1})
Adenine	9.8	1401
Cytosine	12.2	1423
Guanine	10.0	1389
Thymine	9.9	1393
Uracil	9.5	1393
Ubiquitous Molecules	**pK_a**	**Reported Gas Phase Acidity (kJ mol^{-1})**
H_2O	14.0	1595
Halogen Acid	**pK_a**	**Reported Gas Phase Acidity (kJ mol^{-1})**
HF	3.1	1522
HCl	- 6.0	1354
HI	- 9.5	1291
Organic Acid	**pK_a**	**Reported Gas Phase Acidity (kJ mol^{-1})**
CH_3CO_2H	4.8	1429

Based on previously reported data, the order of acid strengths of the nucleobases in solution does not parallel the order of their acid strengths in the gas phase. Due to this phase-dependent property of nucleobases, it would be of great importance to further

investigate the thermochemical properties of the nucleobases to better understand their gas-phase characteristics. Therefore, the gas-phase characterization of the nucleobases should not be limited to their neutral or protonated forms. By investigating the thermochemical interactions of the nucleobases as anions in the gas phase, novel applications of the deprotonated nucleobases may be revealed. Therefore, investigating the gas-phase basicity of the nucleobases, and their ability to complex with cationic sodium, the goal is to produce foundational data on novel systems and geometries that can assist medical-based research in the future. It would stand to reason that if the protonated nucleobases could act as both a strong acid and an effective complexing agent in the gas phase, deprotonated nucleobases may in turn act as strong bases and equally strong complexing agents.

3.2 Methods

Deprotonation of a molecule involves the loss of a proton via ion-molecule reaction where an appropriate anion can abstract a proton from a neutral molecule. In each of the nucleobases, there are several sites from which a proton can be abstracted, leading to several potential isomers of the deprotonated nucleobase monomer. A computational investigation into the thermochemical energetics of these isomers can determine which of the isomers are the most energetically favourable. Isomers with the lowest Gibbs energy are the most energetically favourable and are more likely to be observed.

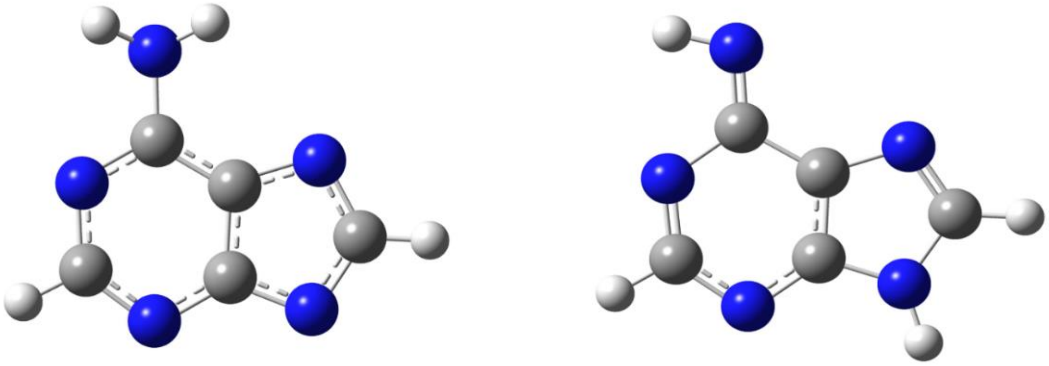
To characterize the complexation of the deprotonated nucleobases and sodium cations, with increased accuracy, properly optimized deprotonated nucleobase monomers must

first be established. By computationally comparing the relative Gibbs energy for the isomers of each nucleobase, the most energetically favourable isomers of each nucleobase were further explored.

Relative Gibbs energy (ΔG^{o}) is determined by comparing the absolute Gibbs energies (G^{o}) for the isomers under investigation, using Equations 5-7 as previously listed, and subtracting them by the lowest G^{o} among the isomers.

$$G^{o}_{\text{ isomer 1, high}} - G^{o}_{\text{ isomer 2, low}} = \Delta G^{o}_{\text{ isomer 1 relative to isomer 2}} \qquad \textbf{(10)}$$

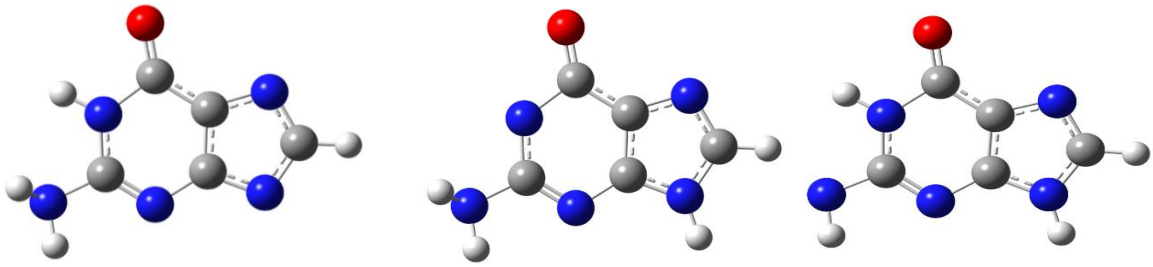
Equation 10 is used as the foundation for comparing the Gibbs energies of the various isomers under investigation and determine the energy difference of each isomer relative to the lowest, most energetically favourable, energy structure. These values can then be used to determine relative populations of the isomers using Equation 4.



ΔG° = 0.000 **ΔG° = 170.371**

Figure 10: Calculated lowest energy isomers for deprotonated adenine. Relative Gibbs energies of deprotonation (298 K) are shown in kJ/mol

As shown in Figure 10, the two more energetically favourable isomers are the most probable isomers for a deprotonated adenine. By determining the Gibbs free energy for each isomer of adenine, a clear difference in their relative stability can be observed. Based on the difference in relative Gibbs energy, the abstraction of the NH2 protons requires sufficiently more energy, such that it would be very unlikely to exist is in significant quantities. The increased energy requirement for abstracting the protons from the NH2 is a characteristic shared among the purines.



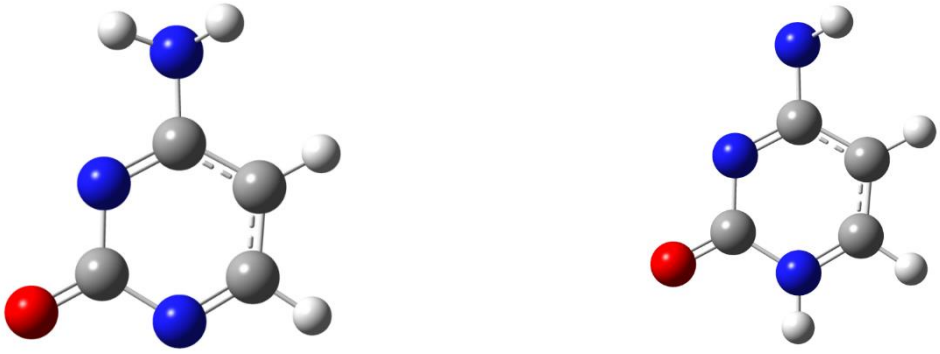
ΔG° = 0.000 **ΔG° = 84.330** **ΔG° = 106.662**

Figure 11: Calculated lowest energy isomers for deprotonated guanine. Relative Gibbs energies of deprotonation (298 K) are shown in kJ/mol.

Figure 11 highlights the Gibbs energy required to abstract the protons in three different positions. The most stable isomer of deprotonated guanine appears to be the result of abstracting the N9 proton, like that of deprotonated adenine. In addition, abstraction of the NH2 protons requires the most energy compared to the other isomers of deprotonated guanine, as with deprotonated adenine.

In the case of the pyrimidines, the differences in energetic favourability are not as consistent. Cytosine is a pyrimidine that differs from the other pyrimidines in that it contains an NH_2 group instead of a double-bonded oxygen on the C4 carbon. This

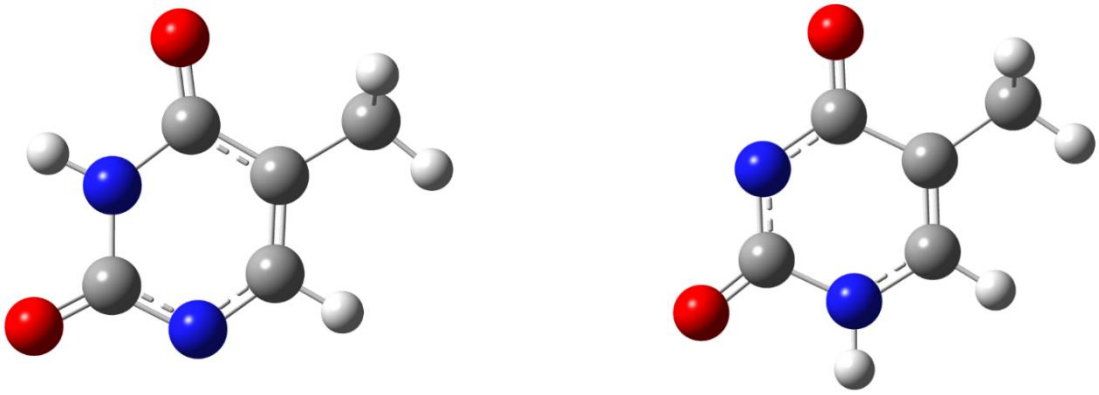
difference is more pronounced in the differences between the isomers of cytosine, than that of uracil and thymine.



ΔG° = 0.000 **ΔG° = 143.399**

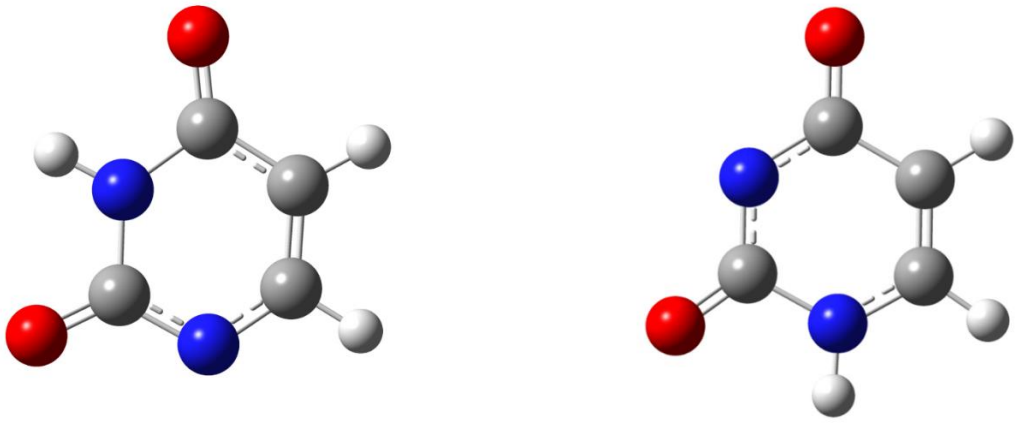
Figure 12: Calculated lowest energy isomers for deprotonated cytosine. Relative Gibbs energies of deprotonation (298 K) are shown in kJ/mol.

The NH_2 protons in cytosine, like the NH_2 protons on the purines, are the least favourably abstracted and require greater than 100 kJ/mol of energy to abstract. The most favourably abstracted proton from cytosine appeared to be the N1 position and was a feature shared between the pyrimidines, as shown in Figure 13 and 14.



ΔG° = 0.000 **ΔG° = 47.746**

Figure 13: Calculated lowest energy isomers for deprotonated thymine. Relative Gibbs energies (298 K) are shown in kJ/mol.



ΔG° = 0.000 **ΔG° = 51.717**

Figure 14: Calculated lowest energy isomers for deprotonated uracil. Relative Gibbs energies (298 K) are shown in kJ/mol.

The monomers of each neutral nucleobase were optimized at a B3LYP/6-311++G(d,p) level of theory. The optimized structures were then used as the candidate structures for the single point energy calculations, using MP2=full/cc-pVTZ, in order to derive a more accurate value. The values determined from these early calculations are extremely important to the accuracy of the data in this chapter. Moreover, the same principles were applied to calculate the thermochemical data of the sodium cation as it too is used in calculations further downstream. Once an isomer has been computationally validated to be the most stable isomer, and by extension the most plausible experimental isomer, a vibrational spectrum of that isomer is extracted and will be used to qualitatively assess against the experimental spectra produced by IRMPD and the FEL.

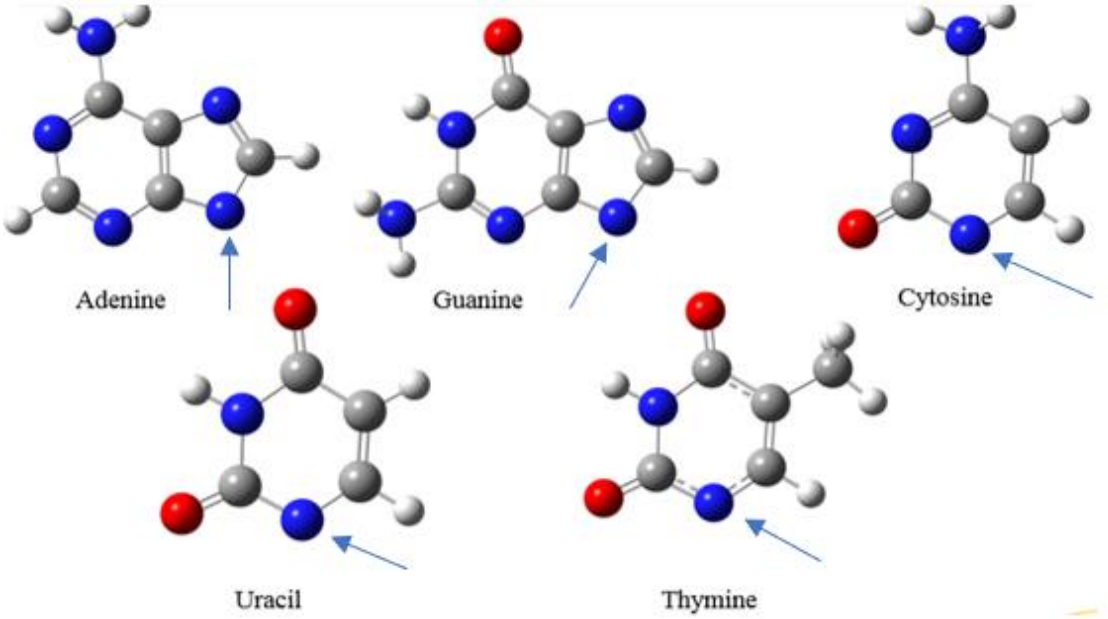


Figure 15: The computationally determined isomers of the most energetically favourable isomers for each nucleobase. The arrows denote the most probable site for proton loss.

The isomers shown in Figure 15, therefore, represent the structures of deprotonated nucleobases that are likely to be experimentally produced and will be used as the foundation of the complexation analysis with sodium cations. To ensure the validity of the computational analysis, an experimental vibrational spectrum that is nearly identical to that of a computational spectrum is required. Nearly, if not completely, identical experimental and computational spectra support the notion that the chemical system which produced the computational vibrational spectrum is in fact the same system seen experimentally.

If the computationally determined deprotonated nucleobase monomer appears to exist experimentally, then the energetics of those monomers can be confidently evaluated and compared to previously reported energetics. For this thesis, the gas-phase acidities of the nucleobases, and other thermochemical properties will be evaluated.

3.3 Results and Discussion

The most energetically favourable isomers for each of the deprotonated nucleobases was computationally determined by assessing the relative Gibbs energies of each of the isomers. The proton loss on the N9 atom in guanine and adenine, as well as the N1 atom in cytosine, thymine, and uracil, were found to be the most energetically favourable nucleobase isomers.

Table 2: The computationally determined thermochemical properties for the nucleobases that were validated both computationally and experimentally.

Reaction	ΔG° (kJ/mol)	ΔH° (kJ/mol)	ΔS° ($J \cdot K^{-1} \cdot mol^{-1}$)
[AH] ➔ $[dA]^-$ + $[H]^+$	1362.415	1381.361	20.155
[CH] ➔ $[dC]^-$ + $[H]^+$	1403.299	1427.621	24.464
[GH] ➔ $[dG]^-$ + $[H]^+$	1360.345	1385.487	25.121
[TH] ➔ $[dT]^-$ + $[H]^+$	1369.953	1395.070	25.102
[UH] ➔ $[dU]^-$ + $[H]^+$	1380.184	1404.998	24.859

As a result, the gas-phase acidity, representing the loss of a proton from the neutral nucleobase can be determined, as shown in Table 2. Before confirming the accuracy of the GPAs and the clustering potential of the deprotonated nucleobases, experimental spectra for each of the nucleobase monomers were produced and compared to their respective theoretical vibrational spectra. Once the most energetically favourable deprotonated isomer for each nucleobase was determined, the clustering behaviour of each deprotonated nucleobase was characterized. The clustering behaviour was evaluated via analyses of vibrational spectra, energetics of formation, geometries, charge, and bond distances. An important standard to compare the energetics of formation for each of the nucleobases is the Gibbs energy of deprotonation of water, in the gas phase. At - 228.61 kJ/mol, this provides a well-documented standard to assess the relative

complexing strength of the nucleobases. After these evaluations are performed the clustering behaviour of these deprotonated nucleobases with cationic sodium will be evaluated. Given that several of these clusters observe rotational symmetry, they will be referred to as Propeller geometries throughout the report.

3.3.1 Characterization of Deprotonated Pyrimidines

The investigation into the nucleobases was unfortunately limited to mostly computational analyses, however, several experimental spectra could be produced for the clusters of deprotonated thymine and deprotonated uracil. The computational and experimental evaluation of thymine and uracil may allow for a precedent to be set for the pyrimidines by illuminating trends within the pyrimidines, such as conserved vibrational modes and geometries. Moreover, the trends observed in the deprotonated clusters of thymine and uracil may help to reinforce the validity of the computational analysis for the clustering behaviour of deprotonated cytosine. Some conservation between the vibrational modes and energetically favourable structures within the pyrimidines is expected due to the similarities in structures between pyrimidines.

3.3.1.1 Deprotonated Thymine

The energy difference of 47.746 kJ/mol between the two most energetically favourable deprotonated isomers of thymine suggests that abstracting the proton from the N1 position of neutral thymine leads to a much more energetically favourable isomer of deprotonated thymine (dT). The structure and partial charges of the N1 isomer of dT are shown in Figure 16.

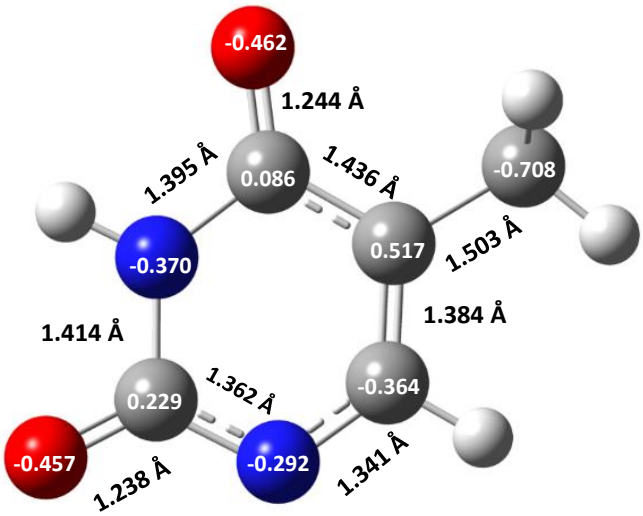


Figure 16: The computationally determined structure for the most energetically favourable isomer of deprotonated Thymine, the N1 isomer (dT-N1).

The charges and bond distances listed in Figure 16, will serve as a scaffold to evaluate the physicochemical changes in the monomers of dT as we introduce cationic sodium atoms, as well as more dT monomers. However, this computational assessment of the isomers of dT alone is not enough to confidently state there will be no other possible isomers produced experimentally. Therefore, an evaluation of the vibrational spectra produced by the two most favourable dT isomers against the experimentally produced vibrational spectrum will allow for a better understanding of the gas phase chemical system under investigation.

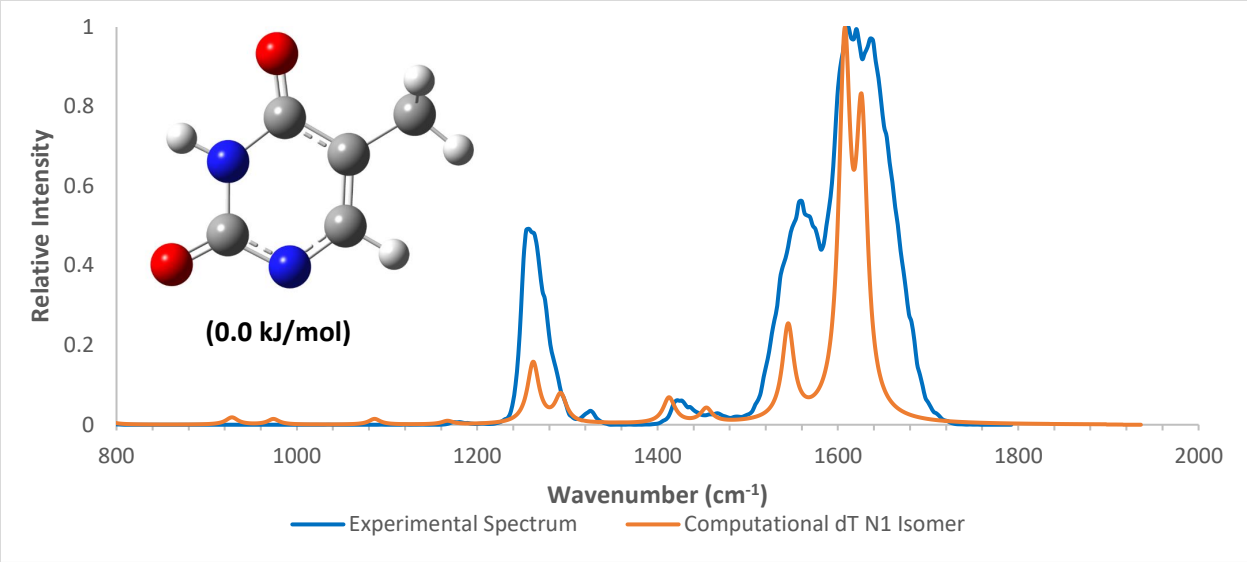


Figure 17: The computationally and experimentally determined spectra for the most energetically favourable isomer of deprotonated Thymine, the N1 isomer (dT-N1).

The dT-N1 isomer, shown in Figure 17, was computationally determined to be the most favourable isomer of dT. Superficially, there appears to be good agreement between the two spectra as 5 out of the 7 features observed in the experimental spectrum seem to be well-defined by the computational spectrum. In addition, the order of relative intensities between the experimental and computational spectra are in good agreement with the most intense peak occurring around 1600 cm^{-1} and the least intense peak seen around 1400 cm^{-1}. It is important to note that there appear to be two experimental peaks around 1320 cm^{-1} and 1640 cm^{-1} that were not well defined by any of the computationally produced peaks. As shown in Table 3, the exact peak positions for the computational and experimental peaks with the best overlap have been summarized.

Table 3: The computed vibrational frequencies for dT-N1 cluster and the experimental vibrational frequencies for dT.

Experimental dT Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1257	**1262**	Stretch of N1/C6 bond
1326		
1421	**1412**	Symmetric in-plane stretching of N1/C2 and C4/C5
1560	**1544**	Stretch of C5/C6 bond
1612	**1608**	Carbonyl stretch of C4/O
1621	**1626**	Carbonyl stretch of C2/O
1635		

The data in Table 3 demonstrates the similarities between the two spectra and describes the vibrational modes of the peaks with the best overlap. Apart from the experimental peak at 1560 cm^{-1}, the agreement between the overlapping peaks was within 1% error, supporting the possibility for the dT-N1 isomer to exist experimentally. However, as previously mentioned, the existence of unassigned experimental peaks can suggest the possibility of an isomer mixture.

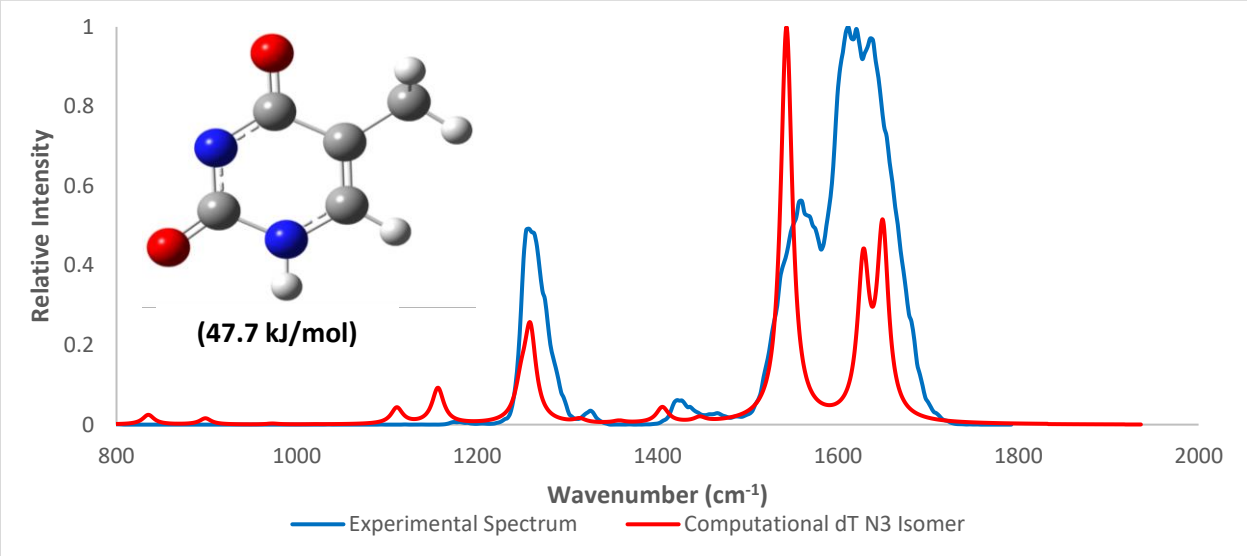


Figure 18: The computationally and experimentally determined spectra for the less energetically favourable isomer of deprotonated Thymine, the N1 isomer (dT-N3).

The experimental spectrum presented in Figure 18, is the same experimental spectrum shown in Figure 17, and can be used to evaluate the possibility of the dT-N3 isomer existing in mixture with the dT-N1 isomer. The spectrum for dT-N3 appears to have comparable agreement, with experimental peak positions, to those in the spectrum of dT-N1. The differences in vibrational mode types seen between the two isomers of dT are the result of different positions for the abstraction of protons at different sites as seen in Table 4.

Table 4: The computed vibrational frequencies for dT-N3 isomer and the experimental vibrational frequencies for dT.

Experimental dT Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1257	**1259**	Symmetric in-plane stretch of N1/C2 and C4/C5 bonds
1326		
1421	**1406**	Rocking of N1 proton
1560	**1544**	Carbonyl stretch of C4/O bond
1612		
1621	**1628**	Carbonyl stretch of C2/O bond
1635	**1645**	Stretch of C5/C6 bond

However, in the case of the dT-N3 isomer, the intensities of the peaks do not parallel the intensities exhibited by the experimental spectrum. Therefore, it is very likely that the dominant chemical system exhibited experimentally is best characterized by the dT-N1 isomer. This does not exclude the possibility of a mixture of isomers, but simply reinforces the conclusion made from the computational energetic analysis that the dT-N1 isomer is in fact the dominantly present isomer of dT.

Since the dominant isomer of dT has been both computationally and experimentally validated as the dT-N1 isomer, it was used to propagate the clustering studies for the clustering behaviour of dT. Alternating stepwise additions of cationic sodium and dT were then introduced to the chemical system to build clusters of increasing complexity and determine their thermodynamic properties, as seen in Table 6.

Table 5: The thermodynamic properties of reaction for the clusters of dT-N1 and Na^+.

Reaction	ΔG° (kJ/mol)	ΔH° (kJ/mol)	ΔS° ($J \cdot K^{-1} \cdot mol^{-1}$)
$[dT]^- + [Na]^+ \rightarrow [dT][Na]$	-540.645	-540.645	-27.475
$[dT][Na] + [dT]^- \rightarrow [dT]_2[Na]^-$	-180.221	-218.770	-58.377
$[Na][dT]_2^- + [Na]^+ \rightarrow [dT]_2[Na]_2$ Propeller Geometry	-490.470	-525.827	-28.343
$[Na][dT]_2^- + [Na]^+ \rightarrow [dT]_2[Na]_2$ Train Geometry	-416.773	-445.017	-81.018
$[dT]_2[Na]_2$ (Propeller) $+ [dT]^- \rightarrow [dT]_3[Na]_2^-$ Propeller A Geometry	-190.095	-234.988	-35.987
$[dT]_2[Na]_2$ (Propeller) $+ [dT]^- \rightarrow [dT]_3[Na]_2^-$ Propeller B Geometry	-171.112	-221.361	-40.281
$[dT]_2[Na]_2$ (Train) $+ [dT]^- \rightarrow [dT]_3[Na]_2^-$ Train Geometry	-223.578	-271.118	-38.109

Despite not being able to experimentally observe the neutral clusters of dT and sodium cations, it is important to include the geometries to observe the stepwise additions and their effects on the clustering behaviour. The reaction for the formation of the Propeller Geometry for the dT_2-Na_2 was shown to be more favourable than that of the Train geometry, for the same type of cluster, having used the same dT_2-Na foundation. Conversely, the Gibbs energy of reaction for the Propeller A and Propeller B geometries of dT_3-Na_2 appeared to be less favourable than that of the Train geometry, however, it is important to note they were created from different dT_2-Na_2 candidate structures. In every case, the absolute Gibbs energy (G°) of the Propeller structure at every level of clustering was lower than the Train structures. This is further discussed in the following sections.

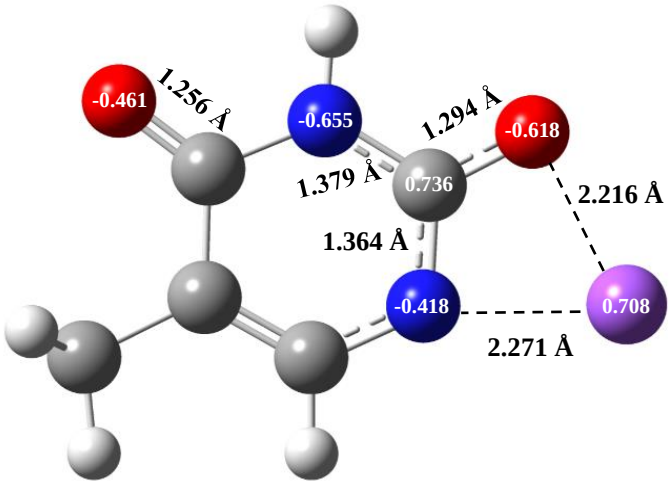


Figure 19: The computationally determined structure for the cluster of dT-N1 and a cationic sodium.

In Figure 19, the geometrically optimized cluster of dT-Na is shown. It was expected that the sodium cation would favourably react with the region of N1/C2/O_{C2} due to the regions large negative charge compared to other regions of the dT-N1 isomer, based on fundamental chemical principles. With the addition of the sodium cation, there appears to be an increase in the negative charges and bond distances of the atoms surrounding the cation. This optimized structure served as the basis for the next cluster through the addition of another dT-N1 isomer. Based on the favourable interaction of the dT-N1 isomer about the N1/C2/O_{C2} region, it is expected that the next dT will similarly interact.

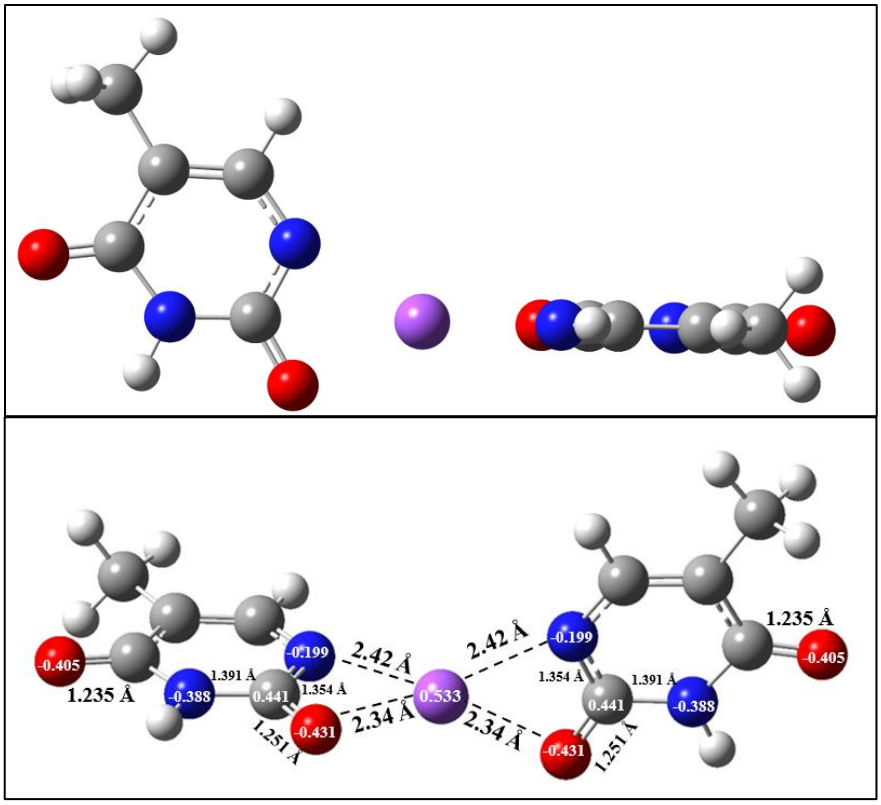


Figure 20: The optimized structure of dT_2-Na cluster.

This dT_2-Na cluster, shown in Figure 20, demonstrated symmetrical interactions between the respective oxygen and nitrogen atoms of one molecule, and the oxygen and nitrogen atoms of the second molecule, at the exact same site. The dihedral angle between the two rings using atoms $N1_A/O_{C2}/N1_B/O_{C2}$, is 85.99° and is nearly orthogonal about the sodium cation. Physiochemical properties such as bond distances and charges are completely equal on both sides of the sodium and reinforces the expectation that the two independent regions of the dT monomers would equally interact about the sodium cation. For this particular cluster, we were fortunate enough to have collected an experimental spectrum, thus, we evaluated the validity of this structure through the comparison of the computational and experimental vibrational spectra.

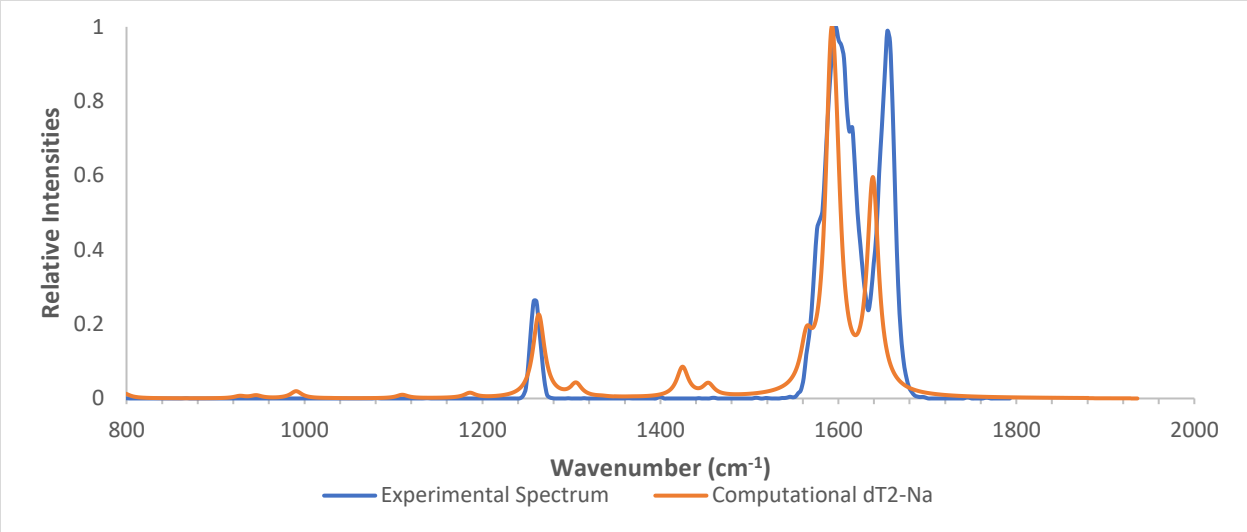


Figure 21: The computationally and experimentally determined spectra for the structure of two deprotonated thymine monomers with a sodium cation (dT_2Na^-).

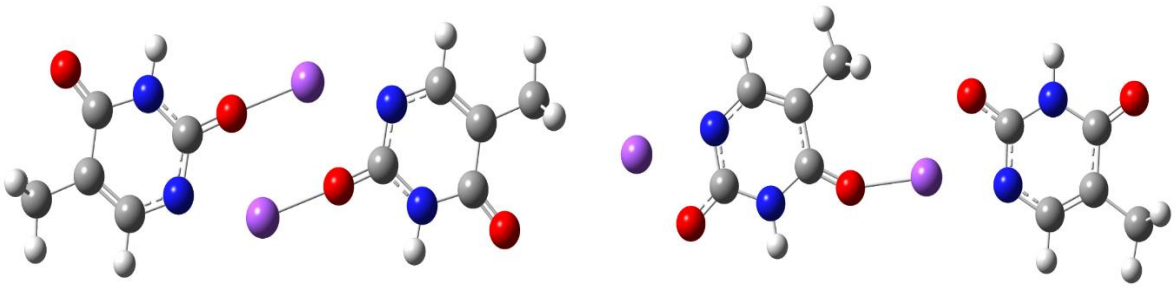
In Figure 21, the experimentally determined vibrational spectrum of dT_2-Na is compared to the computational spectrum of the same structure, seen in Figure 20. The experimental spectrum contains three major peaks and one barely visible peak at 1401 cm^{-1}. Each of the major peaks was shown to have strong agreement with peaks found in the computational spectrum, except for the experimental peak at 1400 cm^{-1}. Table 5, summarizes the major and minor experimental peaks, the computational peaks with the greatest agreement to those experimental peaks and their respective vibrational modes.

Table 6: The experimental and computed vibrational frequencies for dT_2-Na cluster.

Experimental dT_2-Na Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1258	**1263**	Asymmetric rocking of N3 and C6 protons
1401	**1424**	Symmetric in-plane stretch of N1/C2 and C4/C5 bonds
1597	**1592**	Carbonyl stretch at C2/O with additional N3 proton rocking
1655	**1640**	Carbonyl stretch at C4/O bond

The agreement between the three largest peaks in the computational and experimental spectra strongly implies that the geometry presented in Figure 20 effectively reflects the geometric structure most likely dT2-Na system to be experimentally observed. However, this strong agreement is not perfect. In the experimental spectrum, there appears to be a second peak hidden by the major experimental peak at 1597 cm^{-1}, represented by a slight bump along the major peak. Although the computational spectrum does not show a well-defined peak for this particular peak, at 1615 cm^{-1}, it is possible that it was a vibrational mode that was favourably induced for a particular attenuation. The Gibbs energy of reaction for the formation of dT_2-Na from dT-Na and dT, of -182.408 kJ/mol, further supports the viability of the geometry presented in Figure 20. Based on fundamental chemical principles, it is unlikely that there will be another dT_2-Na cluster geometry which favourably interacts about the sodium cation, with increased symmetry. Therefore, the evidence supports favourable clustering of two dT's with a cationic sodium in the gas phase, in a near-orthogonal manner.

The next stepwise addition to evaluate the clustering behaviour of dT was the addition of a second sodium cation. To ensure that the geometric optimizations did not accidentally converge to a local minimum that is not the global minimum energy structure, the two most likely structures for dT_2-Na_2 were evaluated. As seen in Figure 22, there were two computationally determined minimum energy structures for dT_2-Na_2.



ΔG° = 0.000 (Propeller) **ΔG° = 73.697 (Train)**

Figure 22: The computationally determined geometric structures for two arrangements of two deprotonated thymine monomers with two sodium cations (dT_2Na_2). The relative Gibbs energy is measured in kJ/mol.

The significant difference in relative Gibbs energy, seen in Figure 22, suggests that maintaining the symmetry, previously incurred through the addition of the second dT, will benefit the overall stability of the structure and provide a more energetically favourable chemical system. The system that maintains symmetry about two sodium cations lined up between the two monomers of dT was referred to as the 'propeller system', while the other structure was referred to as a 'train'. The neutral clusters of dT_2-Na_2 could not be validated experimentally. As a result, to ensure the conclusions based off of the computational evaluation of the dT_2-Na_2 system, the clusters derived from both variations were evaluated. A third dT monomer was added to each dT_2-Na_2 system, geometrically optimized, and then had their vibrational spectra evaluated against the computational spectrum produced for the dT_3-Na_2 system.

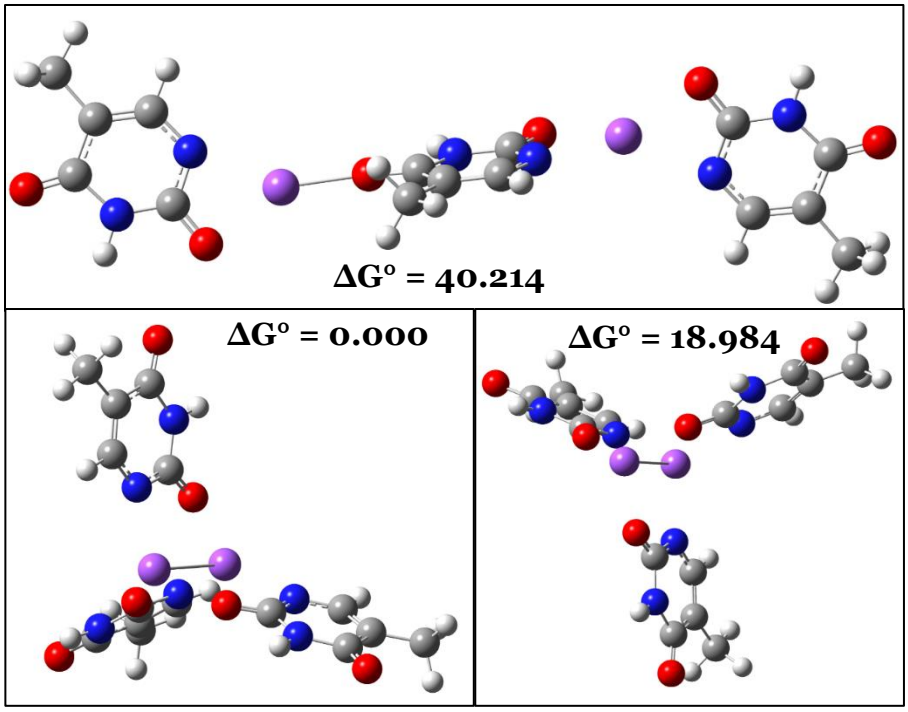


Figure 23: The computationally determined geometric structures for three arrangements of three deprotonated thymine monomers with two sodium cations (dT_3Na_2). Top image is the "train" geometry, bottom-left is Propeller A and bottom-right is Propeller B. The relative Gibbs energy is measured in kJ/mol.

In Figure 23, there are three possible geometric structures that can be achieved through the addition of a dT-N1 monomer to either of the dT_2-Na_2 systems, seen in Figure 22. The dT_3-NA_2 system with the greatest relative Gibbs energy corresponds to the subsequent addition of a dT to the previously mentioned "train" system of dT_2-Na_2. This linear "train" geometry, for the dT_3-Na_2 cluster, utilized each oxygen of one dT to form an ionic bond with the cationic sodium atoms and bond with the N1/C2-oxygen region of the two other dT's. The lack of symmetry in the "train" geometry, exhibited by the two "propeller" clustering geometries, was likely the cause for its increase in Gibbs energy, thus, making the "train" geometry roughly 50 kJ/mol less energetically favourable than either of the two "propeller" systems. This large difference in relative Gibbs energy suggests, that the train geometry of the dT_3-Na_2 system is far less likely to exist experimentally.

To achieve the two different propeller systems, the third dT monomer was added about the Na^+-Na^+ internuclear axis for the energetically favourable dT_2-Na_2 propeller system (Propeller A), and having alternating orientations of the dT monomers, or by rotating the each of the dT monomers to allow for the C2-oxygens of each monomer to face the same direction (Propeller B). It was found that Propeller A was more energetically favourable than Propeller B, computationally suggesting that Propeller A was the most likely to be the dominant form of dT_3-Na_2 seen experimentally. However, to fully address this, we examined the vibrational spectra for dT_3-Na_2 and evaluated the validity of this finding.

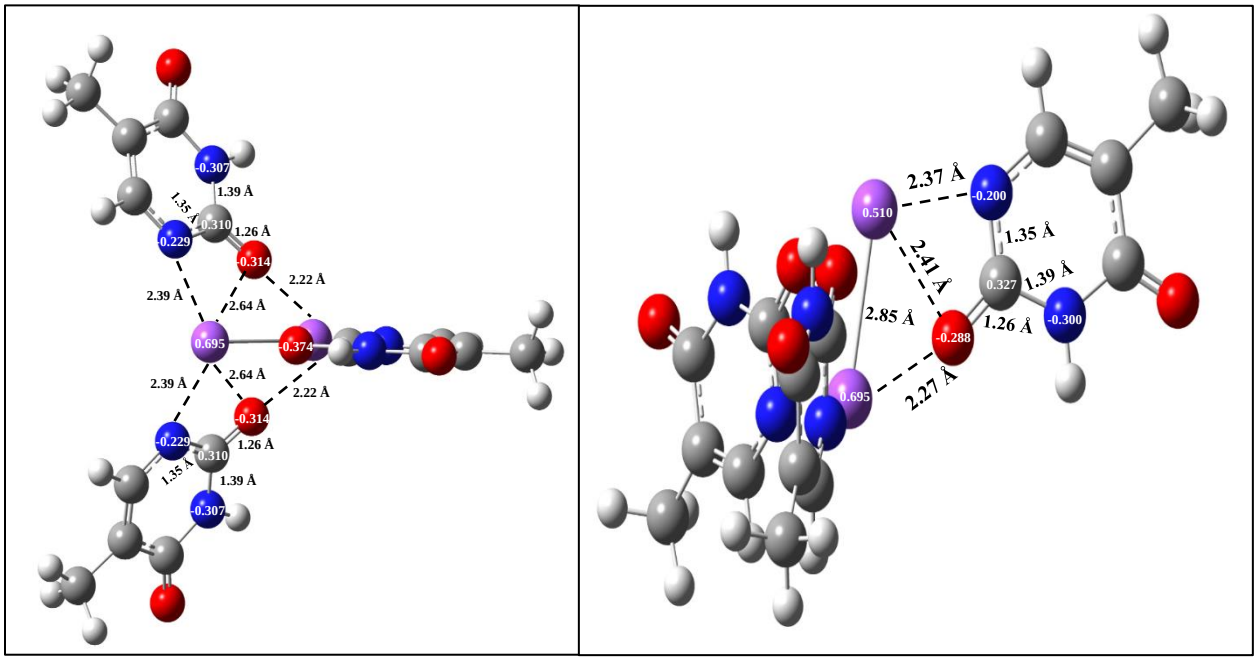


Figure 24: The computationally determined geometry for dT_3-Na_3, Propeller A.

The geometry of Propeller A, displayed in Figure 24, demonstrates the high degree of symmetry (~$C1_v$) between the two monomers of dT that are oriented along the Na^+-Na^+ axis in the same way. In addition, it extremely interesting to note the conservation of bond distances between atoms within the rings of the monomers, despite their orientation. Moreover, the non-symmetric monomer of dT was found to have greater

interaction with one sodium cation, while the symmetric monomers more strongly interacted with the other available sodium cation.

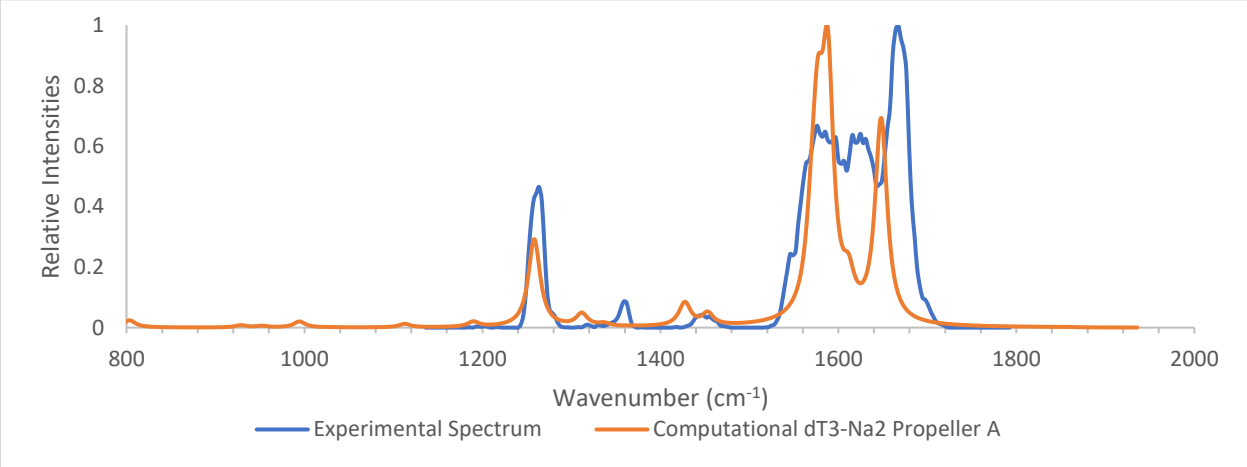


Figure 25: The computationally and experimentally determined spectra for the structure of dT_3-Na_2, the Propeller A geometry.

The spectra demonstrated in Figure 25, for the Propeller A geometry of dT_3-Na_2, shows some agreement on four of the major and minor experimental peaks. As shown in Table 7 below, there are however, two experimental peaks which are not well characterized by the computational spectrum, at 1359 cm^{-1} and 1625 cm^{-1}. The minor experimental peak at 1359 cm^{-1} did not have any computational peaks in its immediate proximity, which may suggest there is a mixture of geometries for dT_3-Na_2 that led to the production of that particular vibrational mode. This possibility can be verified by examining the computed spectra for other geometries of dT_3-Na_2 and seeing whether the experimental peak at 1359 cm^{-1} is better characterized by another geometry. The experimental peak at 1625 cm^{-1} might similarly be better characterized by a vibrational mode existing in one of the other two geometries for dT_3-Na_2.

Table 7: The experimental and computed vibrational frequencies for dT_3-Na_2 Propeller A.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1264	**1259**	Asymmetric rocking of N3 and C6 protons
1359		
1446	**1453**	Scissoring of CH_3 protons
1579	**1577**	Carbonyl stretch at C2/O on symmetric dT's
	1585	Carbonyl stretching at C2/O on asymmetric dT
1625		
1667	**1649**	Carbonyl stretch at C4/O bond

Given the approximate agreement between the experimental spectrum and that of Propeller A, and the computational analysis, it gives sufficient support that Propeller A characterizes the chemical system for dT_3-Na_2.

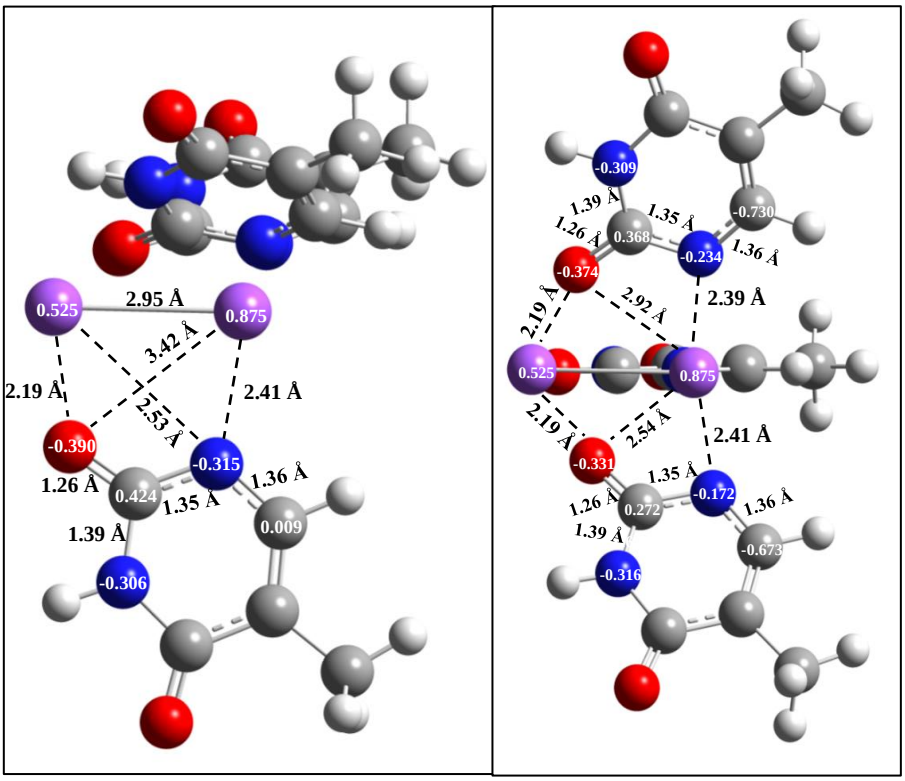


Figure 26: The computationally determined geometry for the structure of dT3-Na_2 in the Propeller B geometry.

The Propeller B geometry summarized in Figure 26 highlights the deviation from geometrical symmetry exhibited by the Propeller A geometry. Each monomer of dT in the Propeller B geometry is oriented in the same direction, with respect to the Na^+-Na^+ internuclear axis. The monomers appeared to shift along the Na^+-Na^+ axis to better accommodate their respective interactions with the sodium cations. Despite having seemingly consistent bond distances within each monomer, the bond distances between the sodium cations, the C2-oxygen and the N3 atoms varied greatly. Although this structure may not be the most energetically favourable, the deviation from symmetry may allow for vibrational modes to be excited, that were not observed in the Propeller A geometry.

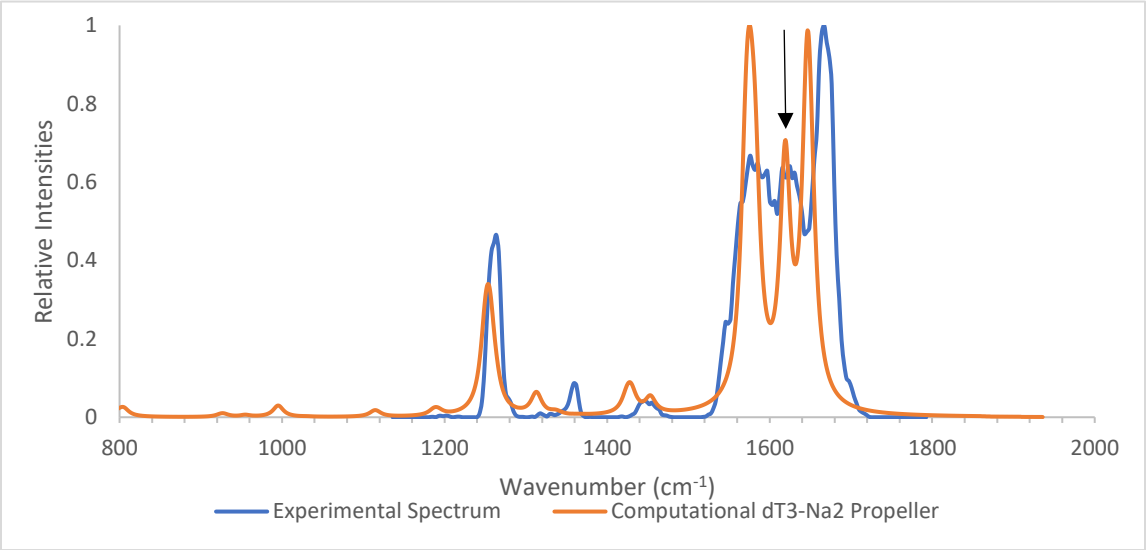


Figure 27: The computationally and experimentally determined spectra for dT_3-Na_2, with Propeller B geometry.

It was expected that there would be conservation of vibrational modes between Propeller A and B, as they are nearly the same system, but it was nonetheless interesting to have seen how much of a similarity the computational spectrum for Propeller B had with Propeller A. However, there was one glaring exception; the peak at 1620 cm^{-1}.

Table 8: The experimental and computed vibrational frequencies for dT_3-Na_2 Propeller B cluster.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1264	**1254**	Asymmetric in-plane stretch of N1/C6 and C2/N3 bonds
1359		
1446	**1453**	Scissoring of CH_3 protons
1579	**1574**	Stretch of N1/C6 bonds
1625	**1620**	Carbonyl stretch at C2/O
1667	**1647**	Carbonyl Stretch at C4/O

The geometry of Propeller B appeared to produce vibrational frequencies, listed in Table 8, at nearly identical positions as Propeller A but demonstrates a difference in the vibrational modes of the experimental peaks of 1579 cm $^{-1}$ and 1625 cm^{-1}. It appeared as though the 180° rotation of the asymmetric dT monomer found in Propeller A led to the production of a vibrational mode unique to the geometry of Propeller B. This would reinforce the possibility that the experimental spectrum obtained is produced characteristic of a mixture of Propeller A and Propeller B geometries.

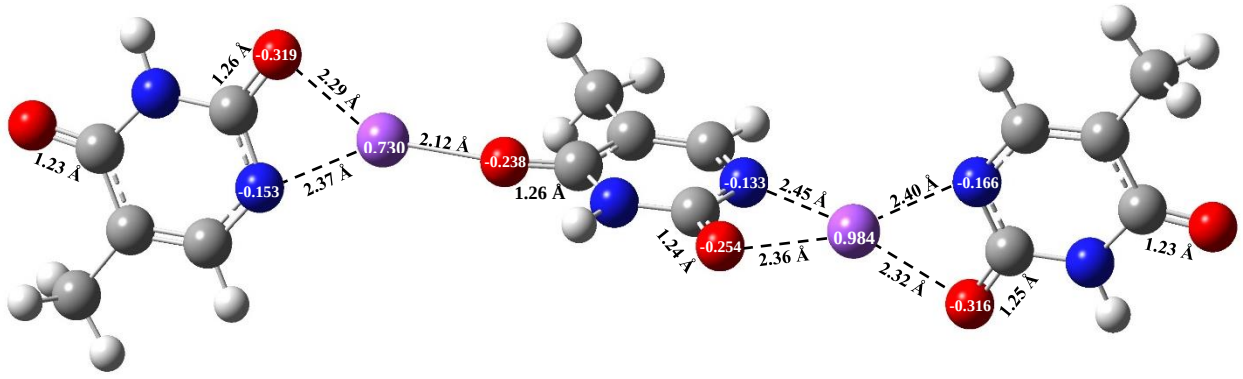


Figure 28: The optimized structure of dT_3Na_2 in their train orientation.

The train orientation shown in Figure 28, demonstrates deviation from the symmetrical interactions observed in Propeller A geometry of dT_3Na_2. The dihedral angles between the two left rings are 67.99°, and 82.83° for the rings on the right; as shown. The large differences in charges and bond distances of the interacting atoms can be attributed to the absence of symmetry as well as the introduction of a dT that is forced to interact with the sodium cations at either oxygen. The difference in geometry would likely have resulted in changes to the computed vibrational spectrum, that would easily distinguish this less energetically favourable geometry of dT_3-Na_2.

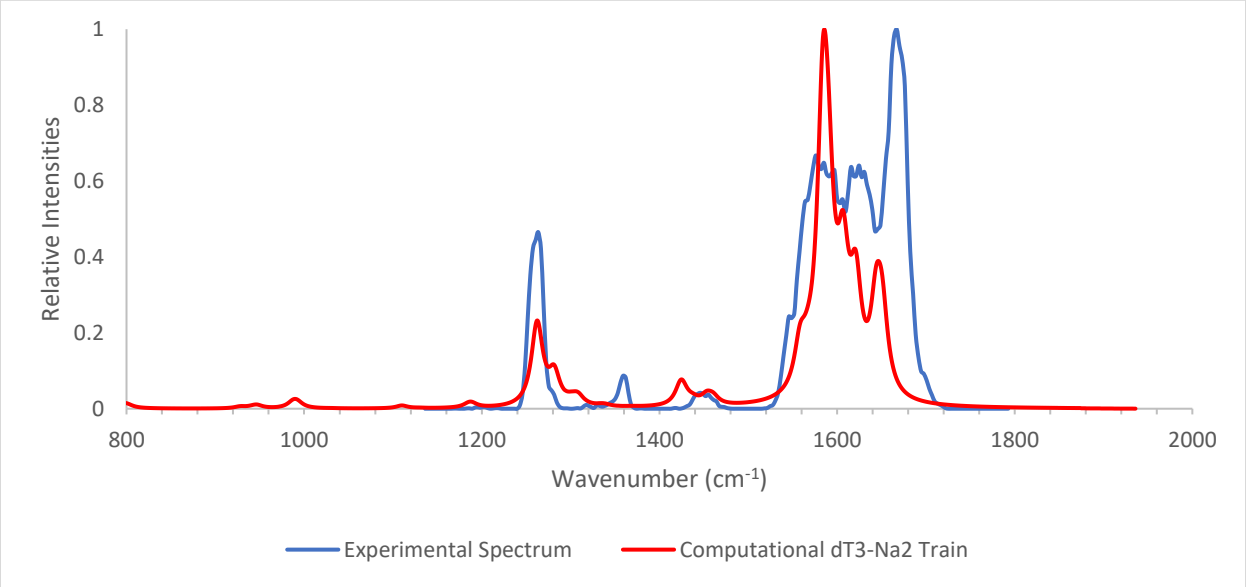


Figure 29: The experimental and computational spectra of dT_3-Na_2. The computed spectrum is for the train geometry.

It is clear that based on the vibrational spectra of the least energetically favourable geometry of dT_3-Na_2, as seen in Figure 29, there is a large overlap between the three computed vibrational spectra of dT_3-Na_2. Therefore, inspection of the vibrational spectra alone would not be enough to confidently distinguish, at this level of clustering complexity, the geometry that best characterizes the chemical system of dT_3-Na_2.

Table 9: The experimental and computed vibrational frequencies for dT_3-Na_2 train cluster.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1264	**1262**	Asymmetric in-plane rocking of N3 and C6 protons on outer dT's
1359		
1446	**1455**	Scissoring of inner monomer CH_3 protons
1576	**1584**	Carbonyl stretch at C2/O
1625		
1667	**1649**	Carbonyl stretch at C4/O

Table 9 summarizes the peak positions of the train geometry for dT_3-Na_2 and clearly demonstrates the differences in vibrational modes for the peaks at 1262 cm^{-1} and 1584 cm^{-1}. The largest intensity peak in the computational spectrum is not paralleled by the computational spectrum. Moreover, the train geometry does not appear to have a well-defined peak for the experimental peak at 1625 cm^{-1}.

The findings in this subchapter strongly suggest that there is a mixture of the propeller geometries found experimentally. Propeller A is computed to be the most energetically favourable geometry for dT_3-Na_2 but does not have the best agreement with the experimentally produced vibrational spectrum. The Propeller B geometry demonstrates the best agreement with the experimental vibrational spectrum and has a sufficiently small relative Gibbs energy, to the Propeller A, that suggests the isomers could have been produced in sufficient quantities to observe experimentally. The train geometry was both the most energetically unfavourable geometry and had the least amount of agreement with the experimental vibrational spectrum and therefore, is not likely to exist.

3.3.1.2 Deprotonated Uracil

Uracil is a nucleobase that is found exclusively in RNA and is replaced by thymine in DNA. The reason for this is that uracil and thymine are two types of pyrimidines, along with cytosine, and they are similar in structure and reaction mechanisms. Through an investigation of the RNA-exclusive nucleobase of uracil, similar to the previous investigation of the DNA-exclusive thymine, the hope is to elucidate differences in clustering behaviour that may be exclusive to the non-polar biological environment they can be found in.

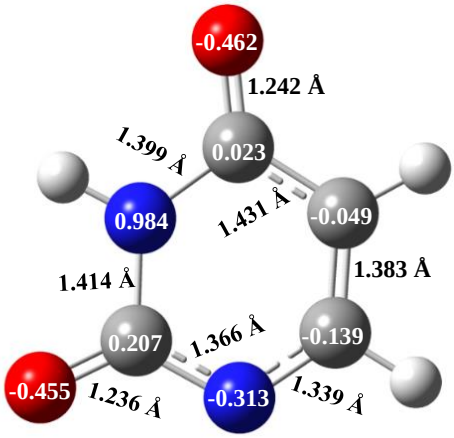


Figure 30: The computationally determined most energetically favourable isomer of dU, the dU-N1 isomer.

In Figure 30, the physiochemical properties of the computationally determined energetically favourable isomer of deprotonated uracil (dU) are shown. Once again, this is a means to observe the type of effects that dU monomers and sodium cations would have on the clustering behaviour and ring structure of each monomer. It was determined that the dU-N1 isomer was 51.717 kJ/mol more energetically favourable to the dU-N3 isomer which abstracted a proton from the N3 position. The computational analysis alone

would not be sufficient to correctly identify the dominant isomer, thus, experimental analysis on the vibrational spectra of these isomers was performed.

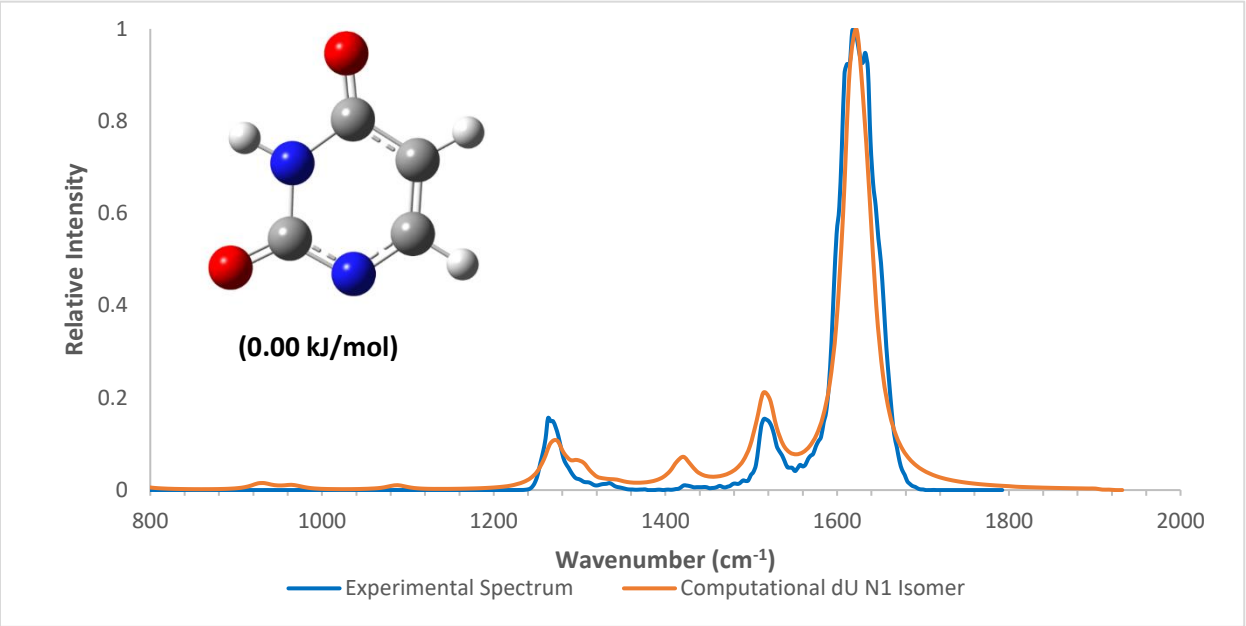


Figure 31: A comparison of the experimentally determined vibrational spectrum for dU and the computational spectrum for the most energetically favourable isomer of dU, the N1 isomer.

The dU-N1 isomer was determined to be the most energetically favourable isomer of dU, and when analyzed, produced vibrational frequencies (in Figure 31) that strongly agreed with the vibrational frequencies produced experimentally. The three largest experimental peaks for the vibrational frequencies of 1264, 1515 and 1621 cm^{-1} appeared to be extremely well characterized by three of the computed vibrational frequencies for the dU-N1 isomer. In addition, the minor experimental peak at 1421 cm^{-1} was also well characterized by the computational peak of the dU-N1 isomer. There is however, one experimental peak, at 1335 cm^{-1}, which was not defined by the computational spectrum.

Table 10: The experimental vibrational frequencies for dU and computed vibrational frequencies for the N1 isomer of dU.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1264	**1266**	Scissoring of C5/C6 protons
1335		
1421	**1422**	Symmetric stretching of N1/C2 and C4/C5
1515	**1514**	Asymmetric stretching of C5/C6/N1 bonds
1621	**1622**	Asymmetric stretching of C2/O and C4/O with slight N3 proton rocking

It is possible that the cause for this unassigned peak, seen in Table 10, was due to the vibrational mode relating to that peak belonging exclusively to another experimentally produced isomer of dU existing alongside the dU-N1 isomer. To further investigate the possibility of a mixture, the second most energetically favourable isomer of dU, the N3 isomer, was experimentally assessed.

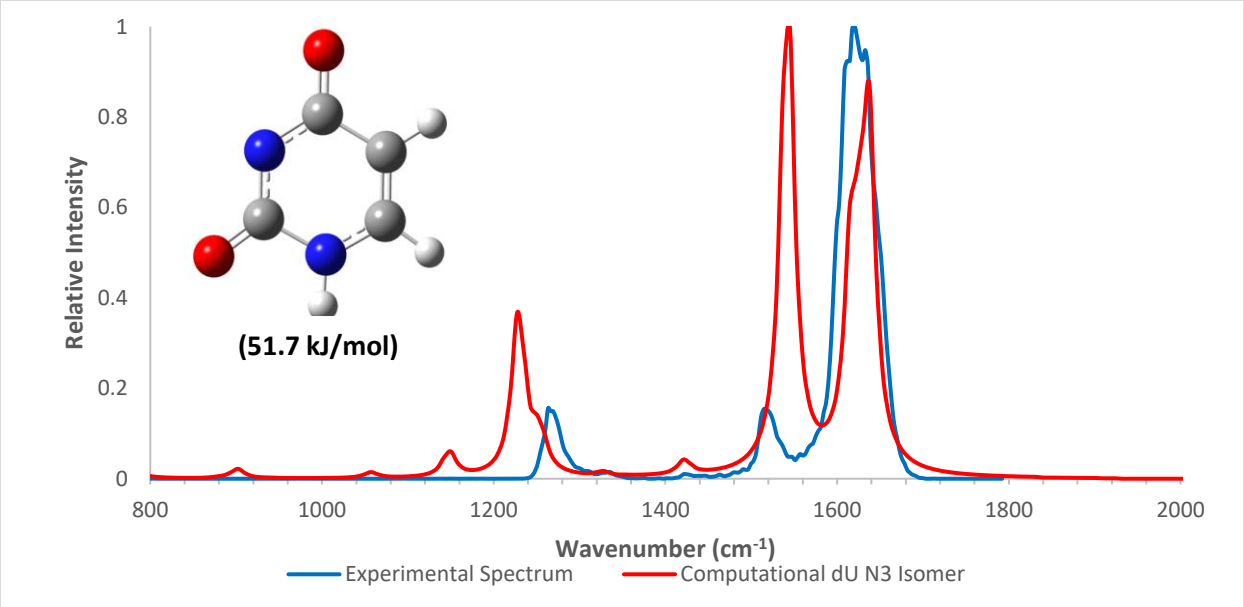


Figure 32: A comparison of the experimentally determined vibrational spectrum for dU and the computational spectrum for the N3 isomer of dU.

It was apparent that the agreement between the spectrum of the dU-N3 isomer, shown in Figure 32, and the experimental spectrum was much poorer than the agreement with the dU-N1 isomer. One of each of the major and minor experimental peaks were in good agreement with the computational spectrum and suggests there is likely not a great amount of dU-N3 produced in tandem with dU-N1.

Table 11: The experimental vibrational frequencies for dU and computed vibrational frequencies for the N3 isomer of dU.

Experimental dU Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1264		
1335		
1421	**1422**	Asymmetric stretching of C4/C5 with C6/N1
1515		
1621	**1630**	Stretching of C2/O bond

The computationally determined vibrational frequencies, listed in Table 11, support the likeliness of the dU-N1 isomer being the dominant isomer. The two peaks conserved across the N1 and N3 isomers of dU might be attributed to vibrational modes that are not greatly affected by the change in the deprotonation site on the neutral dU monomer. In addition, the experimental peak of 1335 cm^{-1} does not appear to be better reproduced by the computed spectrum of the dU-N3 isomer than the dU-N1 isomer. This observation excludes the possibility that a mixture of dU-N1 and dU-N3 isomers was observed experimentally and might be the result of a vibrational mode exclusive to a far less energetically favourable isomer of dU. Given this information, we can exclude other possible isomers of dU from being the dominant isomer based on their significantly reduced energetic favourability as well as the lack of agreement with the experimental vibrational spectrum. Thus, reaffirming the likeliness of dU-N1 being the most abundant isomer of dU.

Having identified and concluded that the dU-N1 isomer is the most abundant and energetically favourable isomer of dU, the dU-N1 isomer was used to build the clusters, shown in Table 12, with sodium cations.

Table 12: The gas phase energetics of clustering reactions for dU and cationic sodium.

Reaction	ΔG° (kJ/mol)	ΔH° (kJ/mol)	ΔS° (J·K^{-1}·mol^{-1})
$[dU]^- + [Na]^+ \rightarrow [dU][Na]$	-530.179	-563.879	-27.015
$[dU][Na] + [dU]^- \rightarrow [dU]_2[Na]^-$	-189.910	-227.729	-30.317
$[Na][dU]_2^- + [Na]^+ \rightarrow [dU]_2[Na]_2$ Propeller Geometry	-482.358	-517.730	-28.355
$[Na][dU]_2^- + [Na]^+ \rightarrow [dU]_2[Na]_2$ Train Geometry	-368.194	-393.295	-20.122
$[dU]_2[Na]_2$ (Propeller) $+ [dU]^- \rightarrow [dU]_3[Na]_2^-$ Propeller A Geometry	-194.199	-238.630	-35.617
$[dU]_2[Na]_2$ (Propeller) $+ [dU]^- \rightarrow [dU]_3[Na]_2^-$ Propeller B Geometry	-182.245	-225.387	-34.584
$[dU]_2[Na]_2$ (Train) $+ [dU]^- \rightarrow [dU]_3[Na]_2^-$ Train Geometry	-243.550	-299.416	-44.784

It is clear by the Gibbs energy of reaction, shown in Table 12, that the formation of each type of cluster was energetically favourable. It was energetically interesting to see how the clustering of each type of geometry, at every level of complexity, followed the same pattern as the clustering of dT and sodium cations, as seen in Table 5. The Gibbs energy of reaction for forming Propeller A dU_3-Na_2 was slightly more favourable than Propeller B. Similar to the clusters of dT, the Gibbs energy of reaction for both dU_3-Na_2 Propeller geometries were still less favourable than the Train geometry but can be explained by the different starting structures used to build them. Lastly, the absolute energy of each Propeller geometry was more energetically favourable than their respective Train geometry counterparts, as observed in the clusters of dT. This suggested a clustering trend might exist across pyrimidines.

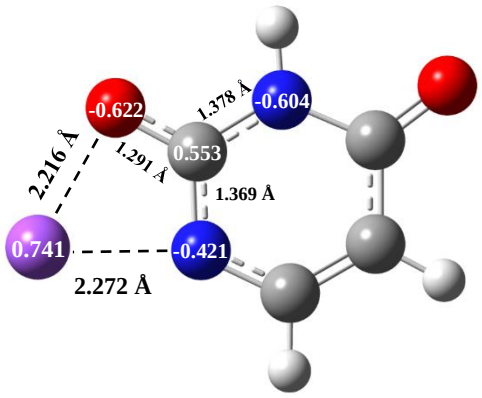


Figure 33: The computationally determined geometry properties of the dU-Na cluster.

The structure shown in Figure 33 demonstrates the most favourable site of interaction between the dU monomer and the sodium cation. Based on this geometric optimization, it was expected that the subsequent addition of a dU monomer would geometrically optimize to a symmetrical system, favouring the N1/C2-oxygen region of dU. To evaluate this assumption, the geometry optimized structure of dT_2-Na was computationally determined and shown in Figure 34.

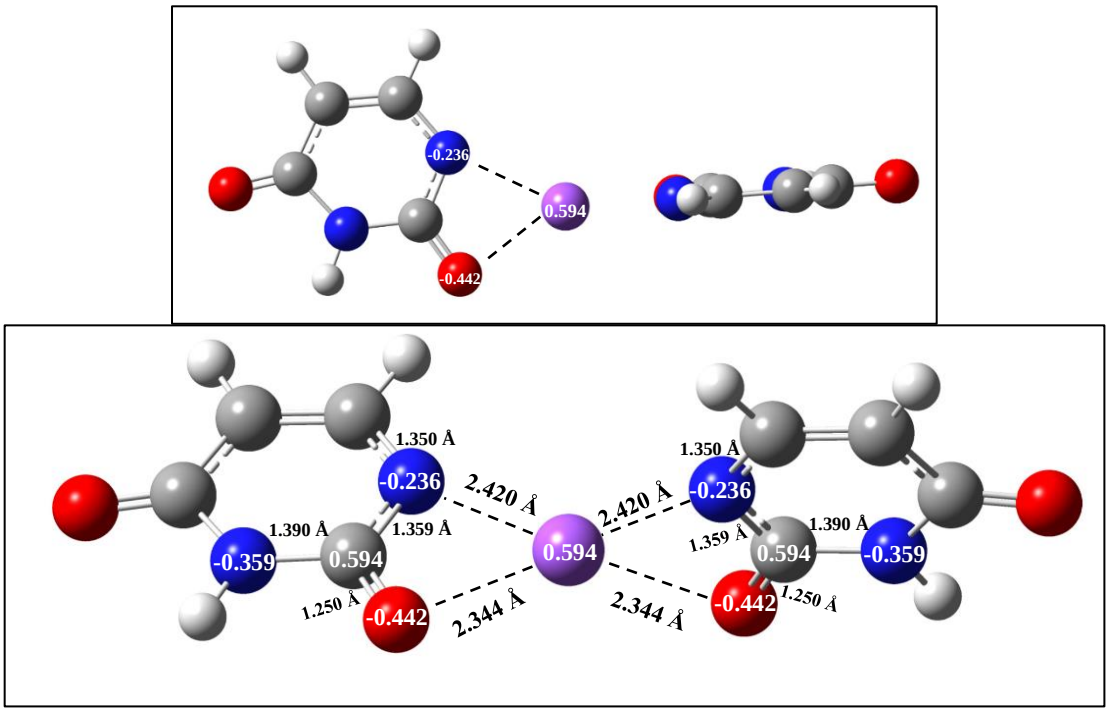


Figure 34: Multiple views of the computationally determined geometry of dU_2-Na.

As expected, the cluster of dU_2-Na shown in Figure 34 demonstrates a high level of symmetrical interactions about the central sodium cation. Both the charges and bond distances of the atoms directly or indirectly interacting with the central sodium cation, appear to be equal in both the dU monomers. A slight increase in the bond distances of the N1 and C2-oxygen was observed and is likely attributable to the increased distribution of positive charge to surrounding atoms in the dT_2-Na cluster compared to the dT-Na cluster. In addition, there is a dihedral angle, of 83.557°, created by the planes of opposing N1 and C2-oxygen atoms. Similar to the dihedral angle of 84.931° in the dT_2-Na system, the dU_2-Na cluster appears to favour approaching orthogonality between deprotonated monomers. The similarities between geometric orientation, dihedral angles and sites of ion-molecule interaction suggest that the mechanisms of clustering previously observed with clusters of dT, will be conserved relatively well across other pyrimidines. To further validate the computationally determined structure for dU_2-Na, an analysis of the vibrational spectra was performed.

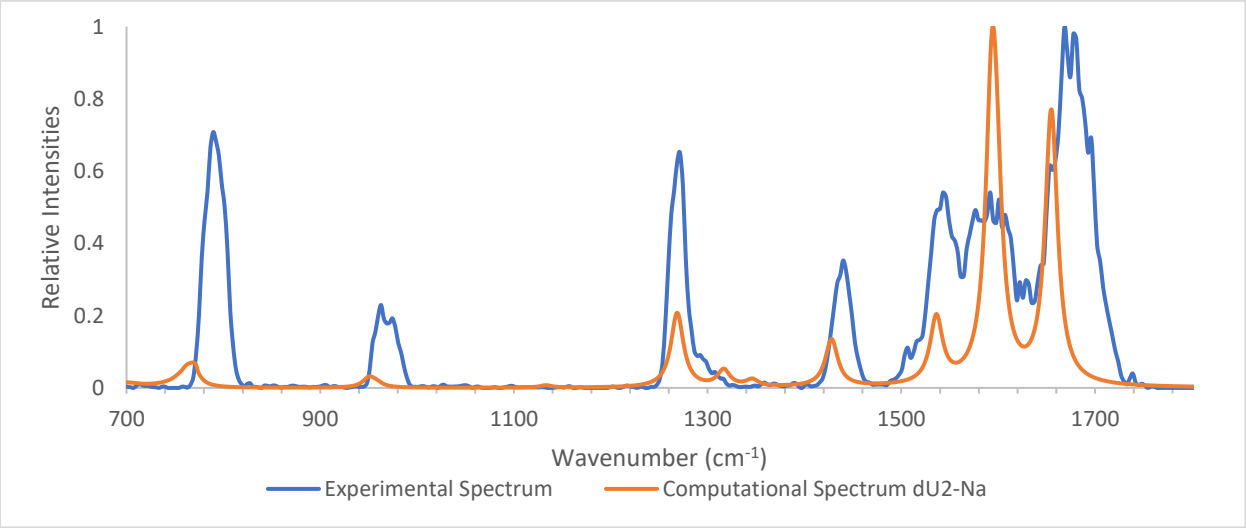


Figure 35: A comparison of the experimentally and computationally determined vibrational spectra for the dU_2-Na cluster.

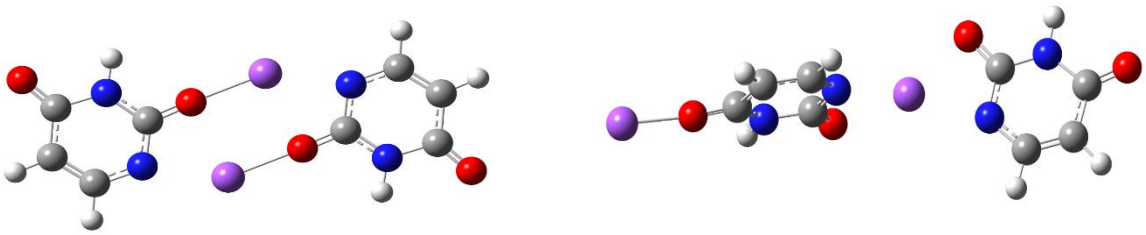
As seen in Figure 35, nearly all the major experimental peaks have been characterized well by the computational spectrum. Interestingly, the order of increasing relative intensities for the computational spectrum did not parallel the experimental spectrum. However, based on fundamental chemical principles and observations made in the dU-Na cluster, it is likely there are no other more energetically favourable dU_2-Na structures.

Table 13: The experimental and computational frequencies for dU_2-Na.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
790	**775**	Asymmetric wagging of C2/C4/C6 and N1/N3/C5 atoms
963	**952**	Scissoring of N1 and N3 atoms about C2
1271	**1269**	Rocking of C6 proton
1440	**1427**	Symmetric stretching of N1/C2 and C4/C5 atoms
1543	**1536**	Asymmetric stretching of N1/C6/C5 bonds
1592	**1594**	Stretching of C2/O bonds
1670	**1657**	Stretching of C4/O bonds

The vibrational frequencies observed experimentally and computationally in dU_2-Na, shown in Table 13, were compared to the vibrational frequencies of the dU-N1 isomer, Table 10, and found to have four vibrational modes that appeared to be relatively conserved. These four slightly red-shifted vibrational modes appear to represent vibrational modes intrinsic to dU, such as the stretching of the C2/O and C4/O bonds. The vibrational frequencies that were observed in the dU_2-Na cluster, and not in the dU-N1 isomer, are the result of interactions between the sodium cation and the dU's. Thus, it is possible that the non-conserved vibrational modes between dU_2-Na and dU might instead be conserved between dU_2-Na and dU_3-Na_2.

To better evaluate the clustering behaviour of dU_3-Na_2, an evaluation of the possible geometrical variations in dU_2-Na_2 was performed. By comparing relative Gibbs energies between the Propeller and Train motifs observed in dU clusters, similar to those of dT, the Propeller geometry of dU_2-Na_2 was computationally determined to be more energetically favourable than the Train geometry.



ΔG° = 0.000 (Propeller) **ΔG° = 114.164 (Train)**

Figure 36: The computationally determined geometry and relative Gibbs energies for the Propeller and Train clusters of dU_2-Na_2.

Figure 36 highlights the large difference in absolute Gibbs energy between the two most energetically favourable motifs of dU_2-Na_2. The two geometries for dU_2-Na_2 parallel the two most energetically favourable geometries of dT and reinforce the validity of the Propeller geometry serving as the more appropriate candidate structure to use to build up to dU_3-Na_2.

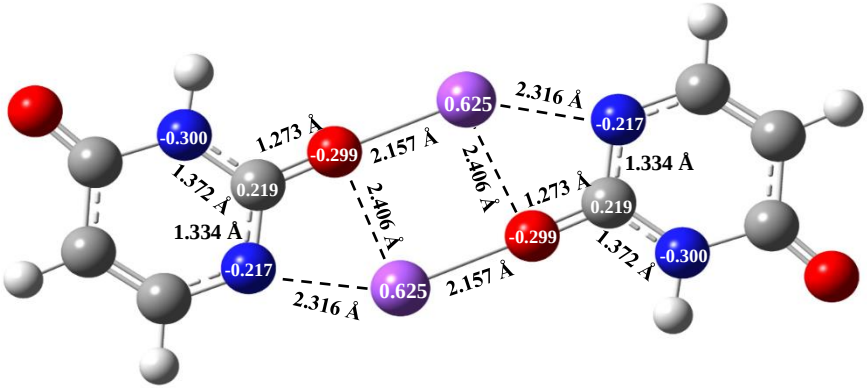


Figure 37: The computationally determined geometry of the planar dU_2-Na_2 Propeller cluster.

With the addition of a second sodium cation, shown in Figure 37, planarity of the chemical system is restored, and symmetry is maintained. In addition, the geometric optimization appears to have determined that ionic bonds between the sodium cations and the C2-oxygens of each dU are formed.

The observations previously made on the dT_3-Na_2 clusters, suggested that there are three energetically favourable variations of the dU_3-Na_2 clusters; Propeller A, Propeller B and the Train geometry. By comparing the relative Gibbs energy between these three motifs, the Propeller A geometry was once again determined to be the most energetically favourable geometry for dU_3-Na_2.

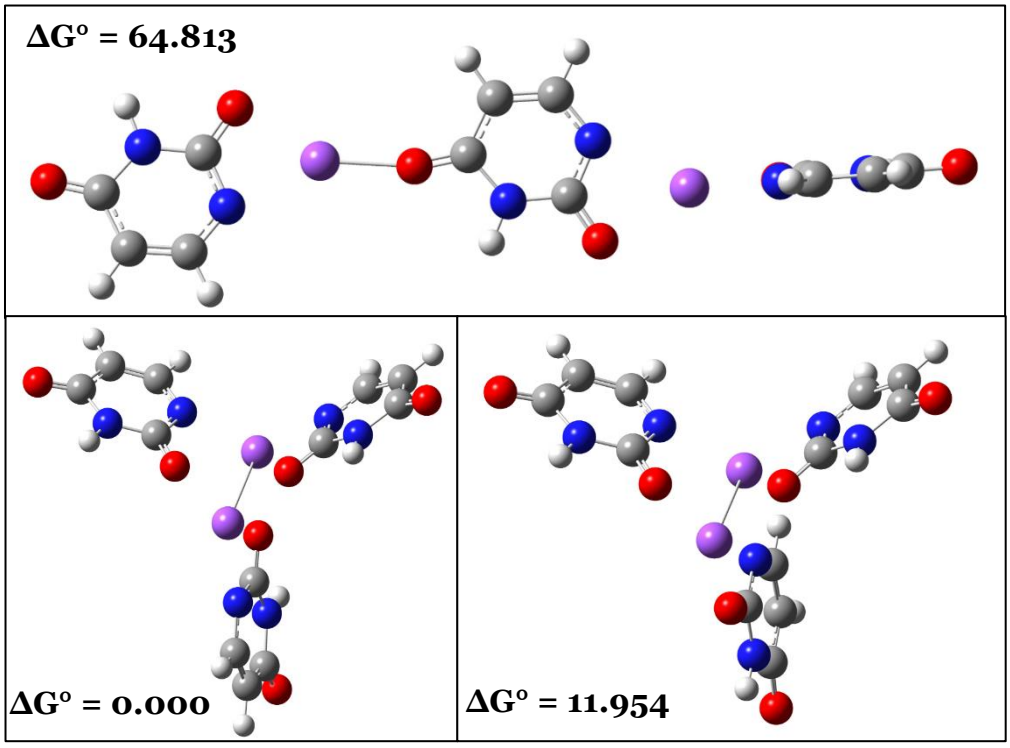


Figure 38: The computationally determined geometric structures for the three arrangements of dU_3Na_2. Top image is the "train" geometry, bottom-left is Propeller A and bottom-right is Propeller B. The relative Gibbs energy is measured in kJ/mol.

Figure 38 highlights the differences in geometries for the three most energetically favourable geometries of dU_3-Na_2. The two Propeller geometries having similar relative energies to each other could be considered to be acceptable experimentally possible isomers of one another. In the case of dT_3-Na_2, the two Propeller geometries differed by ~18 kJ/mol, which is comparable to the difference observed in the dU_3-Na_2 clusters. Therefore, there is more evidence to support the earlier conclusion, regarding the Propeller clusters of dT_3-Na_2, that both Propeller A and Propeller B geometries were observed experimentally.

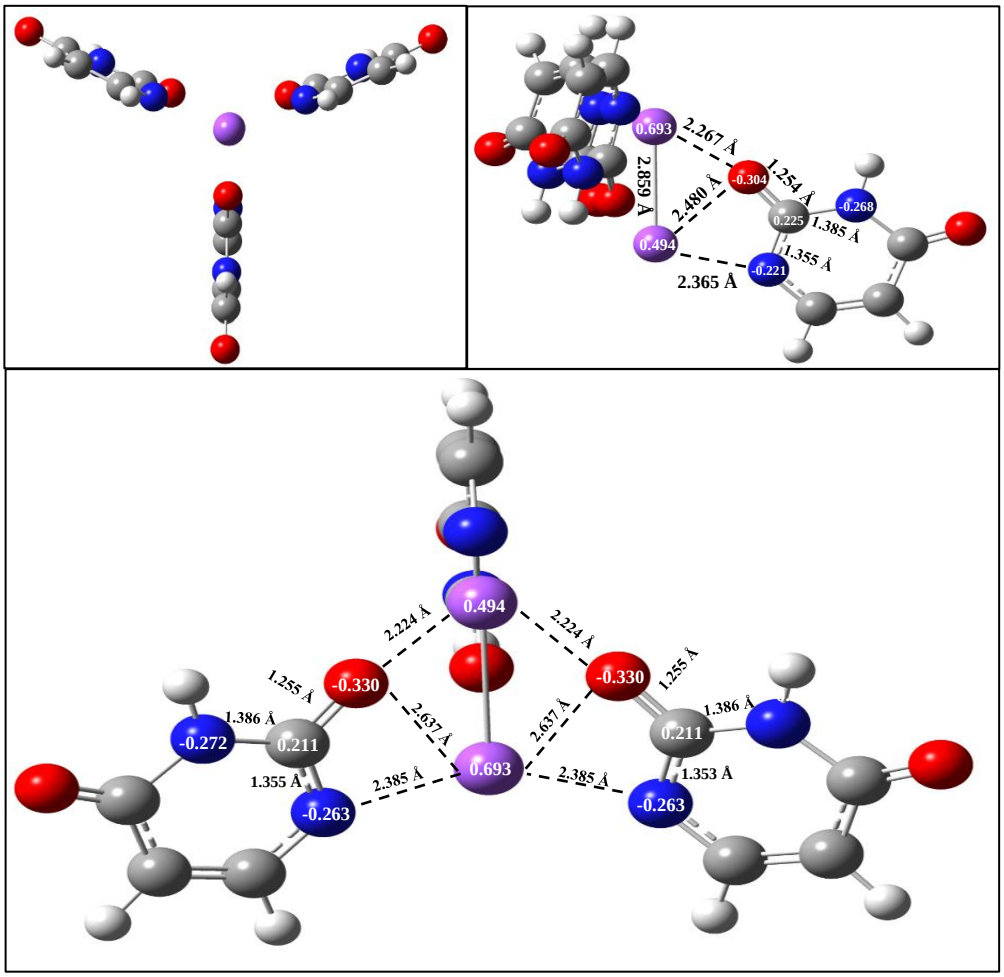


Figure 39: The computationally determined geometric structures for the Propeller A geometry of dU_3-Na_2.

Propeller A, shown in Figure 39, exhibited one plane of symmetry through the Na^+-Na^+ internuclear axis. The two symmetric dU's appear to interact with the sodium cation of charge 0.693 Debye, while the asymmetric dU monomer more favourably interacts with the other sodium cation. Interestingly, this geometry did not minimize the amount of space taken up by the overall system, as with the Propeller A geometry of dT_3-Na_2, which might have contributed to the mixture of dT_3-Na_2 geometries observed experimentally. Following this previous finding, it is likely that dU_3-Na_2 would exhibit similar clustering behaviour; where Propeller A was computationally determined to be the most energetically favourable isomer of dU_3-Na_2 but was experimentally observed to likely be mixed with Propeller B geometry.

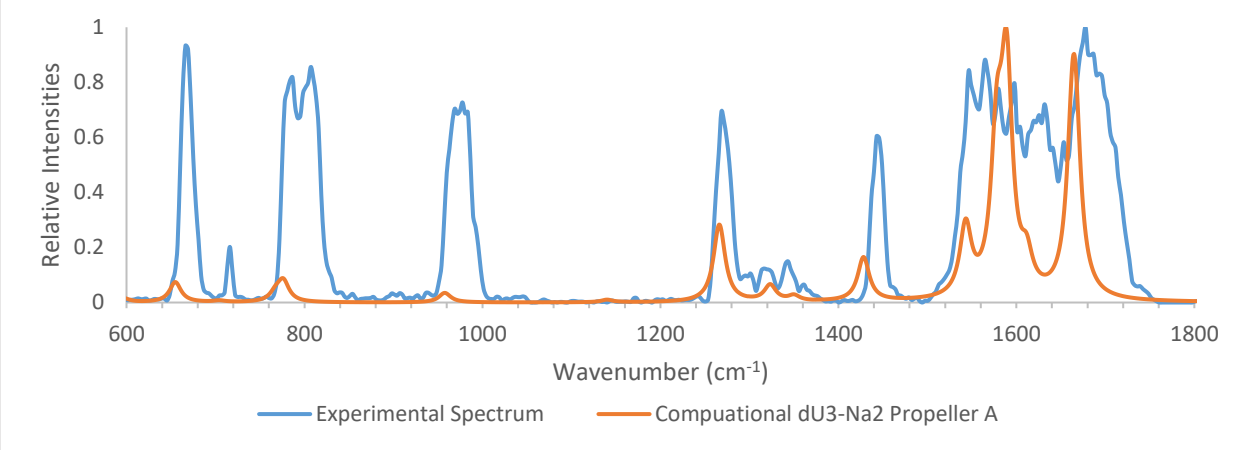


Figure 40: The computationally determined spectrum for the Propeller A geometry of dU_3-Na_2, with the experimental spectrum for dU_3-Na_2.

The vibrational spectrum for the Propeller A geometry of dU_3-Na_2, shown in Figure 40, appeared to have good agreement with a large majority or the major and minor experimental peaks. Apart from the experimental peaks at 717, 1564 and 1631 cm^{-1}, each experimental peak had been matched by a computational peak. The summary of the experimental and computational peaks is shown in Table 14.

Table 14: The experimental and computational frequencies for dU_3-Na_2, the Propeller A geometry.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
666	**652**	Wagging of the N3 protons
717		
786	**776**	Wagging of the C5 protons
978	**958**	Ring breathing
1268	**1266**	Asymmetric rocking of the N3 and C6 protons
1317	**1323**	Rocking of N3 protons
1343	**1351**	Symmetric rocking of C5/C6 protons
1443	**1428**	In-plane symmetric stretching along N1/C2 and N3/C4 bonds
1546	**1542**	Stretching of C5/C6 bond on 2 equivalent dU's
1564		
1580	**1589**	Carbonyl stretching of C2/O
1598		
1631		
1678	**1665**	Carbonyl stretching of C4/O

The Propeller A geometry of dU_3-Na_2 appeared to be missing a peak to characterize the experimental peak of 1631 cm^{-1}. Although they are not equivalent chemical systems, this absence in the computational spectrum around 1630 cm^{-1} was seen in both the dT_3-Na_2 and dU_3-Na_2 Propeller A clusters. Following trends in the clustering behaviour between the dT and dU clusters from lower-complexity clusters, it was expected that Propeller B's geometry would have produced a peak that better mathces the experimental peak around 1631 cm^{-1}. Moreover, this finding in the cluster of dU_3-Na_2 reinforces earlier conclusions that the clustering behaviour between pyrimidines is mostly conserved.

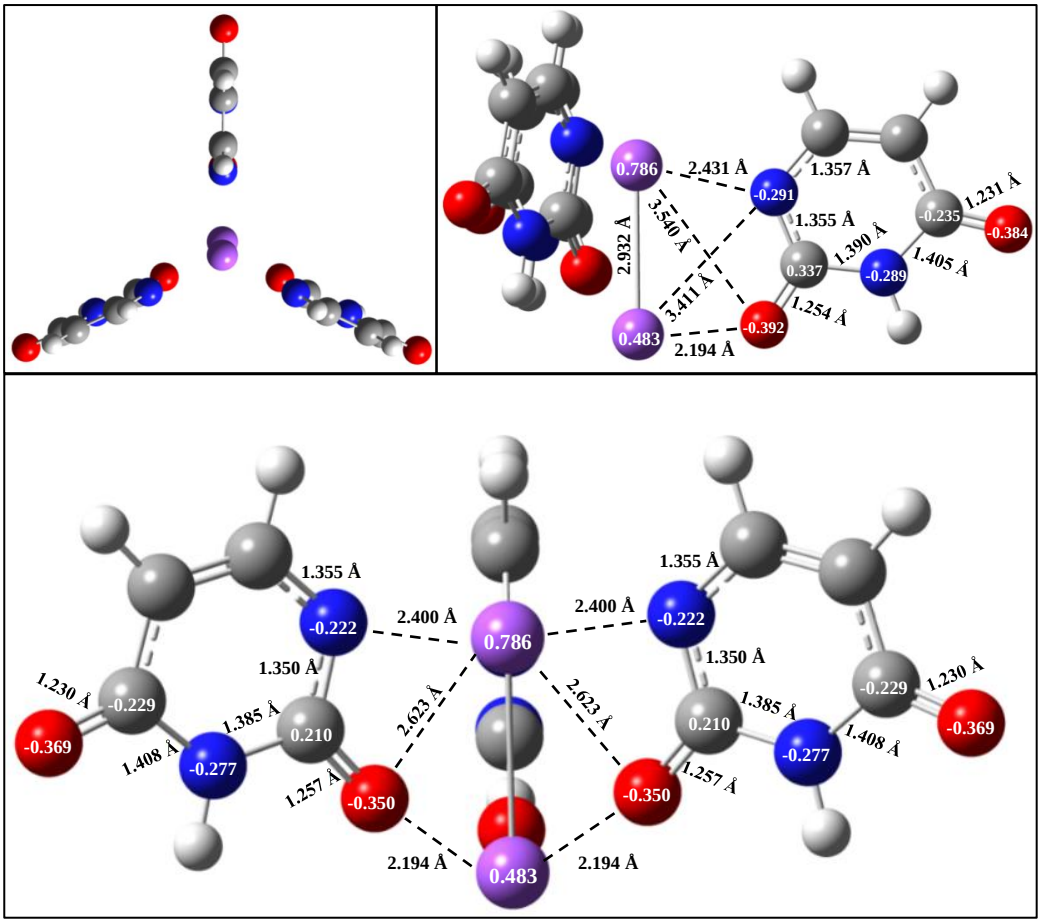


Figure 41: The computationally determined geometry of the dU_3-Na_2 Propeller B cluster.

The Propeller B geometry, found in Figure 41, demonstrates how the energetics of the dU_3-Na_2 chemical system can change significantly through a simple rotation of one of its monomers. The 180° rotation of the asymmetrical monomer in the Propeller A geometry leads to fewer ion-molecule bond interactions as the asymmetric monomer in Propeller B oriented itself in such a way that the N3 and C2-oxygen atom only interact with a single sodium cation at a time. This difference is likely the defining characteristic which allows for the unassigned experimental peak of 1631 cm^{-1} to now be assigned in Propeller B, as seen below in both Figure 42 and in Table 15.

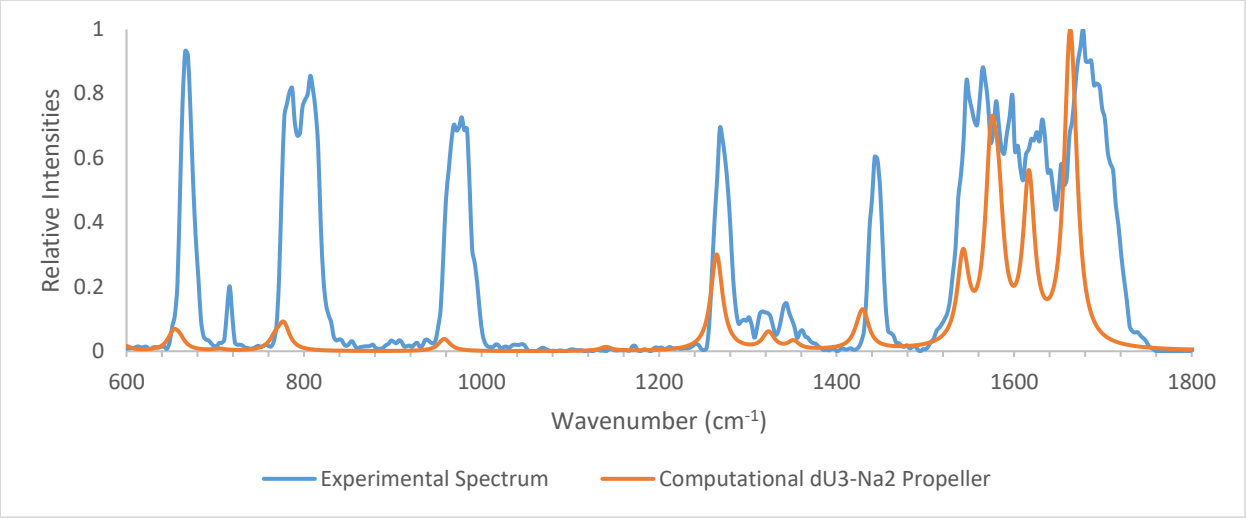


Figure 42: A comparison of the experimentally and computationally determined vibrational spectra for the dU_3-Na_2 Propeller B cluster.

Table 15: The experimental and computational frequencies for the dU_3-Na_2 Propeller B geometry.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
666	**654**	Wagging of the N3 protons on the 2 equivalent dU's
717		
786	**776**	Wagging of the non-equivalent C5 proton
978	**958**	"Pinching" of the N1/N3/C5 atoms on each dU
1268	**1265**	Asymmetric rocking of the N3 and C6 protons on 2 equivalent dU's
1317	**1324**	Rocking of N3 protons on each dU
1343	**1351**	Rocking of N3/C5/C6 protons on non-equivalent dU
1443	**1430**	Symmetric stretching along C2/N3/C4 bond
1546	**1543**	Stretching of C5/C6 bond on 2 equivalent dU's
1564		
1580	**1575**	Stretching of C2/O in 2 equivalent dU's
1598		
1631	**1616**	Stretching of C2/O in non-equivalent dU
1678	**1663**	Stretching of C4/O in each dU

Based on the peak positions listed in Table 15, the rotation of the asymmetric dU monomer led to the splitting of the Propeller A peak at 1590 cm^{-1} into two individual Propeller B peaks at 1575 and 1616 cm^{-1}. In each case, the original Propeller A peak and the two new Propeller B peaks, give vibrational frequencies which are described by the same type of vibrational mode, a C2-oxygen carbonyl stretch. However, in the case of the computational peak at 1616 cm^{-1} in Propeller B, this vibrational mode describes the interaction between a C2-oxygen exclusively interacting with a single sodium cation. This can be supported by observing the geometry of Propeller A, since each of its C2-oxygens appear to be oriented in such a way that they interact with both sodium cations. Given this interesting observation, it is extremely likely that this is the reason why the Train and Propeller B geometries of both dT_3-Na_2 and dU_3-Na_2 appear to have comparable vibrational spectra in the 1500-1670 cm^{-1} region.

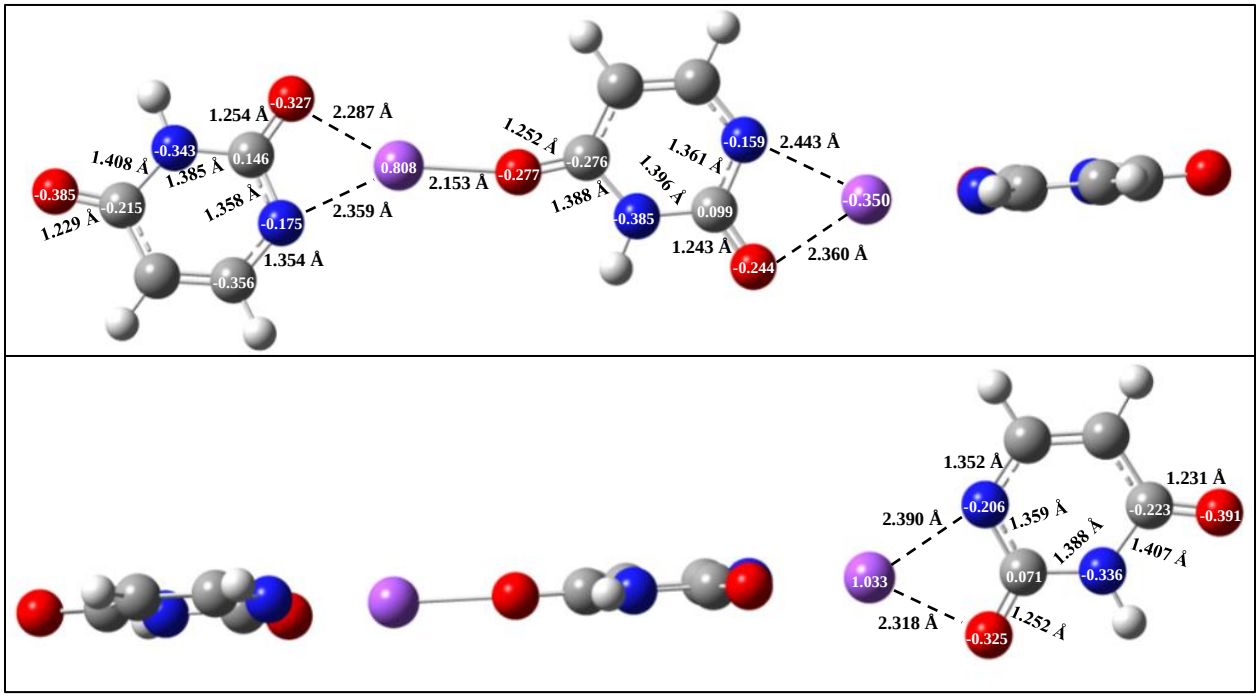


Figure 43: The computationally determined geometry of the dU_3-Na_2 cluster with train geometry.

Through a geometric observation of the Train geometry shown in Figure 43, it was apparent how likely this characterization of the experimental peak at 1630 cm^{-1} was to hold true of all pyrimidine clusters containing a C2-oxygen that exclusively interacts exclusively with one sodium cation. This characterization seemed to hold true as the vibrational spectrum for the Train geometry, in Figure 44, similarly exhibited a peak comparable to the experimental peak at 1630 cm^{-1}.

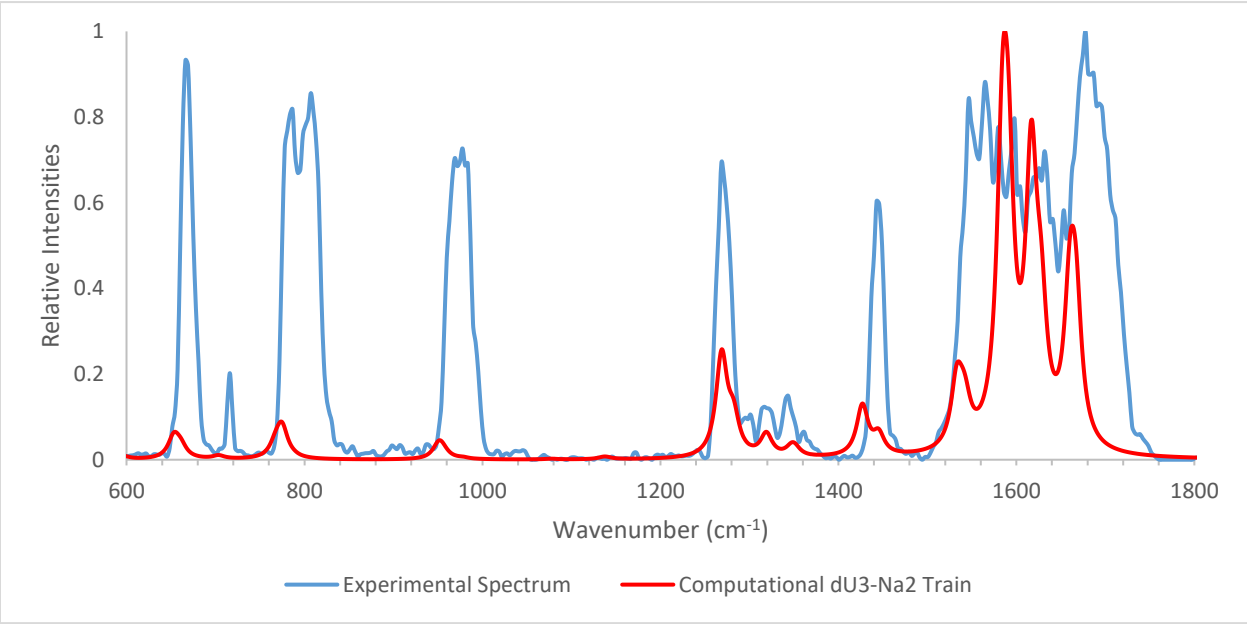


Figure 44: A comparison of the experimentally and computationally determined vibrational spectra for the dU_3-Na_2 train cluster.

Given the similarities between the vibrational frequencies between the Propeller B and Train geometry for dU_3-Na_2, in Tables 15 and 16, it is very likely that the experimental vibrational frequency at 1630 cm^{-1} is uniquely defined by the interaction of a C2-oxygen atom with a single sodium cation, among each of the pyrimidines. Therefore, we would expect to see these trends in the spectra for the clusters of deprotonated cytosine (dC).

Table 16: The experimental and computational frequencies for dU_3-Na_2, the train geometry.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
666	**654**	Wagging of the N3 protons on the 2 equivalent dU's
717		
786	**773**	Wagging of the non-equivalent C5 proton
978	**952**	"Pinching" of the N1/N3/C5 atoms on each dU
1268	**1269**	Asymmetric rocking of the N3 and C6 protons on 2 equivalent dU's
1317	**1318**	Rocking of N3 protons on each dU
1343	**1348**	Rocking of N3/C5/C6 protons on non-equivalent dU
1443	**1426**	Symmetric stretching along C2/N3/C4 bond
1546	**1535**	Stretching of C5/C6 bond on 2 equivalent dU's
1564		
1580	**1582**	Stretching of C2/O in 2 equivalent dU's
1598		
1631	**1615**	Stretching of C2/O in non-equivalent dU
1678	**1663**	Stretching of C4/O in each dU

The observations made for the clustering behaviour of dU parallel with those of dT. These experimental and computational findings strongly support the conclusion that the clustering behaviour of all pyrimidines are conserved among them, and thus can be reliably predicted based off the trends of any one deprotonated pyrimidine. In short, with the limited experimental data we were able to collect, there is strong evidence to support the parallels in clustering behaviour between dT, dU and dC. dC is, therefore, expected to produce two potential clusters for dC_2-Na_2 and three for dC_3-Na_2 in either the Propeller or Train geometry. The Propeller geometries for dC_3-Na_2 will be more energetically favourable than that of the Train, and we would expect to observe a mixture of the two Propeller geometries experimentally.

3.3.1.3 Deprotonated Cytosine and Clusters

Cytosine is a pyrimidine that contains an NH_2 group on the C4 atom and a double bonded oxygen on the C2 group, with three protons that are likely to be favourably abstracted. Two of these three protons are equivalent as they are both NH_2 protons, which means that if either of those two protons were abstracted, the resulting molecule would be equivalent isomers of each other. Therefore, there are only two potential isomers of note, the N1 isomer and the NH_2 isomer. As seen in Figure 12, the isomer deprotonated at the NH_2 site, the NH_2 isomer, was computationally determined to be 143.4 kJ/mol greater in energy than that of the N1 deprotonated cytosine (dC). Energetically, this difference in relative Gibbs energy suggests that the most stable isomer of dC is the N1 isomer and would then require further validation via experimental vibrational spectra. An exclusively computational assessment, with expectations from the investigation on the two other pyrimidines, is therefore made as there were no experimental spectra for the clusters of dC to evaluate.

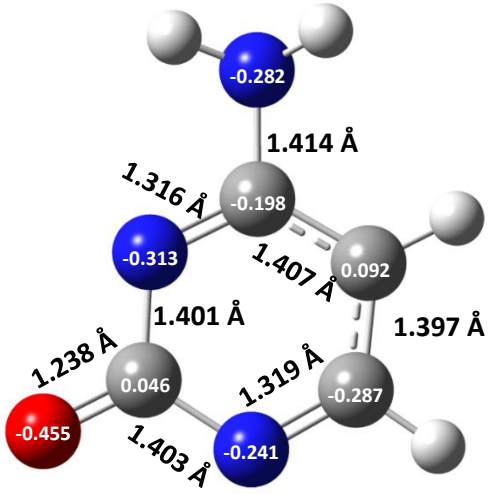


Figure 45: The computationally determined geometry for the most energetically favourable isomer of deprotonated cytosine (dC).

Figure 45 highlights the electrostatics of the monomer for the most energetically favourable dC. Before moving forward with this monomer, in order to build the clusters of dC with sodium cations, the two isomers must be further validated experimentally. Figures 46 and 47 will compare the vibrational frequencies of the NH_2 and N1 isomer to the experimental vibrational frequencies, respectively.

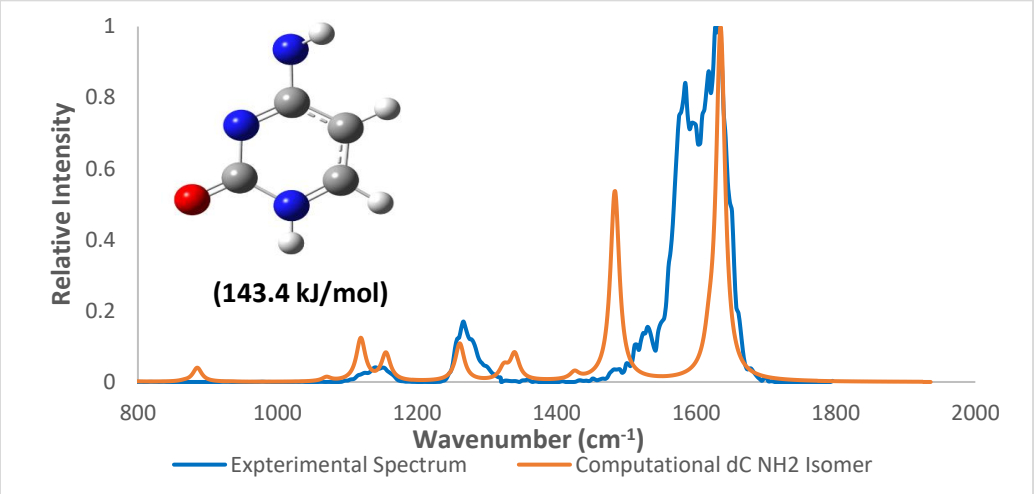


Figure 46: The computationally and experimentally determined vibrational spectra for deprotonated cytosine, with the computational spectrum referring to the loss of a proton from cytosine's NH_2 group.

Despite the energetic unfavourability of the dC-NH2 isomer compared to the dC-N1 isomer, as shown in Figure 46, three of the five major experimental peaks of dC appear to be modelled by some of the major computational peaks in the NH2 isomer. However, there appeared to be two experimental peaks, at 1533 and 1585 cm^{-1}, that were not in good agreement with the other computational peaks from dC-NH2. As a result, there remained a possibility for the experimental spectrum to have captured the vibrational frequencies for a mixture of dC isomers. Given that the two most energetically favourable isomers of dC were the dC-N1 and dC-NH_2 isomer, it is possible that despite their large difference in relative Gibbs energy, a mix of these isomers were observed experimentally.

Table 17: The computed vibrational frequencies for dC NH_2 isomer and the experimental vibrational frequencies for dC.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1151	**1119**	Rocking of the NH proton
	1156	Symmetric rocking of the N1 and C6 protons, asymmetric to the rocking of the C5 proton
1266	**1262**	Stretching along C2/N3 bond
1533		
1585		
1633	**1635**	Carbonyl stretch along C2/O bond

Two of the major computational peaks at 1262 and 1635 cm^{-1}, shown in Table 17, have excellent agreement with the experimental peaks of 1262 and 1633 cm^{-1}. Therefore, it is likely that the two experimental peaks are best characterized by the stretching of the C2/N3 bond at 1262 cm^{-1} and by the in-plane stretch of the C2/O bond at 1635 cm^{-1}. Moreover, the experimental peak at 1151 cm^{-1} appears to be well-characterized by two computational peaks at 1119 and 1156 cm^{-1}. There is, however, a relatively intense computed peak at 1482 cm^{-1} that does not appear to characterize any experimental peak shown in Figure 46. This might be due to an issue with attenuation that could have prevented that vibrational mode from being sufficiently excited in order to match the intensity predicted by the computational spectrum. Interestingly, as shown in Figure 46, the computational peak around 1482 cm^{-1} is also found in the computational spectrum for the NH1 dC isomer. Therefore, it was necessary to further explore the differences in the computational vibrational spectra of these two isomers to better characterize the vibrational frequencies observed experimentally.

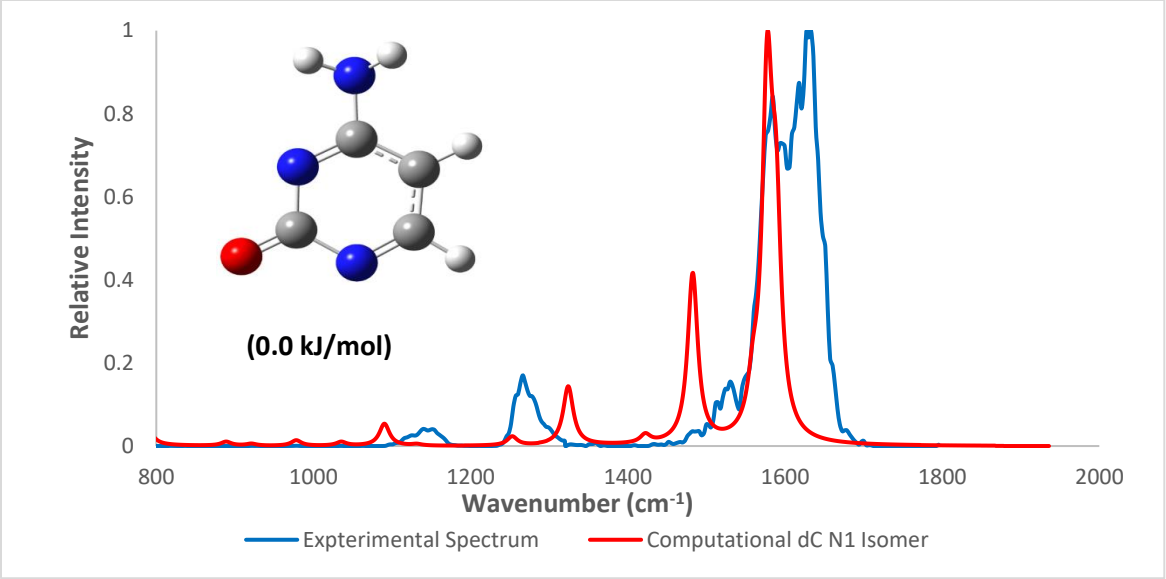


Figure 47: The computationally and experimentally determined vibrational spectra for deprotonated cytosine, with the computational spectrum referring to the loss of a proton from cytosine's N1 atom.

The major experimental peak of dC at 1585 cm^{-1} appears to show excellent agreement with the computational peak of the dC-N1 isomer at 1578 cm^{-1}, shown in Figure 47. Due to the experimental spectrum of dC appearing to be characterized by the most intense peak in the dC-N1 isomer, there is strong evidence to support that the experiment very likely captured characteristics of a mixture of dC isomers.

Table 18: The computed vibrational frequencies for dC-N1 isomer and the experimental vibrational frequencies for dC.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1151		
1266		
1533		
1585	**1578**	Symmetric in-plane stretching of N1/C6, C2/O and N3/C4 bonds, with slight proton scissoring.
1633		

Initially, as shown in Table 18, there appears to be excellent agreement between the most intense computational peak in the NH1 isomer and the experimental peak at 1585 cm^{-1}. The overall agreement between the computational spectrum of the dC-NH_2 isomer and the experimental spectrum is much better. However, the increased agreement in the spectrum for the dC-NH_2 isomer does not mean the isomers must exist independently of each other. Therefore, based on the energetic analysis and vibrational spectra for dC and its isomers, there was a mixture of isomers observed experimentally and supported the dC-N1 isomer as the dominant isomer between the two. This provided the foundational isomer used to build higher-complexity clusters of dC and Na^+. The clusters listed in Table 19 were computationally evaluated based on their energetics.

Table 19: The thermodynamic properties of reaction for the clusters of dC-N1 and sodium cations.

Reaction	ΔG° (kJ/mol)	ΔH° (kJ/mol)	ΔS° ($J \cdot K^{-1} \cdot mol^{-1}$)
$[dC]^- + [Na]^+ \rightarrow [dC][Na]$	-576.637	-610.582	-27.211
$[dC][Na] + [dC]^- \rightarrow [dC]_2[Na]^-$	-181.539	-219.996	-30.828
$[Na][dC]_2^- + [Na]^+ \rightarrow [dC]_2[Na]_2$ Propeller Geometry	-563.878	-600.020	-28.973
$[dC]_2[Na]_2$ (Propeller) + $[dC]^- \rightarrow$ $[dC]_3[Na]_2^-$ Propeller A Geometry	-154.293	-201.311	-37.691
$[dC]_2[Na]_2$ (Propeller) + $[dC]^- \rightarrow$ $[dC]_3[Na]_2^-$ Propeller B Geometry	-112.497	-155.297	-34.310

Unfortunately, there are no experimental spectra for the clusters of deprotonated cytosine, however, further analysis of the clusters based on their computationally determined energetics, geometries and vibrational modes was investigated.

Despite not being able to evaluate the experimental vibrational spectra of the formed neutral clusters, the computational analysis revealed clusters of dC would greatly favour complexation with small cationic species, as shown in Table 19. In addition, when comparing the complexing favourability for forming the neutral clusters as well as the ionic clusters, as the cluster increases in size and complexity, the favourability of formation decreases slightly. As this trend was not observed in either cluster of dT or dU, it would be interesting to see if this trend is continued in other deprotonated nucleobases. In future work, it would be best to determine if the favourability of reaction continues to decrease as we continue to increase the size and complexity of the clusters.

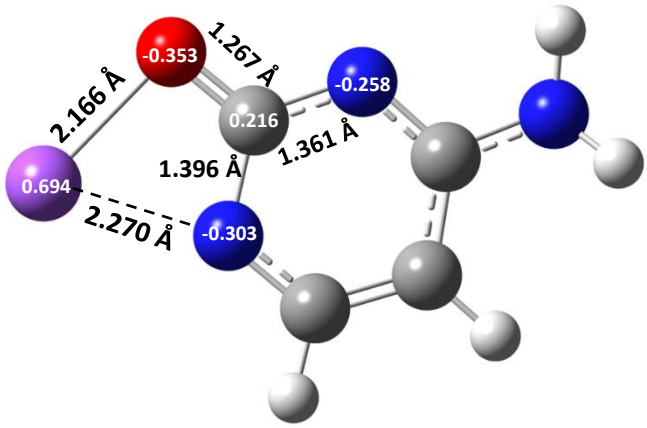


Figure 48: The computationally determined geometry for dC-Na.

The geometric optimization of dC-Na, shown in Figure 48, revealed the most favourable site of interaction between a dC monomer and a sodium cation to be the N1/C2-oxygen region. Like the other pyrimidines, this region appeared to be the basis for ion-molecule interactions within each cluster of dC, as we will see in the following sections. It was interesting to note that steric hindrance likely played a large role in determining the most favourable site of interaction in the dC-Na cluster, seeing how the C2-oxygen/N3 region on the dC-N1 isomer, exhibited more negative charge than the N1/C2-oxygen region.

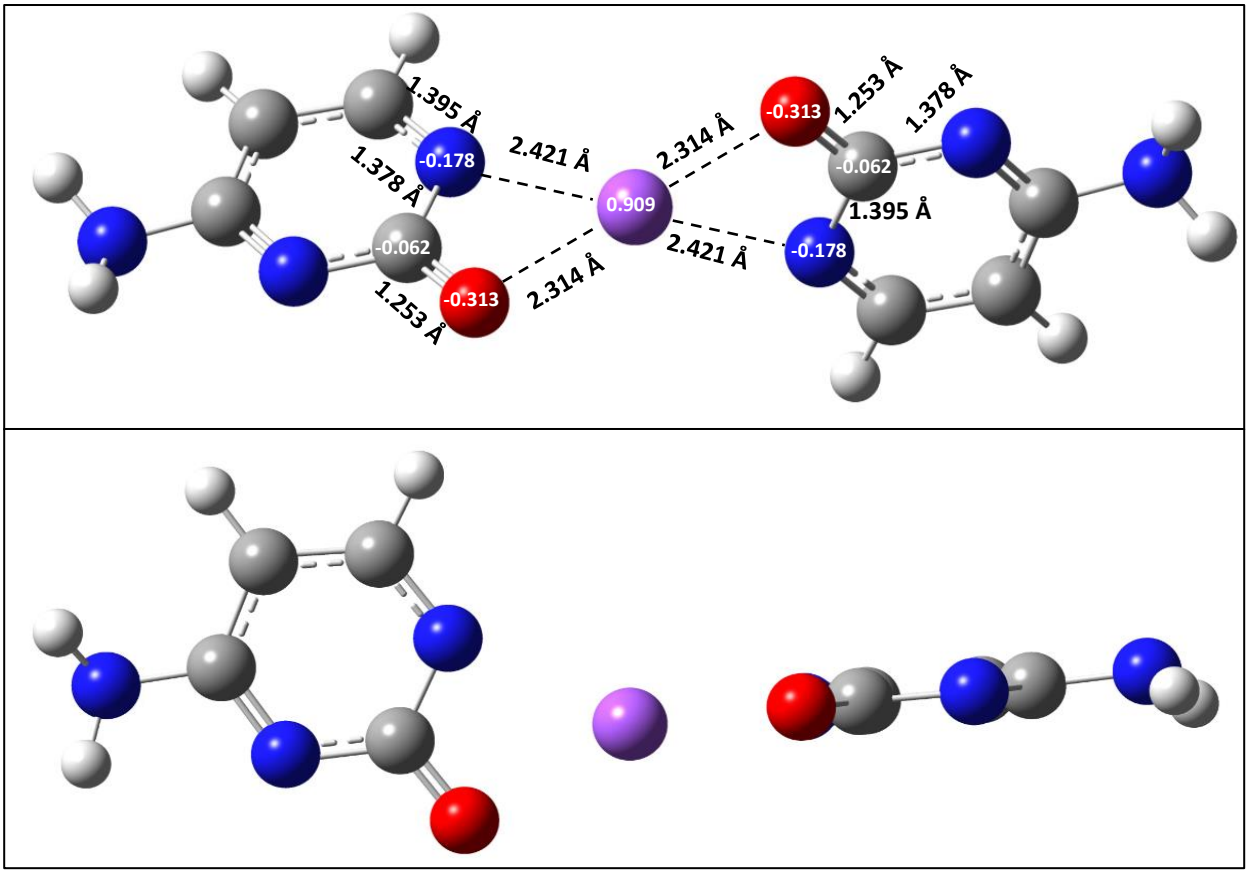


Figure 49: The computationally determined structure for two deprotonated cytosines with one cationic sodium (dC2-Na).

The addition of another dC, shown in Figure 49, has led to increased symmetry about the cationic sodium. The symmetry exhibited by the dC2-Na cluster is a trait shared among the pyrimidines when forming this particular type of cluster. The bond distances of the Na^+/O and Na^+/N1 interactions have increased compared to those of the dC-Na cluster. This would suggest that the ion-molecule interactions between the sodium cation and each individually interacting atom has became weaker through the addition of a dC. In addition, the bond distances between the C2 atom and its neighbours has not significantly changed, although they are different. Based on the observations from the dT and dU evaluations, it would make sense to expect vibrational modes that are simultaneously present in both the dC's because of the increased symmetry in this cluster.

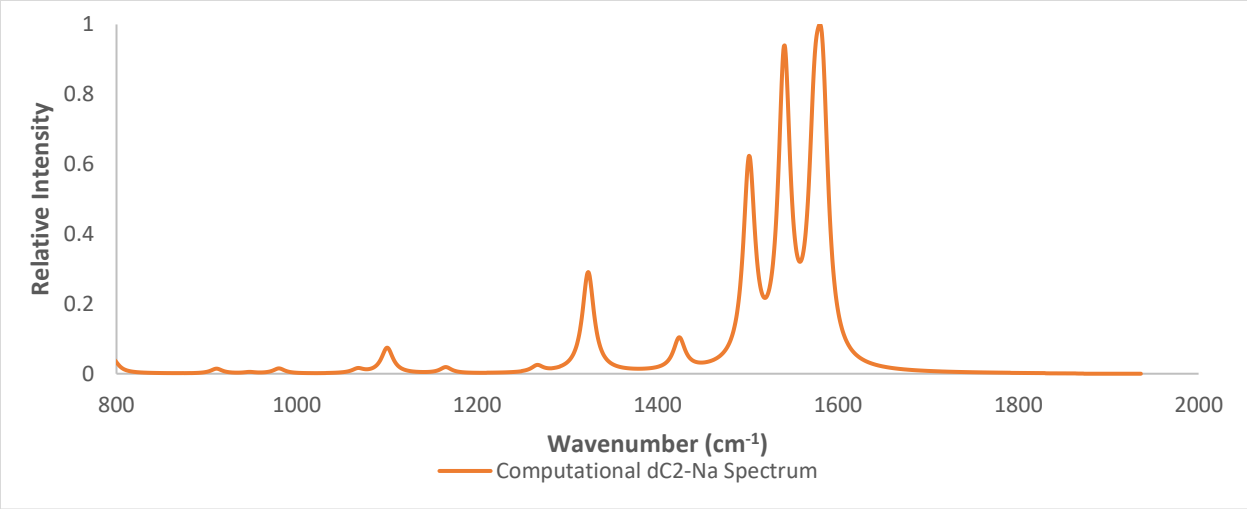


Figure 50: The computationally determined vibrational spectra for dC2-Na.

The computationally determined vibrational spectrum for dC2-Na is shown in Figure 50. The assumption, based on our earlier findings with dT and dU, that each major peak would describe the same vibrational motion on both the dC's was correct. However, in some cases, there were two nearly identical peaks, making up a major peak, that described the same vibrational motion but in one dC monomer. Therefore, it could be possible to see the well-defined peaks of the computational spectrum, at peak positions shown in Table 20, each split into two nearly identical peaks in the experimental spectrum.

Table 20: The computed vibrational frequencies for dC_2-Na cluster.

Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1100	Proton rocking of NH_2 protons on both dC's
1323	Scissoring of the C5/C6 protons
1424	Asymmetric stretching of the C6/N1 and N3/C4
1501	Asymmetric stretching along the C4/C5/C6 bond
1540	Stretching of C2/O bond
1582	Scissoring of NH_2 protons

Based on the vibrational frequencies shown in Table 20, some peak positions from the experimental dC spectrum were conserved in the dC-Na computational spectrum. For example, the dC experimental peak of 1431 cm^{-1} was shown as a low-intensity computational peak describing the asymmetric stretch of the N1/C6 and N3/C4 bonds and might have been conserved in the computational dC_2-Na peak at describing the same motion. Similarly, the asymmetric in-plane stretching along the C4/C5/C6 bonds previously characterized by a peak at 1482 cm^{-1} in the computational dC-N1 spectrum, appears in the dC_2-Na spectrum as a peak at 1501 cm^{-1}. There remains a possibility that these two vibrational modes would likely be experimentally seen throughout all the dC clusters, as intrinsic peaks unique to the dC monomers. Interestingly, other peaks with complex vibrational modes, in the dC experimental spectrum, appear to have been split-up in the computational spectrum of dC_2-Na. The experimental peak of dC at 1579 cm^{-1}, was previously assigned as a frequency describing the symmetric in-plane stretching of the N1/C6, N3/C4 and C2/O bonds, with slight NH_2 proton rocking. It appeared as if that singular peak has been broken up into two new peaks found in the vibrational spectrum for dC_2-Na at 1540 and 1582 cm^{-1}. The 1540 cm^{-1} peak now exclusively describes the carbonyl stretch of the C2/O bond, while the 1582 cm^{-1} peak now exclusively describes NH_2 scissoring. It is possible that the additional symmetry in the chemical system, resulting from interactions with a cationic sodium, led to fewer degrees of vibrational freedom and in turn made several more vibrational modes more pronounced.

The previous investigations into the clusters of dT and dU revealed both experimentally and computationally that the two-pyrimidine monomer with two sodium cation clusters (dX_2-Na_2) likely favour the formation of the propeller geometry over the Train geometry.

Therefore, only the Propeller geometries for the dX_2-Na_2 and dX_3-Na_2 clusters will be considered. In Figure 51, the Propeller geometry and charges for dC_2-Na_2 are shown.

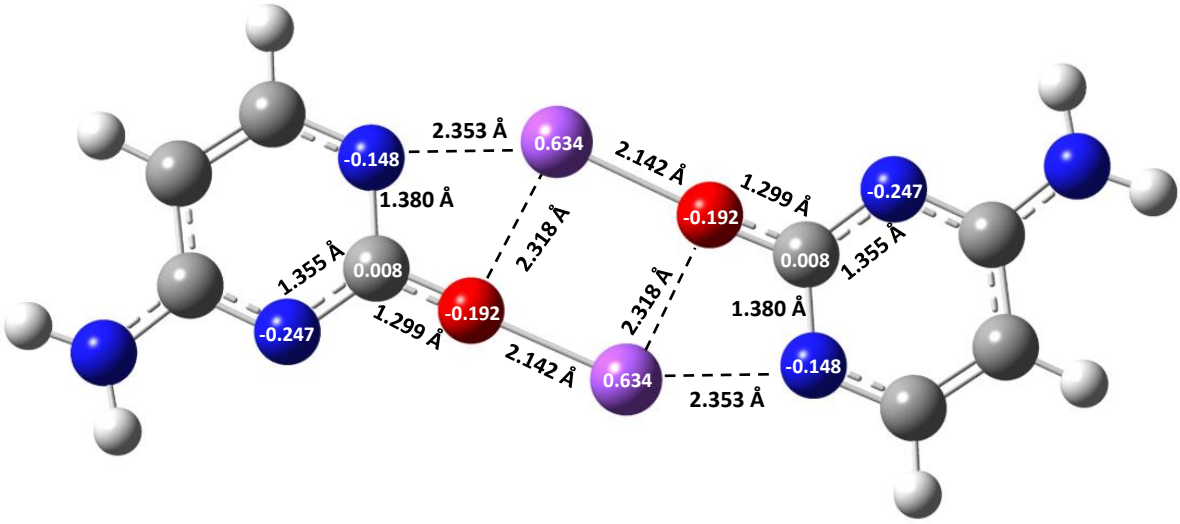


Figure 51: The computationally determined Propeller geometry for the dC_2-Na_2 cluster.

The structure for dC_2-Na_2, shown in Figure 51, demonstrated perfect symmetry in its geometric orientation and charge distribution, similar to the other two pyrimidines. Once again, the computational analysis of dC clusters, parallels the trends observed in the other pyrimidine clusters.

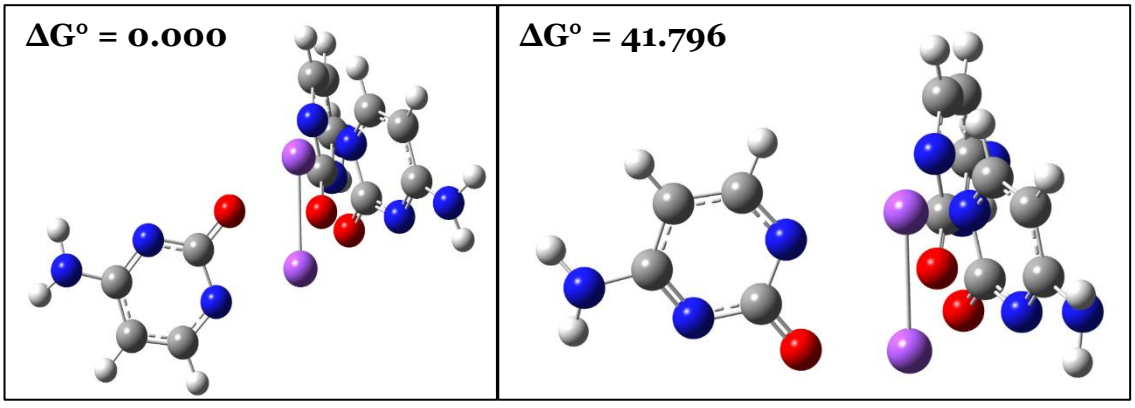


Figure 52: The computationally determined relative Gibbs energies (kJ/mol) for the dC_3-Na_2 clusters. Propeller A geometry on the left and Propeller B geometry on the right.

The difference in relative Gibbs energy between the Propeller A and B geometries, shown in Figure 52, is far more pronounced in the dC_3-Na_2 clusters than in the clusters of dT and dU. In both the other two pyrimidine clusters, Propeller A and B differed in relative energies by less than 20 kJ/mol. It made sense for them to have smaller relative energies to one another as it was previously observed that there was a mixture of the Propeller geometries for both dT and dU clusters. However, in the case of dC_3-Na_2, the difference in relative energies was nearly 42 kJ/mol. This can imply that it would be far less likely for a mixture of these two isomers of the dC_3-Na_2 clusters to exist in a mixture, and that only the Propeller A geometry would be observed experimentally.

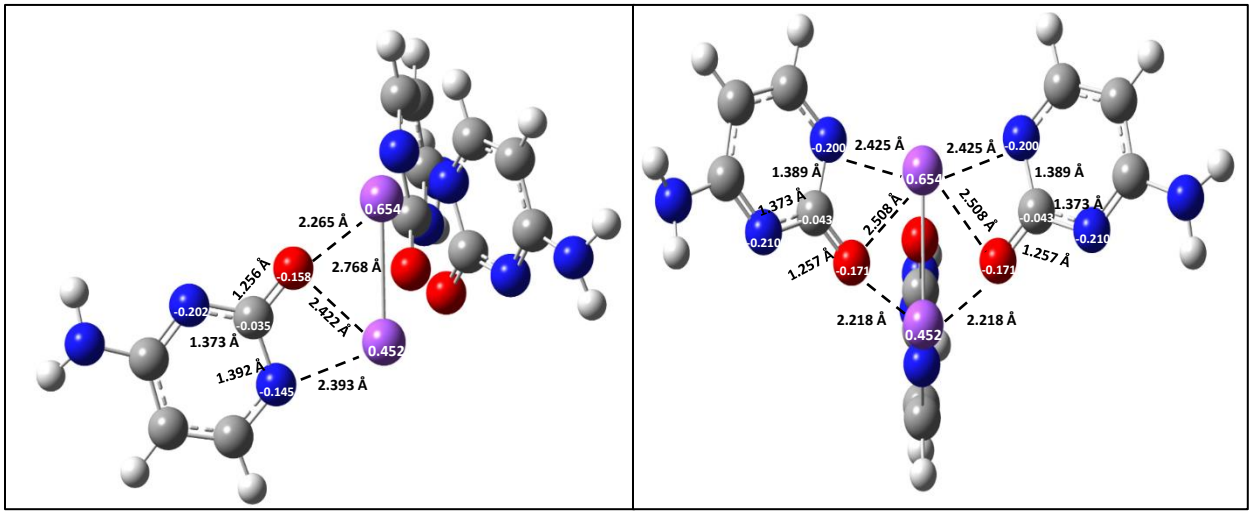


Figure 53: The computationally determined Propeller geometry for the dC_3-Na_2 Propeller A cluster.

The same characteristics of the other pyrimidines for the Propeller A geometry was observed in the geometry shown in Figure 53. A plane of symmetry can be made through the plane of the non-equivalent dC monomer and equally distribute charge among the other two dC monomers. In addition, it is clear that every Propeller A geometry for the dX_3-Na_2 clusters of the pyrimidines observes all C2-oxygens interacting with both of the

sodium cations. Therefore, we can expect to observe a missing peak around 1630 cm^{-1} in the vibrational spectrum for dC_3-Na_2 found in Propeller B, but not Propeller A.

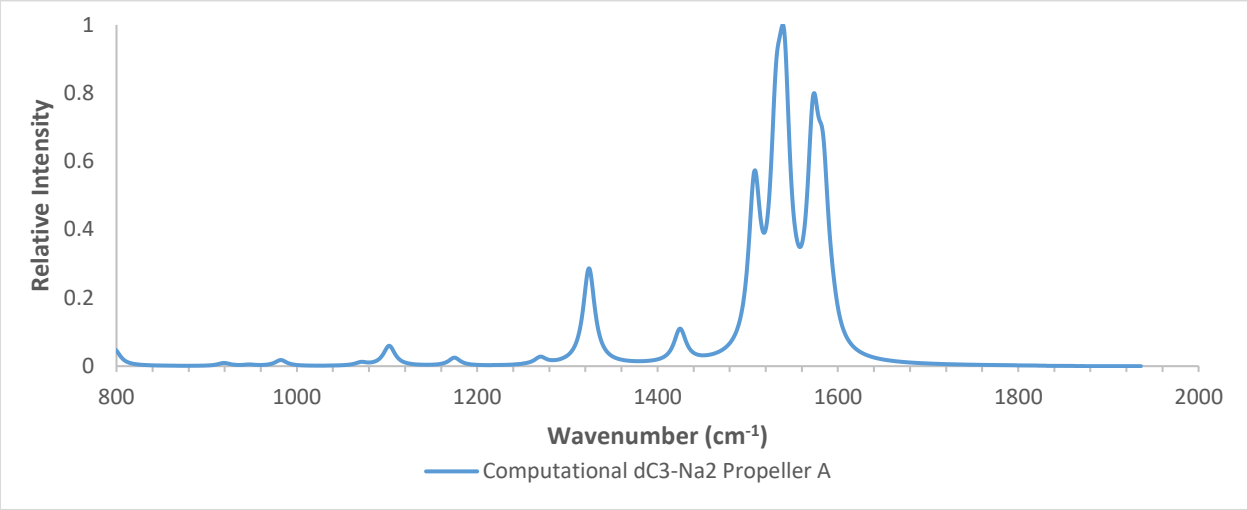


Figure 54: The computationally determined vibrational spectrum for the Propeller A geometry of the dC_3-Na_2 cluster.

Without the experimental spectrum for the dC_3-Na_2 cluster it was impossible to validate any of the peak assignments for the vibrational spectrum in Figure 54. The specific peaks observed in the computational spectrum can be seen in Table 21 but are described by exclusively computational vibrational modes.

Table 21: The computed vibrational frequencies for dC_3-Na_2 Propeller A cluster.

Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1107	Proton rocking of NH_2 protons on both dC's
1327	Ring deformation on equivalent dC's
1430	Rocking of C5/C6 protons on equivalent dC's
1510	Stretching along the C4/C5 on equivalent dC's
1541	Stretching of C2/O bonds
1575	Scissoring of NH_2 protons

It would be difficult to compare the structure of the two Propeller geometries for dC_3-Na_2 without the experimental spectrum to validate our computational analysis. Therefore, the geometry and vibrational spectrum for the Propeller B geometry are briefly evaluated against the Propeller A geometry.

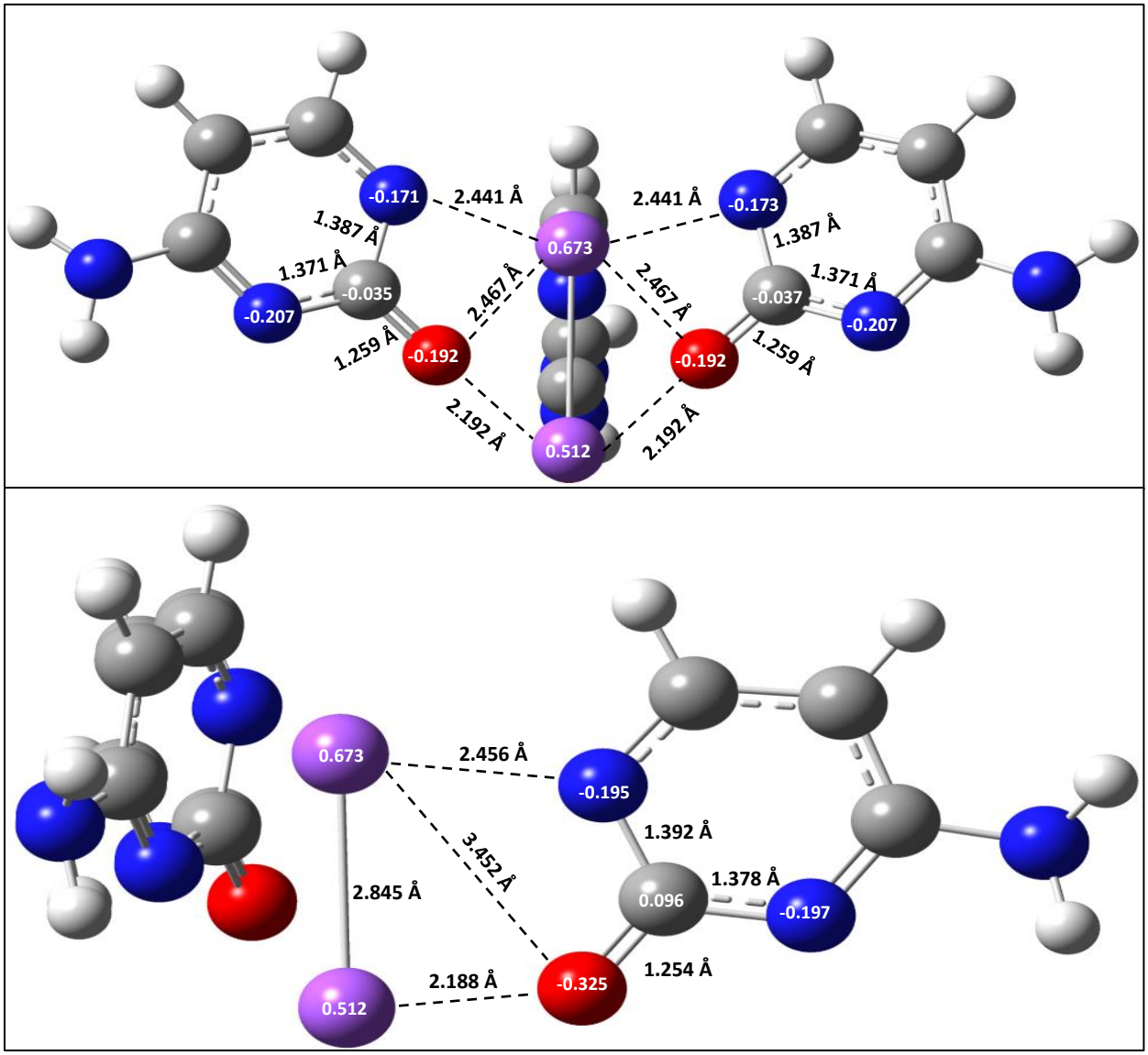


Figure 55: The computationally determined geometry for the dC_3-Na_2 Propeller B cluster.

The cluster shown in Figure 55 demonstrates the same symmetry observed in the other Propeller B pyrimidine clusters and appears to similarly demonstrate interaction between

the C2-oxygen and N3 atoms, in its non-equivalent dC monomer, to one sodium cation at a time. Thus, there should be a vibrational frequency seen in the Propeller B vibrational spectrum that was not seen in the Propeller A spectrum.

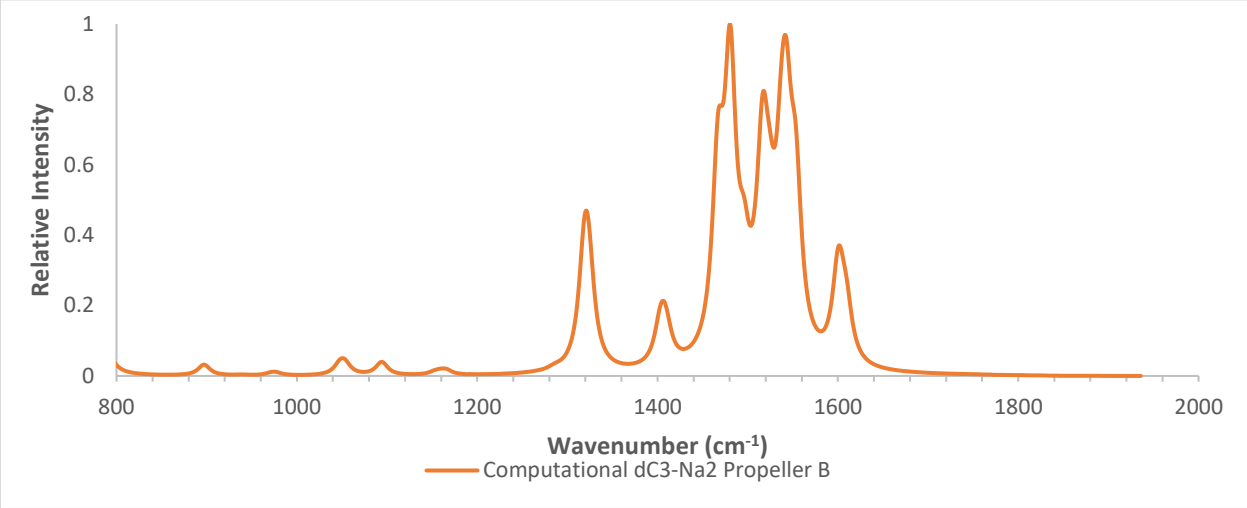


Figure 56: The computationally determined vibrational spectrum for the dC_3-Na_2 Propeller B cluster.

There appear to be 8 major and minor peaks of note in the computational spectra for the Propeller B geometry of the dC_3-Na_2 cluster, shown in Figure 56, as opposed to the 6 major and minor peaks seen in the Propeller A geometry, as seen in Figure 39. The peak positions for the Propeller B geometry are listed in Table 22.

Table 22: The computed vibrational frequencies for dC_3-Na_2 cluster, for the Propeller B geometry.

Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1057	Scissoring of the C5/C6 protons
1099	Rocking of NH_2 protons
1324	Ring deformation
1407	Symmetric C5/C6 proton rocking equivalent dC's
1480	Symmetric C5/C6 proton rocking non-equivalent dC
1518	Stretching of C5/C6 bond
1541	Carbonyl stretch of C2/O with slight NH_2 scissoring
1602	Symmetric in-plane stretch of N1/C6, C2/O and N3/C4 bonds

The vibrational frequencies presented by Table 22 and Table 21 demonstrate the intrinsic differences in the Propeller A and Propeller B geometries highlighted in the earlier observations into the dT and dU clustering behaviour. In every case of clustering with pyrimidines, the most energetically favourable geometry of the higher-complexity clusters was found to be the Propeller geometry. However, it was experimentally observed that there was a mixture of Propeller geometries in the vibrational spectra of the dT and dU clusters, therefore, this same trend is to be expected when the experimental vibrational spectra for the clusters of dC are taken.

3.3.2 Characterization of Deprotonated Purine-anion Clusters

The investigation into the clustering behaviour of the deprotonated nucleobases has so far focused on the pyrimidines, as the only experimental high-complexity cluster spectra were taken for clusters of dT and dU. It revealed several interesting trends and novel geometries that appear to be conserved between pyrimidines. Thus, an investigation into the clustering behaviour of the deprotonated purines, such as adenine and guanine, might reveal trends and clustering behaviours exclusive to the purines.

3.3.2.1 Deprotonated Adenine & Clusters

Of the two potential isomers for deprotonated adenine, the NH_2 and the N9 isomers, the N9 isomer (Figure 57) was found to be roughly 170 kJ/mol more energetically favourable than that of its NH_2 isomer.

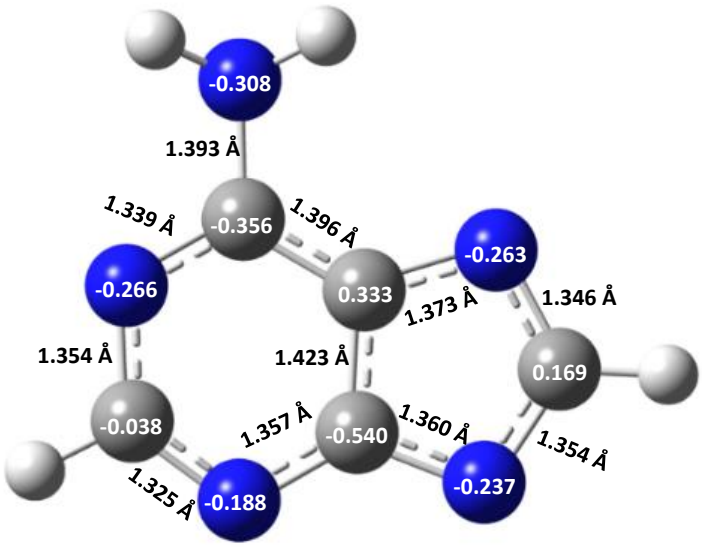


Figure 57: Calculated lowest energy isomer for deprotonated adenine, the N9 isomer.

This difference in energetic favourability suggests that there should experimentally exist a greater population of the N9 isomer than that of the NH_2 isomer. The relative population of the NH_2 isomer, relative to the more energetically favourable N9 isomer, would be roughly 1:(7.04×10^{29}). To confirm this energetic unfavourability, a negative mode vibrational spectrum was obtained for the deprotonated adenine species and was evaluated against the computational vibrational spectra of both the N9 isomer (Figure 58) and the less favourable NH_2 isomer (Figure 59).

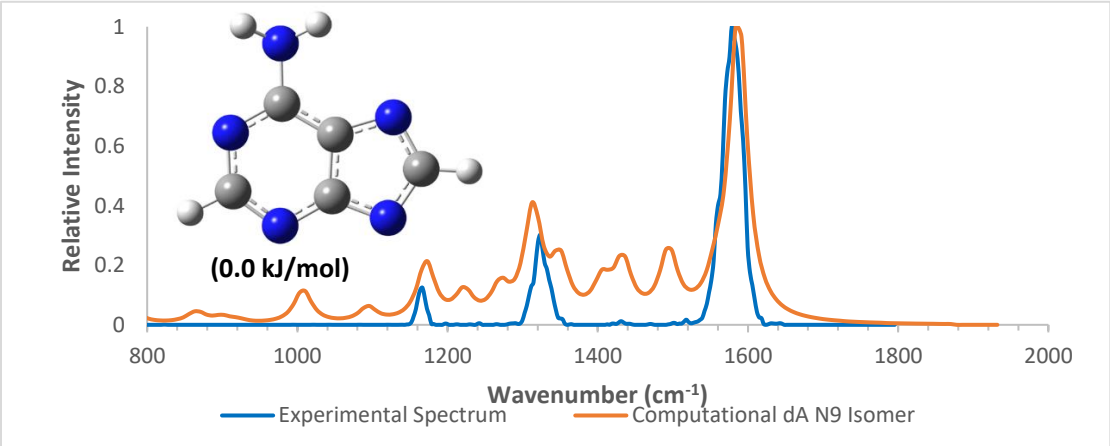


Figure 58: The experimentally determined vibrational spectrum for dA compared to the computational vibrational spectrum of the dA-N9 isomer.

Figure 58 demonstrates the best match between the computational spectrum of the N9 isomer and the experimental spectrum of deprotonated adenine out of all the isomers of dA. All three of the major peaks in the experimental spectrum at 1157 cm^{-1}, 1328 cm^{-1} and 1583 cm^{-1} are accounted for in the computationally generated spectrum. In addition, the two remaining minor experimental peaks, at 1425 and 1520 cm^{-1}, were found to have good agreement with some of the other computationally determined peaks for the N9 isomer, shown in Table 23 below.

Table 23: The experimental peak positions for dA, with assigned peaks from the dA-N9 isomer computational spectrum.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1157	**1172**	C8 proton rocking
1328	**1314**	Stretch of the C4/C5
1425	**1433**	Rocking of C2 proton
1520		
1583	**1587**	Scissoring of NH_2 protons with stretch of C5/C6

Despite the good agreement between the three largest peaks between the computational and experimental spectra, there are several peaks in the computational N9 isomer spectrum that do not appear in the experimental spectrum of deprotonated adenine. Namely, the peaks at 1008 cm^{-1} and 1094 cm^{-1} do not appear to exist in the experimental spectrum. It is possible that the experimental peaks were not clearly observed due to attenuation issues, however, these discrepancies can suggest that there might be some potential for the NH_2 isomer to exist in tandem with the N9 isomer.

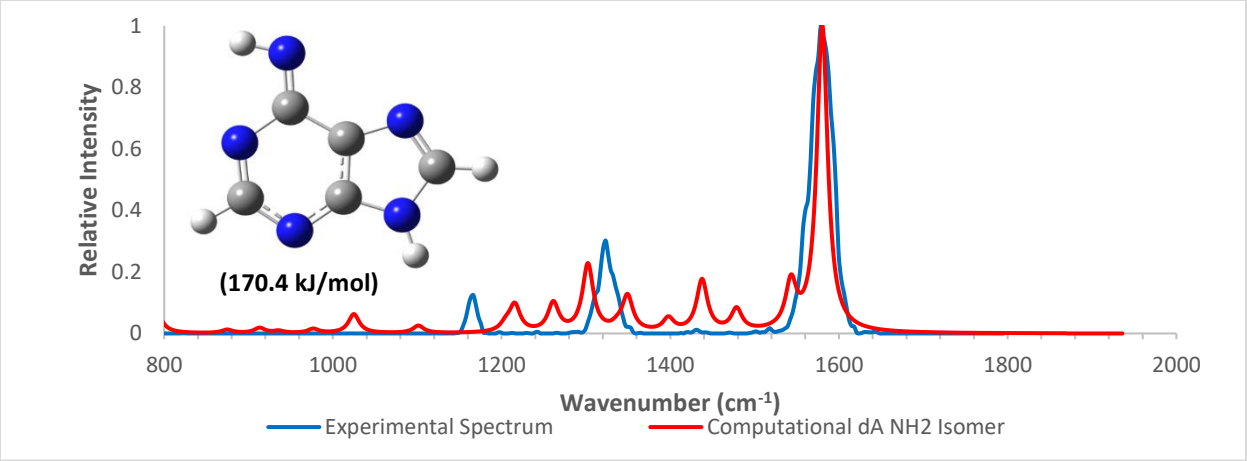


Figure 59: The experimentally determined vibrational spectrum for dA, compared to the computationally determined vibrational spectrum of dA, with proton loss at NH_2.

Figure 59 compares the vibrational spectrum of the second isomer of adenine (the NH_2 isomer) against the experimental spectrum of deprotonated adenine. Interestingly, the agreement between the largest peaks of the experimental and computational spectra is almost exact to one another but represent completely different vibrational modes. This phenomenon might imply that there is some of the NH_2 isomer existing simultaneously to the N9 isomer, however, the lack of agreement throughout the rest of the computational spectrum of NH_2 with the experimental spectrum appears to refute that possibility.

Table 24: The experimental peak positions for dA and the assigned peak positions from the dA-NH_2 isomer.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1157		
1328		
1425	**1437**	Stretch of N7/C8
1520		
1583	**1580**	Symmetric stretch of N1/C2, C4/C5 & C6/NH

As seen in Table 24, there is far less agreement between the experimental peaks of dA and the computational peaks for the dA-NH_2 isomer. This would suggest that the deprotonation at the N9 position, is the unique identifier for the experimentally produced deprotonated adenine. In other words, the dA-N9 isomer is the most energetically favourable isomer of dA and was effectively predicted by the computational evaluation of relative Gibbs energies. Therefore, the dA-N9 isomer was used as the foundation for building clusters of dA and sodium cations, producing energetic property values as seen in Table 25.

Table 25: The computed thermodynamic properties of reaction, for the formation of clusters at increasing levels of complexity.

Reaction	$\Delta G°$ (kJ/mol)	$\Delta H°$ (kJ/mol)	$\Delta S°$ ($J \cdot K^{-1} \cdot mol^{-1}$)
$[dA]^- + [Na]^+ \rightarrow [dA][Na]$	-527.356	-558.660	-0.214
$[dA][Na] + [dA]^- \rightarrow [dA]_2[Na]^-$	-168.038	-206.134	-70.651
$[Na][dA]_2^- + [Na]^+ \rightarrow [dA]_2[Na]_2$	-462.001	-493.206	-25.015
$[dA]_2[Na]_2 + [dA]^- \rightarrow [dA]_3[Na]_2^-$	-183.722	-229.777	-36.919

An investigation into the clustering behaviour of deprotonated adenine revealed the high degree of complexing favourability between deprotonated adenine and sodium cations. As seen in table 25, based on the Gibbs energy, enthalpy and entropy of formation, deprotonated adenine would serve as an effective clustering agent in the gas phase. In addition, the general trend of decreasing energetic favourability in higher-complexity clusters, observed in the clusters of deprotonated pyrimidines, does not appear to be observed in the clusters of dA. It is possible this is due to a structural difference between the pyrimidines and purines; however, it should be a deviation in the trend that is maintained in the clusters of dG.

To further evaluate the effectiveness of these computational evaluations, experimental spectra for each of these clusters must be evaluated against their computationally determined vibrational spectra, however, experimental spectra for deprotonated adenine clusters were not possible to be produced due to travel restrictions from COVID-19.

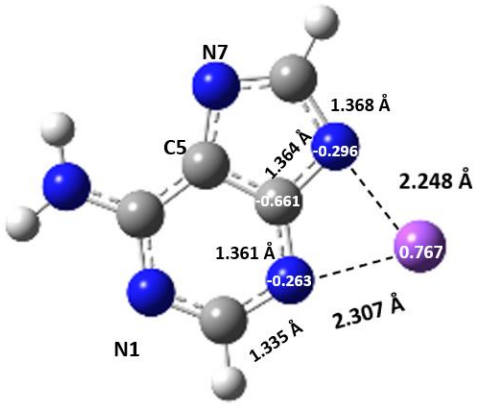


Figure 60: The computationally determined vibrational spectrum for deprotonated adenine with one sodium cation.

The neutral cluster of dA-Na, shown in Figure 60, reveals the most favourable site of interaction between a dA monomer and a sodium cation. It was expected, based on this geometry, that subsequent clusters involving dA and Na^+ will follow this same mechanism.

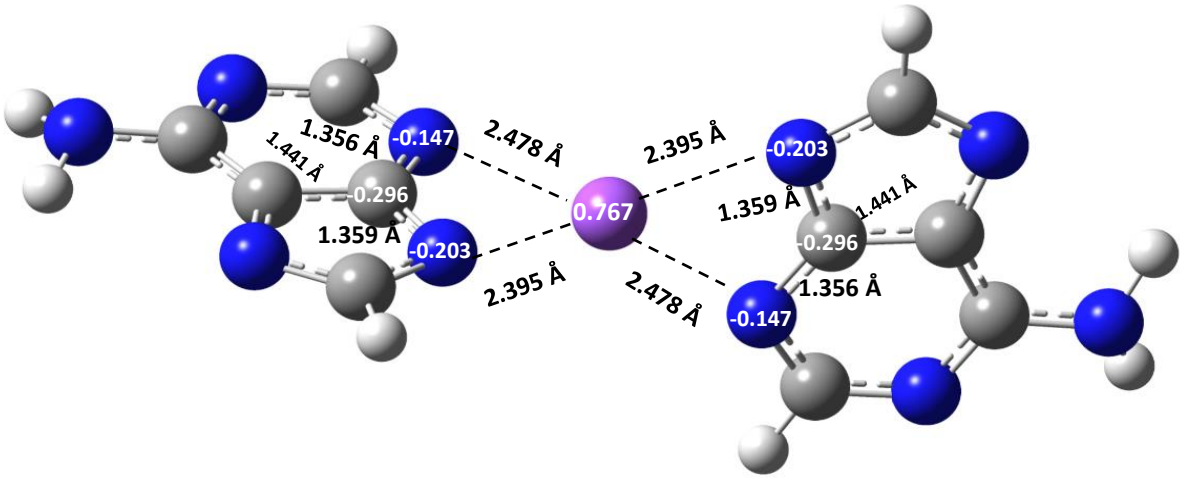


Figure 61: The geometrically optimized structure for the dA_2-Na cluster.

The structure shown in Figure 61 demonstrates the deviation from planarity as the two deprotonated adenines now approach orthogonality. With a dihedral angle of 86.8°, about the planes of each deprotonated adenine, the electronic repulsion is almost completely minimized. In addition, much like the clusters of this complexity with the pyrimidines, this structure is symmetrical about the cation in terms of the bond distances and charges of the atoms directly or indirectly interacting with the sodium cation. This structure, dA2-Na, is one that would be possible to experimentally observe as it has an overall charge of negative one, thus, would be of interest when validating our computational characterization for this species.

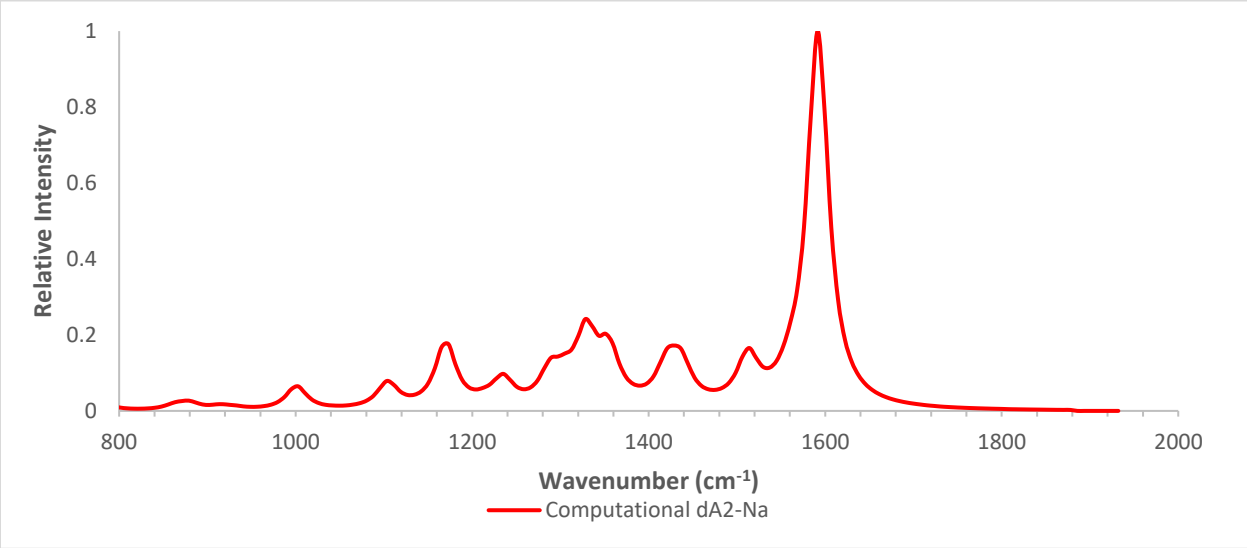


Figure 62: The computationally determined vibrational spectrum for two deprotonated adenines with one sodium cation.

The computational vibrational spectrum in Figure 62 shows the computational peaks of the vibrational modes that we would expect to produce experimentally. In this computed spectrum, we can see a total of 8 major and minor peaks that we expect to experimentally observe in the vibrational modes of the dA_2-Na ion-molecule cluster.

Table 26: The computed dA_2-Na cluster vibrational frequencies.

Computational dA_2-Na Peak Positions (cm^{-1})	Computational dA_2-Na Vibrational Modes
1001	Rocking of NH_2 protons
1104	Rocking of NH_2 and C8 protons
1173	Rocking of C8 proton
1235	Stretch of N1/C2 with symmetric stretch of N7/C8/N9
1328	Stretch of C4/C5 causing rocking of NH_2/C2/C8 protons
1429	Rocking of C2 proton
1513	Stretch of N3/C4
1592	NH_2 scissoring mixed with C1/C6 stretching

Some conservation between the experimental peaks of dA and the dA_2-Na peaks shown in Table 26 has been observed. It is likely due to vibrational modes being unaffected by the new interactions with sodium cations. This conservation of vibrational modes is expected to continue into other clusters of dA, such as dA_2-Na_2 and dA_3-Na_2.

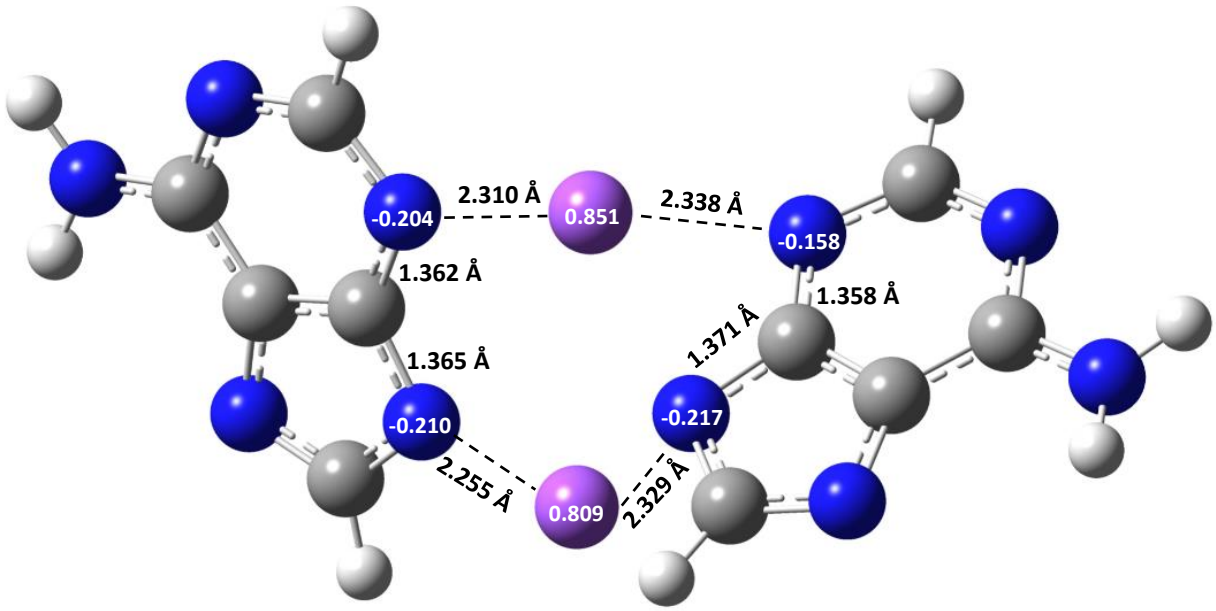


Figure 63: The computationally determined structure for dA_2-Na_2.

Figure 63 highlights the most energetically favourable geometry of the dA_2-Na_2 determined through geometric optimization. It was found that there was only one possible energetically favourable geometry of dA_2-Na_2 and no other energetically comparable "Train" geometry for this cluster. This would suggest that the structure of the purines prevents favourable anion-cation interactions in regions of dA apart from the N3 and N9 region. The dA_2-Na_2 cluster appears to have deviated from symmetrical interactions about the Na^+-Na^+ internuclear axis.

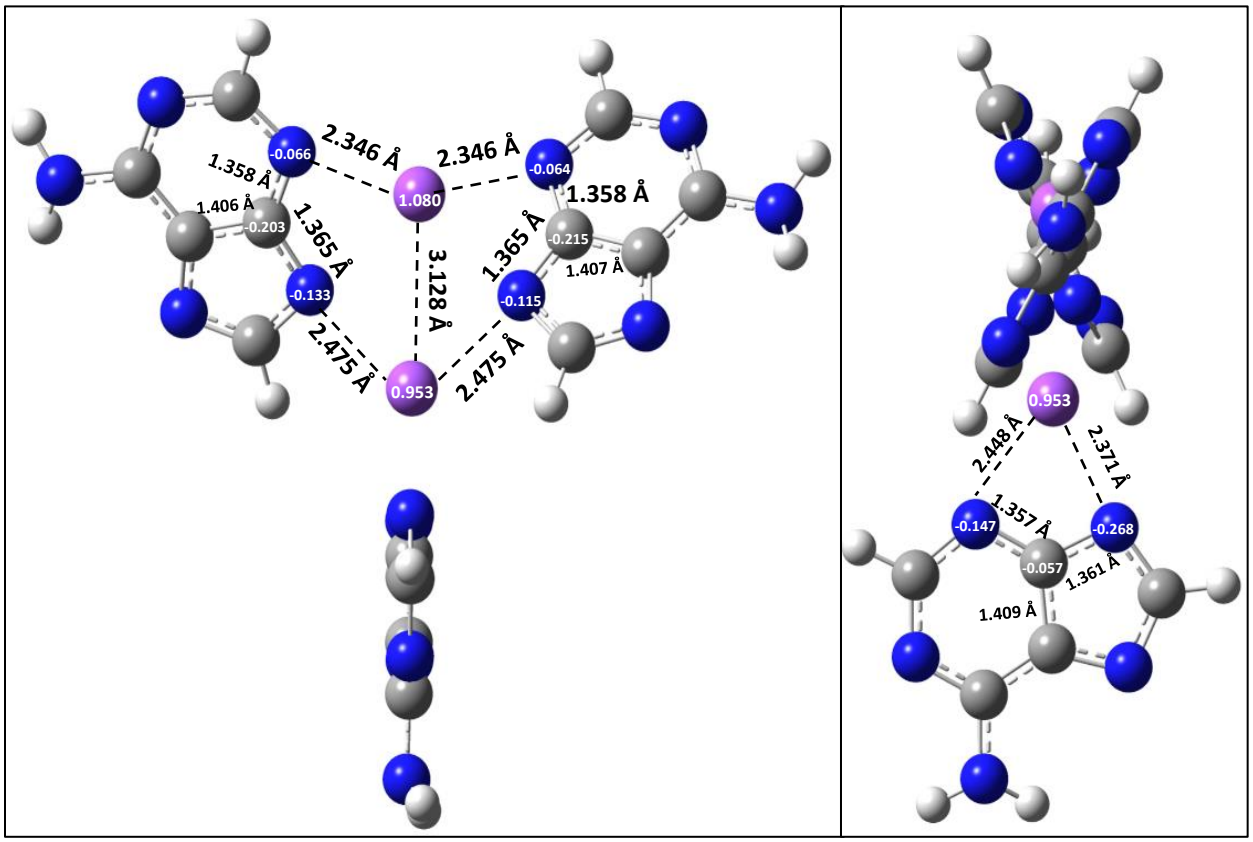


Figure 64: Two different angles for the computationally determined structure for the dA_3-Na_2 cluster.

Figure 64 demonstrates the high degree of symmetry in the dA3-Na2 cluster, with two deprotonated adenines out of the Na^+-Na^+-dA plane. In spite of this high degree of symmetry, the charges and intermolecular interactions between the two seemingly

symmetrical monomers of dA are not equivalent. This deviation from perfect symmetry varies from the clustering behaviour and geometry seen in the pyrimidine clusters. Therefore, it is possible that the purines will instead cluster with these asymmetric interactions, at every level of clustering. It is likely that this high degree of symmetry would be translated to the vibrational modes observed computationally and experimentally.

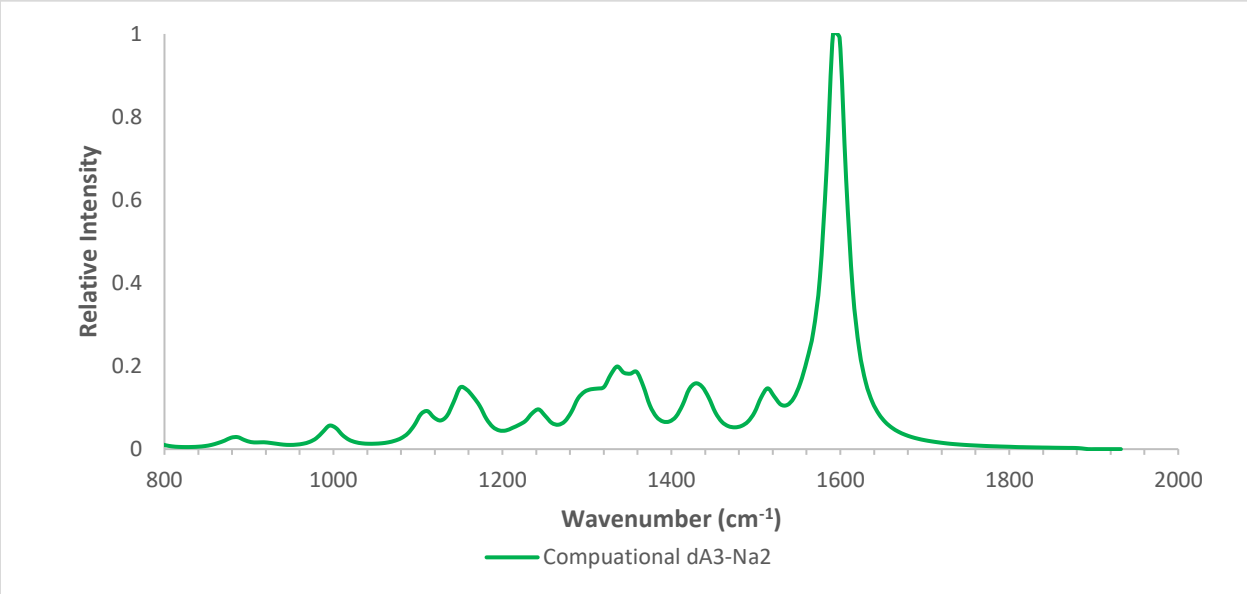


Figure 65: The computationally determined vibrational spectra for three deprotonated adenines, with two sodium cations (dA3-Na2).

In this computational spectrum, from Figure 65, there appear to be several vibrational modes maintained from lower-complexity clusters of dA. The specific peak positions and their corresponding vibrational modes are listed in Table 27 and summarize the vibrational frequencies we expect to experimentally observe.

Table 27: The computed dA_3-Na_2 cluster vibrational frequencies.

Computational Peak Positions (cm^{-1})	Computational dA_3-Na Vibrational Modes
996	Rocking of NH_2 protons
1111	Rocking of NH_2 and C8 protons
1150	Rocking of C8 proton with C8/N9 stretch on 2 equivalent dA's
1242	Stretch of N1/C2 and N7/C8 on 2 equivalent dA's
1335	Stretch of C4/C5 and C6/N1 causing rocking of NH_2/C2/C8 protons on 2 equivalent dA's
1428	Rocking of C8 proton
1513	Stretch of N3/C4 on 2 equivalent dA's
1591	NH_2 scissoring mixed with C5/C6 stretching on 2 equivalent dA's

In short, although experimental spectra for the charged clusters of dA2-Na and dA3-Na2 could not be obtained, there is a wealth of foundational information as to the clustering behaviour of deprotonated adenine with cationic sodium relating to possible vibrational modes, novel geometries, and extremely favourable complexing reactions.

3.3.2.2 Deprotonated Guanine & Clusters

Guanine's molecular structure is similar to that of adenine, seen in Figure 15, with the exception of the position of the amino group as well as the addition of an oxygen group in guanine. As both guanine and adenine are purines, it would stand to reason there should be several similarities in their vibrational modes as well as their clustering behaviour. Moreover, these similarities should extend to the deprotonated version of deprotonated guanine (dG) and deprotonated adenine (dA). The first step to evaluating these similarities between purines is to validate the computational assessment for the most energetically favourable isomer of deprotonated Guanine, as reported in Figure 65.

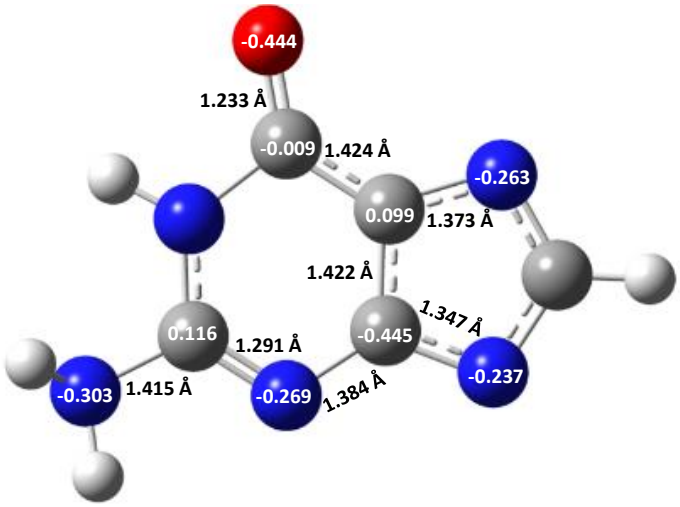


Figure 66: The computationally determined energetically favourable structure of deprotonated guanine (dG), the N9 isomer.

The structure provided by Figure 66 highlights some of the physiochemical properties for the most energetically favourable isomer of deprotonated guanine (dG) and will act as the foundation for analysis of dG clustering behaviour. The dG-N9 isomer, shown in Figure 66, was calculated to be 106.662 kJ/mol more energetically favourable than the dG-N1 and dG-NH_2 isomers, respectively. Based on the chemical charges, there are two regions on dG where a cationic sodium could favourably interact, the N3/N9 region and the O/N7 region. Based on the potential for electrostatic interaction between negatively charged atoms and the positively charged sodium atom, once optimized, and evaluated computationally, we would expect to see the clustering of dG and Na^+ focused on one of these two regions. Once again, there were limitations in terms of experimental spectra we were able to obtain, and unfortunately, the clusters of dG with Na^+ were not among those spectra.

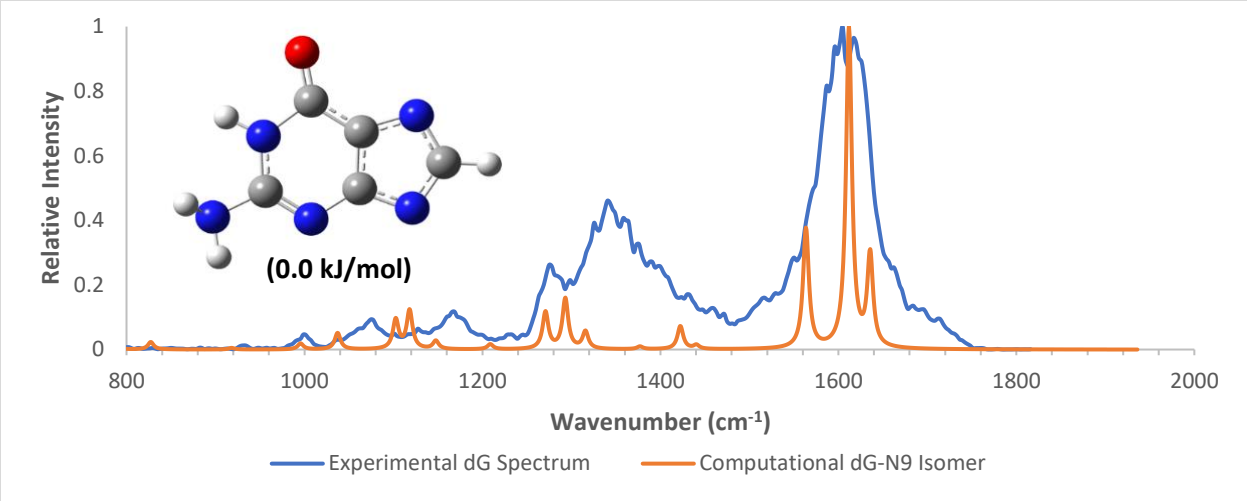


Figure 67: The computationally and experimentally determined vibrational spectra for the N9 isomer of deprotonated guanine (dG-N9).

The spectra shown in Figure 67 highlights the vibrational modes of the experimentally produced dG and the computational spectrum of the dG isomer with an abstracted proton from the N9 site (dG-N9). The experimental spectrum appears to have 6 defining major and minor peaks. Due to the high degree of noise, likely a result of sample contamination, the peaks reported will reflect the value of frequencies at the apex of each group of vibrational modes. The experimental peaks reported in this chapter are therefore, ~999, ~1074, ~1168, ~1274, ~1340 and ~1604 cm^{-1}. The experimental peaks are listed in Table 28 and assigned by the computational peaks of the dG-N9 isomer. The added broadness in the experimental peaks might be attributed to increased laser power or a change in attenuation, compared to the other deprotonated nucleobase monomers, as a result of the laser being operated right after a recalibration at the start of the working day. With time, the laser power is likely to decrease until it stabilizes which may have cause the unusual peak broadening experienced in this experimental spectrum.

Table 28: The computed dG N9 isomer vibrational frequencies and the experimental vibrational frequencies for dG.

Experimental Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
999	**996**	Scissoring of N7/N9 along N7/C8/N9
1074		
1168		
1274	**1270**	Asymmetric stretch of C5/C4/N9
1340		
1604	**1563**	Scissoring of NH_2 protons with slight stretch of C2/N3
	1611	Stretching of C2/N3 with slight stretch of C2/N3
	1635	Stretching of C6/O

The 5 computational peaks, found at ~996, ~1270, ~1563, ~1611 and ~1635 cm^{-1} (Table 28), were in rough agreement with several of the peaks seen in the experimental spectrum. Although not in perfect agreement, the largest and third-largest experimental peaks appeared to be well-characterized by the dG-N9 isomer. There is some degree of overlap between the experimental dG and the dG-N9 isomer spectra, but not enough to be completely certain that the dG-N9 isomer was produced experimentally, without the presence of other isomers of dG. However, the slight agreement of two of the three largest experimental peaks, coupled with the relative Gibbs energy evaluation, would suggest that the dG-N9 isomer is the most dominant form of the dG isomer. Therefore, it would be prudent to assess the other two isomers of dG to ensure our observations.

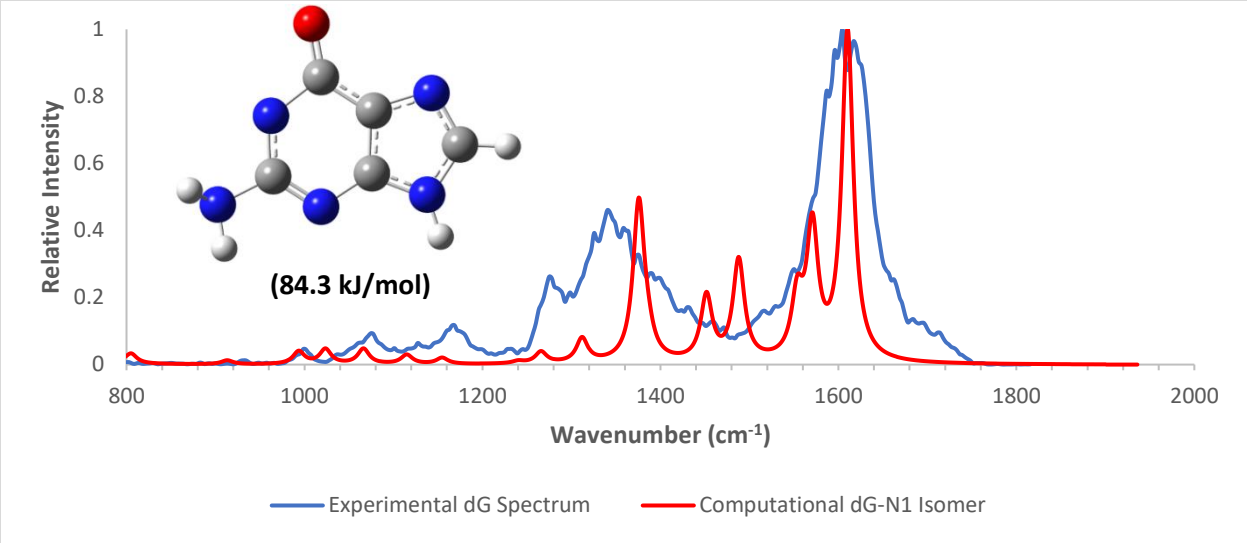


Figure 68: The computationally and experimentally determined vibrational spectra for the N1 isomer of deprotonated guanine (dG-N1).

It is clear from the spectra in Figure 68 that there is some agreement between the vibrational modes of the experimental dG and the computationally determined N1 isomer outside the major peaks around ~1600 cm^{-1} and the minor peaks around 1074 and 1168 cm^{-1}. There are, however, three peaks that do appear to somewhat agree with the experimental spectra and can account for some of the experimental peaks that were not in agreement with the dG-N9 isomer.

Table 29: The computed dG N1 isomer vibrational frequencies and the experimental vibrational frequencies for dG.

Experimental dG Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
999	**994**	Symmetric stretch of N1/C2/N3 bond
1074	**1067**	Rocking of NH_2 proton
1168	**1155**	Rocking of NH_2 protons with stretching of N1/C6 bond
1274	**1267**	Rocking of C8 proton
1340		
1604	**1571**	Stretching of N1/C2 and N3/C4 bonds
	1611	Stretching of C6/O

Table 29 highlights some of the agreement between the vibrational spectra of dG-N1 isomer and the experimental spectrum of dG. There appears to be slightly more agreement between the dG-N1 isomer and the experimental spectrum for dG, suggesting there is some dG-N1 present. However, given the energetic unfavourability of dG-N1 it is much more likely that we experimentally observed some dG-N1 in mixture with a larger proportion of dG-N9.

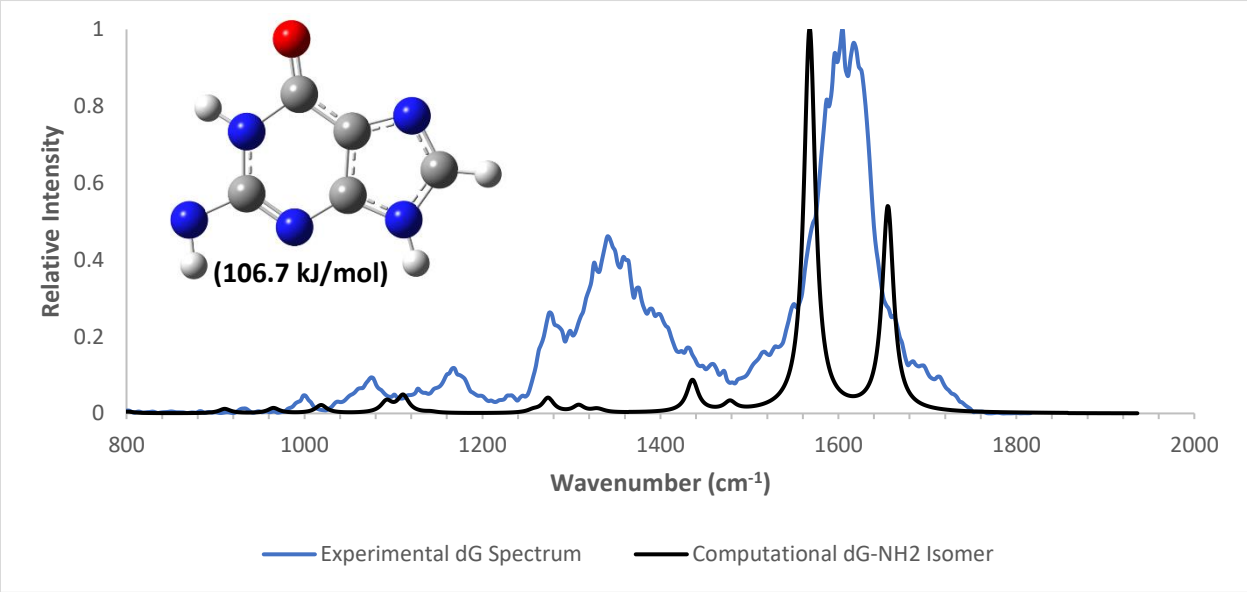


Figure 69: The computationally and experimentally determined vibrational spectra for the NH2 isomer of deprotonated guanine (dG-NH2).

Evidently, there is little agreement whatsoever between the two spectra presented in Figure 69, and vibrational frequencies in Table 30. Given the extremely large relative energetic unfavourability between this isomer and the dG-N9 isomer, it is extremely unlikely that there is enough of the dG-NH_2 isomer to assume it was produced experimentally.

Table 30: The computed dG-NH_2 isomer vibrational frequencies and the experimental vibrational frequencies for dG.

Experimental dG Peak Positions (cm^{-1})	Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
999		
1074	**1092**	Rocking of NH proton
1168		
1274	**1273**	Rocking of NH and C8 proton
1340		
1604		

Based on the sufficient agreement between the computational and experimental spectra, as well as the computational evaluation of the relative Gibbs energies, it is very likely that the dominant dG isomer is dG-N9. However, it is important to note, as stated before, that there is some evidence to suggest that the experiment produced a mixture of the N9 and N1 isomers. Given this computational evidence, the dG-N9 isomer was used as the foundation for subsequent clusters with Na^+.

Table 31: Thermochemical properties of the dG-sodium cation clusters.

Reaction	ΔG° (kJ/mol)	ΔH° (kJ/mol)	ΔS° ($J \cdot K^{-1} \cdot mol^{-1}$)
$[dG]^- + [Na]^+ \rightarrow [dG][Na]$	-510.603	-542.494	-25.565
$[dG][Na] + [dG]^- \rightarrow [dG]_2[Na]^-$	-184.622	-222.783	-30.591
$[Na][dG]_2^- + [Na]^+ \rightarrow [dG]_2[Na]_2$	-425.254	-462.367	-29.751
$[dG]_2[Na]_2 + [dG]^- \rightarrow [dG]_3[Na]_2^-$	-172.909	-233.906	-48.897

The data shown in Table 31, is a summary of the thermodynamic properties of formation for each of the clusters produced from different amounts of dG and Na^+. This table demonstrates how each cluster favourably complexes with the addition of a new cation or dG, which speaks to the overall favourable clustering potential of dG. In addition, as

shown among some of the other nucleobases, there is a slight decrease in energetic favourability as the complexity of the cluster increases. These thermochemical properties all suggest that dG is indeed an extremely favourable complexing agent in the gas phase.

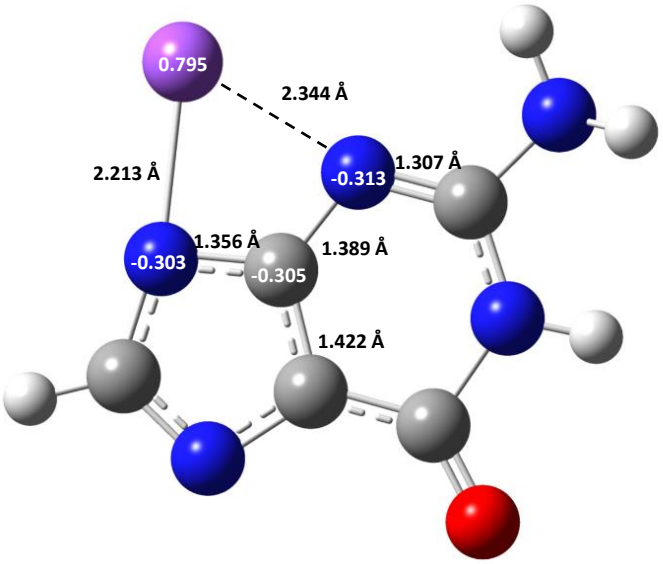


Figure 70: The computationally determined geometry and physiochemical properties for the dG-Na cluster.

The cluster depicted in Figure 70, demonstrates the most favourable site of interaction between the dG-N9 monomer and the sodium cation. Similar to the interaction site shown in dA, based on this geometric optimization, the higher-complexity clusters were predicted to form about Na^+ and these N3/N9 regions. Evidence of this clustering behaviour, consistent with the other purine, can be seen in Figure 70.

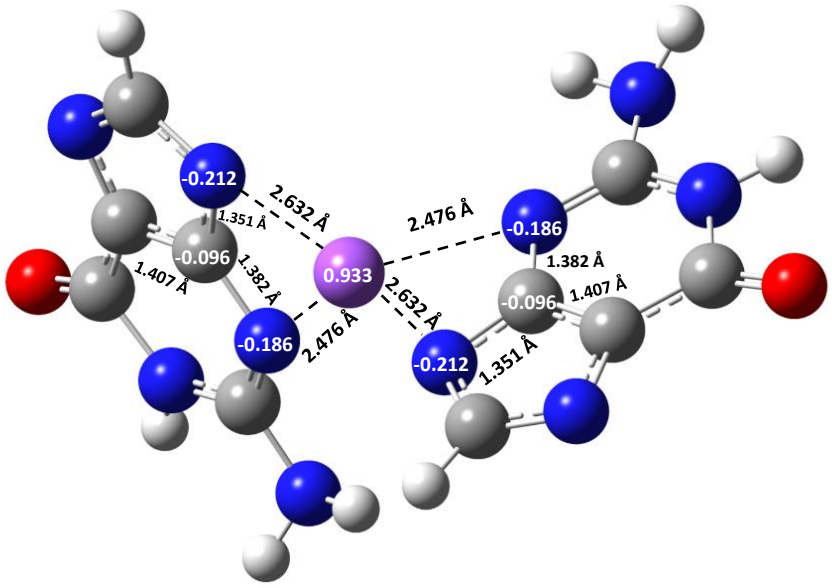


Figure 71: The computationally determined energetically favourable structure for the dG_2-Na.

With the addition of a second dG, in Figure 71, the clustering species had increased in its symmetry creating a rotational symmetry about the sodium cation that produced the equivalent charges and bond distances observed. With an optimized structure exhibiting a high degree of symmetry, as seen in the experimental spectra for the dX_2-Na clusters of dT and dU, it is expected that there will be experimental peaks characterized by either a single intense peak or two nearly identical smaller peaks. Each of these smaller peaks would be expected to describe equivalent vibrational modes, but only describe one of the two monomers at a time. The computationally determined vibrational spectrum for the dG_2-Na cluster that we would have expected to see good agreement with the experimental spectrum, is found in Figure 72. The computationally determined peak assignments can then be found in Table 31 below.

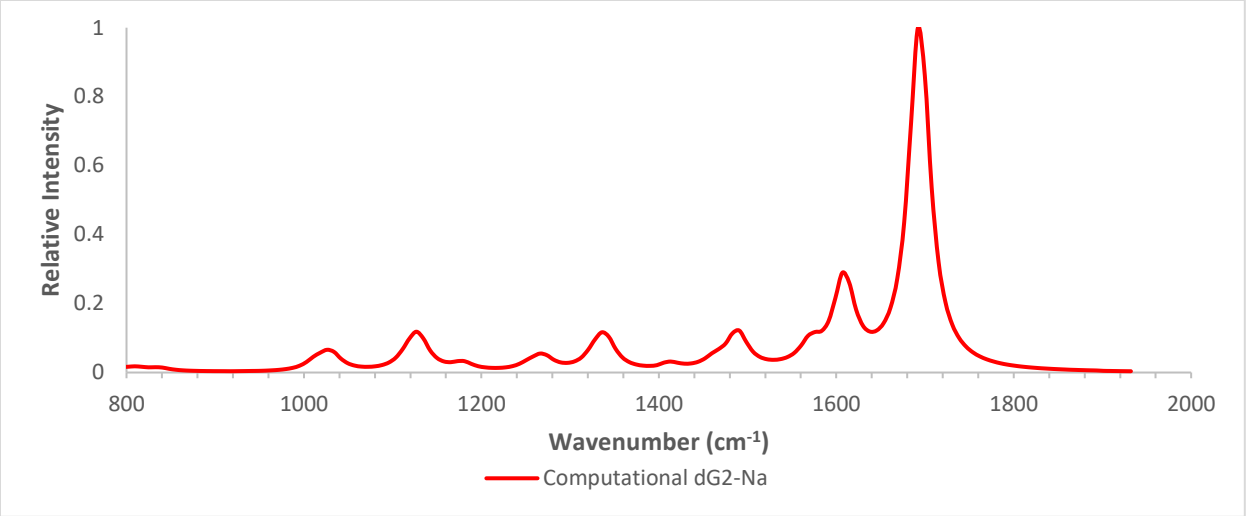


Figure 72: The computationally determined energetically favourable structure of two deprotonated guanines and a sodium cation (dG2-Na).

The vibrational spectrum shown in Figure 72 contains 7 peaks that computationally define this chemical system which are reflected in Table 32. The carbonyl stretch observed past the 1600 cm^{-1} appeared to be characteristic in all oxygen-containing nucleobases, each with relatively high-intensity peaks for that stretch.

Table 32: The computed vibrational frequencies for dG_2-Na cluster.

Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
1029	Rocking of the NH_2 protons
1129	Rocking of C8 protons
1269	Rocking of N1 proton
1335	Stretching of C4/N9 bond
1490	Symmetric stretch of N1/C2 and C4/C5 bonds
1607	Scissoring of NH_2 bonds
1692	Stretching of C6/O bond

It is not surprising to see this level of conservation between the dG2-Na and dG clusters as many of the vibrational modes, that are not interacting with the sodium cation, should not have been affected by the addition of another dG that would interact with the Na^+ in nearly the same way.

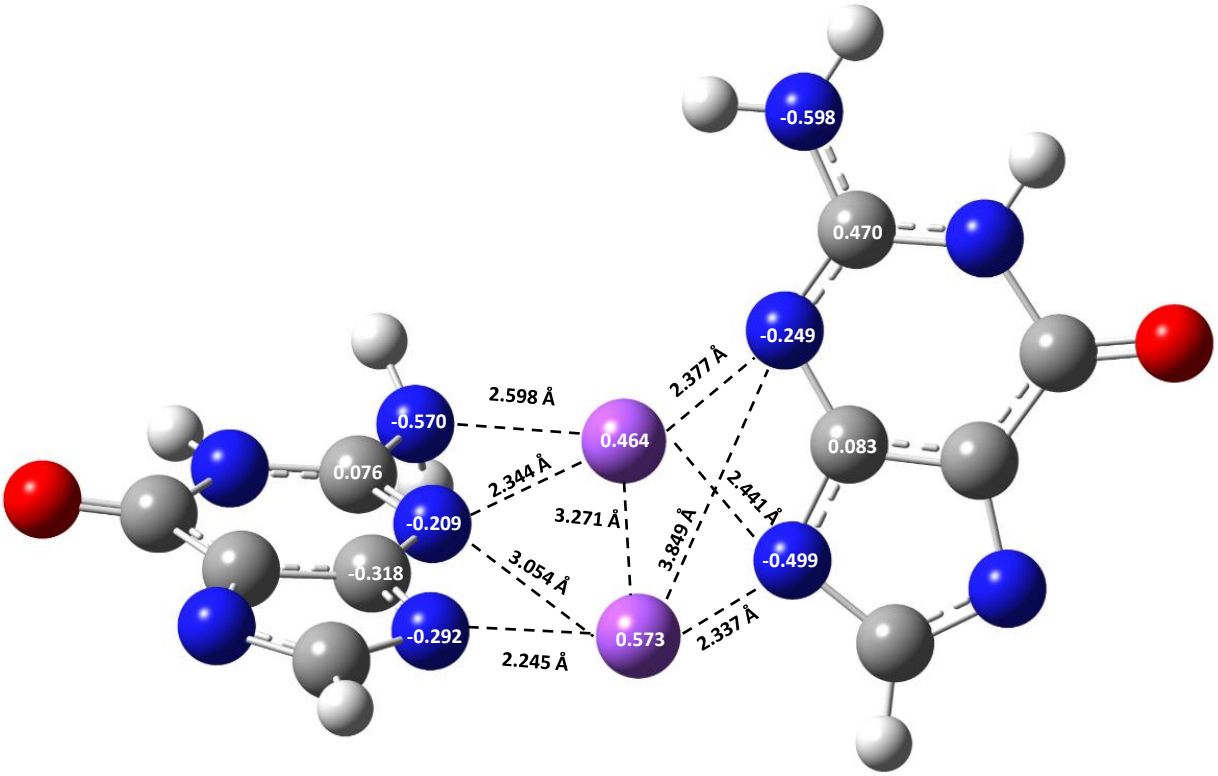


Figure 73: The computationally determined energetically favourable structure for the dG_2-Na_2.

After several attempts at optimizing the cluster, the structure shown in Figure 73 was optimized to a geometry that deviates from symmetry, similar to the dA_2-Na_2 structure previously investigated. It was extremely interesting to observe how one dG monomer appeared to exhibit interaction between three nitrogen atoms, including the NH_2 group, while the other monomer only exhibits interaction between the N3 and N9 nitrogen atoms. Given that the other nucleobases optimized towards a dX_2-Na_2 structure that approached symmetrical interactions more closely than this structure, including a fellow purine, it is quite possible there is still an orientation of this structure that is yet to be observed. Therefore, it might be necessary in the future to evaluate many more candidate structures at a lower level of theory to approximate a reasonable amount of accuracy in geometric optimization before applying a higher level of theory for evaluation.

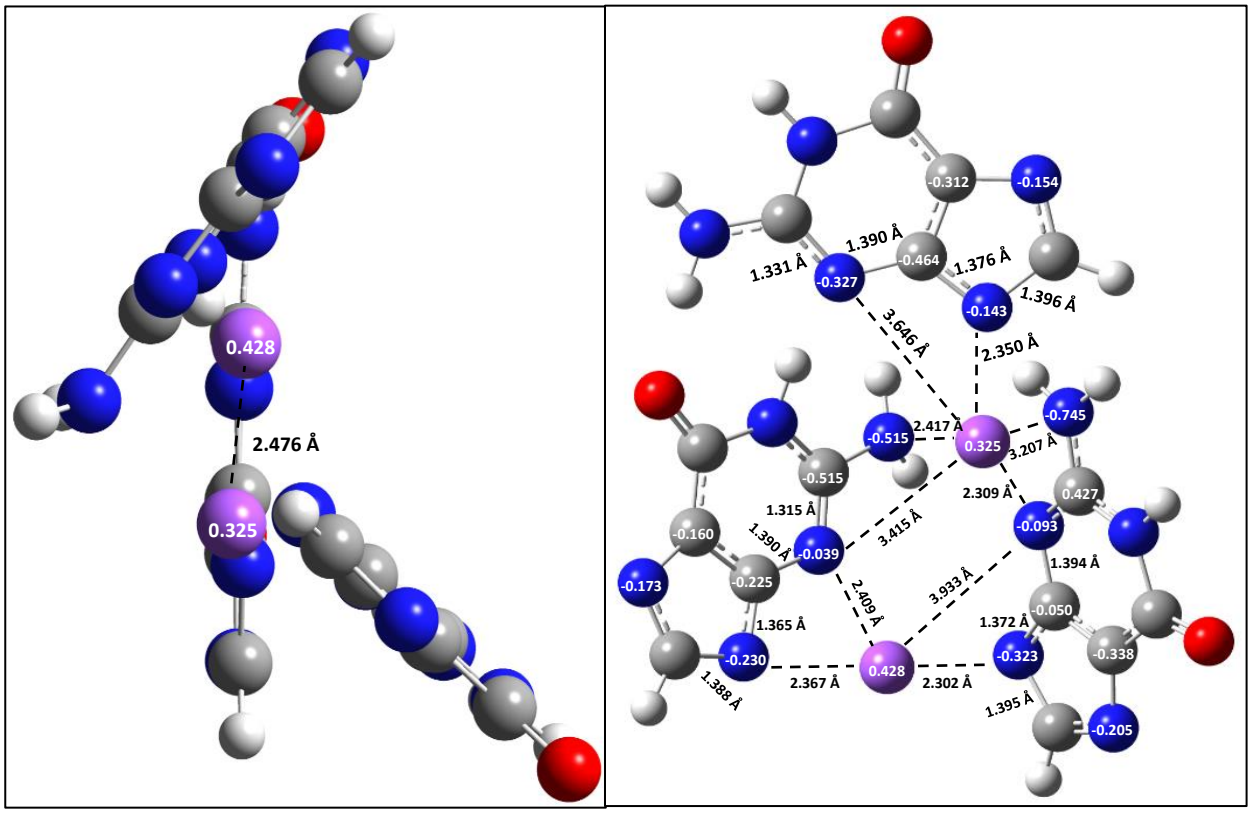


Figure 74: The computationally determined energetically favourable structure of two deprotonated guanines and a sodium cation (dG_3-Na_2).

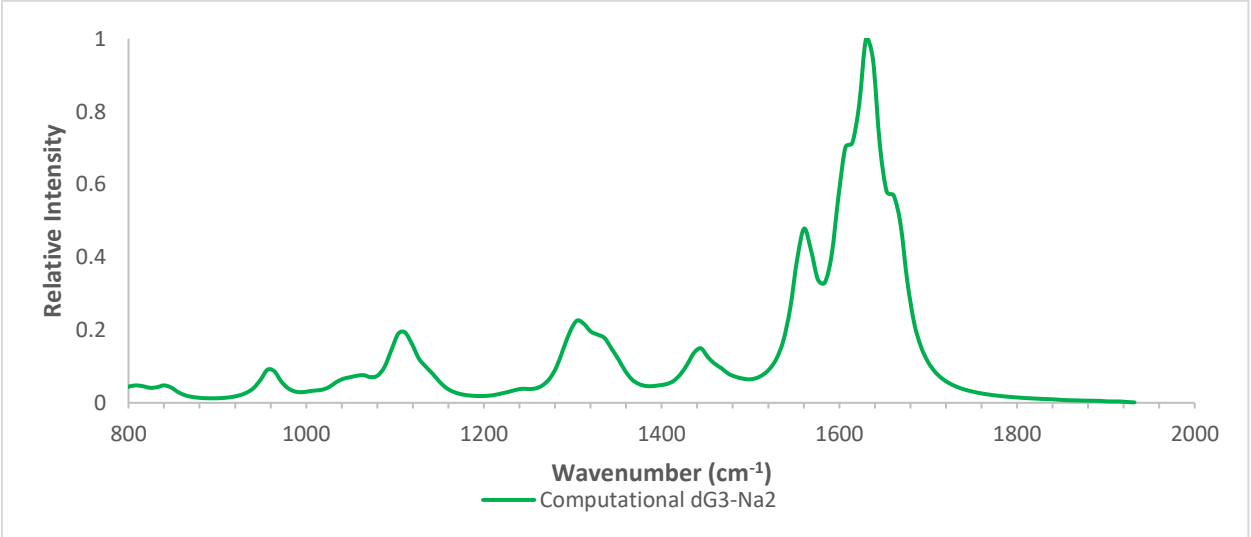


Figure 75: The computationally determined vibrational spectrum for two deprotonated guanines and a sodium cation (dG_3-Na_2).

Figures 74 and 75 provide context for the geometric structure and vibrational frequencies we hope to observe experimentally observe for the dG_3-Na_2 cluster. Interestingly, the deviation from symmetrical interactions about the Na^+-Na^+ internuclear axis was exhibited in the structure demonstrated in Figure 73. Thus, there was a large difference in computationally determined geometries between the two purines and might account for some of the discrepancies between the vibrational modes assigned to peaks that were expected to be conserved between purines sharing similar structures. The computationally determined vibrational frequencies, and their respective modes, are listed in Table 33 below.

Table 33: The major and minor peaks of the computed vibrational frequencies for dG_3-Na_2 cluster.

Computational Peak Positions (cm^{-1})	Computational Vibrational Modes
959	Rocking of the NH_2 protons
1102	Stretching of all C8/N9 bonds
1306	Stretching of C4/N9 and C5/N7 bonds on dG with charge of -0.225 Debye on C4
1335	Stretching of C5/C6 bonds
1443	Asymmetric stretch of N3/C4 and C4/C9 bonds
1560	Asymmetric stretching of C2/N3 and C2/NH_2 on all dG's
1630	Stretching of C6/O bonds on all dG's

In summary, the purines exhibited similar clustering behaviour up until the dX_2-Na_2 clusters. It is possible this is the result of the chemical differences between adenine and guanine, where guanine contains a carbonyl group, however, it was not expected that these two nucleobases would differ so greatly between one another at the higher-level clusters.

3.3.4 Gas Phase Acidities

One of the two main questions this investigation attempted to answer was whether: "the pK_as of the nucleobases reliably predict the acid strength of the nucleobases in the gas phase". Thus, the GPAs were computationally determined for the most energetically favourable isomer of each nucleobase, which were experimentally and computationally validated through analyses of the vibrational spectra, and revealed to exhibit an acid strength comparable to other well-documented organic acids, seen in Table 34 below.

Table 34: The gas phase acidities of the nucleobases. [11-16, 62-65]

Nucleobase	pK_a	Reported Gas Phase Acidity (kJ mol^{-1})	Calculated Gas Phase Acidity (kJ mol^{-1})
Adenine	9.8	1401	1362
Cytosine	12.2	1423	1403
Guanine	10.0	1389	1360
Thymine	9.9	1393	1370
Uracil	9.5	1393	1380
Ubiquitous Molecules	pK_a	Reported Gas Phase Acidity (kJ mol^{-1})	Calculated Gas Phase Acidity (kJ mol^{-1})
H_2O	14.0	1595	1611
Halogen Acid	pK_a	Reported Gas Phase Acidity (kJ mol^{-1})	Calculated Gas Phase Acidity (kJ mol^{-1})
HF	3.1	1522	1522
HCl	- 6.0	1372	1385
HI	- 9.5	1291	1293
Organic Acid	pK_a	Reported Gas Phase Acidity (kJ mol^{-1})	Calculated Gas Phase Acidity (kJ mol^{-1})
CH_3CO_2H	4.8	1459	-

The investigation was quite revealing in that it strongly suggested that pK_a values are not reliable predictors of the acid strength of the nucleobases, in the gas phase, as the order of acid strength in the solution phase did not parallel the order of gas-phase acid

strengths. It was previously reported that the order of increasing acid strength for the nucleobases was as follows: guanine, thymine, uracil, adenine and cytosine. It was interesting how this investigation similarly described the weakest acid as cytosine and the strongest as guanine but differed in the order of remaining nucleobases. Through this investigation, there is strong evidence to support the observation of the purines being stronger gas phase acids than the pyrimidines. It is likely this is due to the increased potential for charge delocalization in a two-ringed molecule compared to a one-ringed molecule, resulting in the purines being weaker bases than the pyrimidines. Based on fundamental chemical principles, the order of increasing gas phase acidity described in this investigation makes sense and further supports the validity of this investigation to accurately assess the GPA of the nucleobases.

3.3.5 Sodium-Nucleobase Clusters

The second question this investigation attempted to answer was to see if "the deprotonated nucleobases could favourably interact and cluster with sodium cations in the gas phase". This question aimed to provide context for possible novel clustering geometries and interactions that would hopefully provide foundational information for potential future medically focused applications. Table 35 summarizes the energetic favourability for the formation of the clusters from iterative additions of sodium cations and deprotonated nucleobases.

Table 35: The gas phase Gibbs energy of reaction for each of the nucleobases.

Product of Formation	Adenine ΔG° (kJ/mol)	Cytosine ΔG° (kJ/mol)	Guanine ΔG° (kJ/mol)	Thymine ΔG° (kJ/mol)	Uracil ΔG° (kJ/mol)
$[dX]^- + [Na]^+ \rightarrow [dX][Na]$	-527.356	-576.637	-510.603	-540.645	-530.179
$[dX][Na] + [dX]^- \rightarrow [dX]_2[Na]^-$	-168.038	-181.539	-184.622	-180.221	-189.910
$[dX]_2[Na]^- + [Na]^+ \rightarrow [dX]_2[Na]_2$	-462.001	-563.878	-425.254	-416.773	-368.194
$[dX]_2[Na]_2 + [dX]^- \rightarrow [dX]_3[Na]_2^-$	-183.722	-154.293	-172.909	-190.095	-194.199

Each of the nucleobases appears to favourably cluster with cationic sodium atoms to form seemingly stable clusters at various levels of complexity. Although there is no apparent trend in the magnitude of energetic favourability across the nucleobases, this thesis accomplished its goal to demonstrate the favourable clustering potential of the deprotonated nucleobases computationally and experimentally.

3.4 Conclusions

This investigation revealed novel clustering geometries, such as the Propeller and Train geometries, that appeared to be exhibited by all the nucleobases, to some degree. In particular, the computational and experimental analysis of the clusters for dT and dU revealed the high degree of symmetry in many of their clusters that set a precedent for the clustering behaviour of the remaining nucleobases. Furthermore, it is likely that given enough time and resources, the experimental vibrational spectra for the nucleobases, and their other clusters with sodium cations that have not been observed, will eventually be produced, and corroborate with the computational data produced in this research. It was revealed that the clusters of dU have the greatest clustering potential among all the

nucleobases, for the clusters that can be experimentally observed (i.e. the anionic clusters). The clusters of each nucleobase were demonstrated to favour the Propeller geometry over the Train geometry and is seen across all the clustering geometries for each nucleobase.

In addition, this investigation determined the acid strengths of the nucleobases and revealed them to be stronger gas phase acids than H_2O, HCl and acetic acid. Moreover, the order of nucleobase gas phase acid strengths produced in this investigation did not parallel the previously reported order of acids strengths for the nucleobases. This deviation from literature, however, follows more closely with fundamental chemical principles and revealed that the acid strength of the purines was greater than the pyrimidines.

The results in this investigation support the potential for deprotonated nucleobases to act as effective charged species carriers, in the gas phase. By characterizing the gas phase characteristics and acid strengths of the deprotonated nucleobases, this work will provide thermodynamic research that may be used to better focus efforts to determine the effectiveness of deprotonated nucleobases as charged species carriers in solution. Future work will involve traveling to FEL facilities which can experimentally produce vibrational spectra for the untested complex nucleobase clusters. It is expected that these vibrational spectra will agree well with our computed vibrational spectra and confirm our initial thoughts on the efficacy and potential of deprotonated nucleobases as charged species carriers.

Bibliography

1. Krishnamurthy, R. (2012). Role of pK_a of Nucleobases in the Origins of Chemical Evolutions. *Accounts of Chemical Research*. **45:** 2035-2044

2. Miller, C. R.; Williams, C. R.; Buchsbaum, D. J.; and Gillespie, G. Y. (2002). Intratumoral 5-fluorouracil produced by cytosine deaminase/5-fluorocytosine gene therapy is effective for experimental human glioblastomas. *Cancer Research*. **62:** 773-780.

3. Finehout, E. J.; and Lee, K. H. (2003). An Introduction to Mass Spectrometry Applications in Biological Research. *Biochemistry and Molecular Biology Education*. **32 (2)**: 93-100.

4. Hofstadler, S. A.; Sannes-Lowry, K. A.; and Harris, J. C. (2003). Analysis of Nucleic Acids by FTICR MS. *Mass Spectrometry Reviews*. **24:** 265-285.

5. Polfer, N. C. (2010). Infrared multiple photon dissociation spectroscopy of trapped ions. *The Royal Society of Chemistry*. **40:** 2211-2221.

6. Marshall, A. G.; Hendrickson, C. L.; and Jackson, G. S. (1998). Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrometry Reviews*. **17**: 1-35.

7. Roca-Sanjuan, D.; Rubio, M.; Merchan, M.; and Serrano-Andres, L. (2006). Ab initio determination of the ionization potentials of DNA and RNA nucleobases. *The Journal of Chemical Physics*. **125**. 084302.

8. Wu, R.; and McMahon, T. B. (2008). Structures Energetics and Dynamics of Gas-Phase Ions Studied by FTICR and HPMS. *Mass Spectrometry Reviews*. **28:** 546-585

9. Giorgi, G.; Piccionello, A. P.; Pace, A.; and Buscemi, S. (2008). Effect of protonation and deprotonation on the gas-phase reactivity of fluorinated 1,2,4-Triazines. *Journal of the American Society for Mass Spectrometry*. **19:** 686-694.

10. Oomens, J.; Sartakov, B. G.; Meijer, G.; and von Helden, G. (2006). Gas-phase infrared multiple photon dissociation spectroscopy of mass-selected molecular ions. *International Journal of Mass Spectrometry*. **254:** 1-19.

11. Topol, I. A.; Tawa, G. J.; Caldwell, R. A.; Eissenstat, M. A.; and Burt, S. K. (2000). Acidity of organic molecules in the gas phase and in aqueous solvent. *Journal of Physical Chemistry A*. **104:** 9619-9624.

12. Huang, Y.; and Kenttamaa, H. (2003). Theoretical Estimations of the 298 K gas-phase acidities of the pyrimidine-based nucleobases uracil, thymine and cytosine. *Journal of Chemical Physics A*. **107:** 4893-4897.

13. Huang, Y.; and Kenttamaa, H. (2004). Theoretical Estimations of the 298 K gas-phase acidities of the pyrimidine-based nucleobases uracil, thymine and cytosine. *Journal of Chemical Physics A*. **108:** 4485-4490.

14. Kumari, S.; Devi, C. L.; Prabhakar, S.; Bhanuprakash, K.; and Vairamani, M. (2009). Estimation of gas-phase acidities of deoxyribonucleosides: an experimental and theoretical study. *Journal of the American Society for Mass Spectrometry*. **21:** 136-143.

15. Kurinovich, M. A.; and Lee, J. K. (2002). The acidity of Uracil and uracil analogs in the gas phase: four surprisingly acidic sites and biological implications. *Journal of the American Society for Mass Spectrometry*. **13:** 985-995.

16. Lee, J. K. (2004). Insights into nucleic acid reactivity through gas-phase experimental and computational studies. *International Journal for Mass Spectrometry*. **240:** 261-272.

17. Wu, R.; and McMahon, T. B. (2006). Investigation of proton transport tautomerism in Clusters of Protonated nucleic acids bases (cytosine, uracil, thymine, and adenine) and ammonia by high-pressure mass spectrometry and ab initio calculations. *Journal of the American Chemical Society*. **129:** 569-580.

18. Wu, R.; and McMahon, T. B. (2007). Infrared multiple photon dissociation spectroscopty as structural confirmation for $GlyGlyGlyH^+$ and $AlaAlaAlaH^+$ in the gas phase. Evidence for amide oxygen as the protonation site. *Journal of the American Chemical Society*. **129:** 11312-11313.

19. Wu, R.; and McMahon, T. B. (2007). Infrared multiple photon dissociation spectra of proline and glycine proton-bound homodimers. Evidence for zwitterionic structure. *Journal of the American Chemical Society*. **129:** 4864-4865.

20. Wu, R.; and McMahon, T. B. (2008). Infrared multiple photon dissociation mechanisms of peptides of glycine. *Chemistry A European Journal*. **14:** 7765-7770.

21. Wu, R.; and McMahon, T. B. (2008). IRMPD spectra of $Gly\text{-}NH_4^+$ and proton-bound betaine dimer: evidence for the smallest gas phase zwitterionic structures. *Journal of Mass Spectrometry*. **43:** 1641-1648.

22. Fridgen, T. D.; MacAleese, L.; McMahon, T.B.; Lemaire, J.; and Matire, P. (2006). Gas phase infrared multiple-photon dissociation spectra of methanol, ethanol and propanol proton-bound dimers, protonated propanol and the propanol/water proton-bound dimer. *Physical Chemistry Chemical Physics*. **8:** 955-966.

23. Henderson, J. P.; Byun, J.; and Heinecke, J. W. (1999). Chlorination of nucleobases, RNA and DNA by myeloperoxidase: a pathway for cytotoxicity and mutagenesis by activated phagocytes. *Redox Report: Communications in Free Radical Research*. **4 (No. 6)**

24. Rochut, S.; Pepe, C.; Paumard, J. P.; and Tabet, J. C. (2004). A computational and experimental study of cation affinity (Na^+) of nucleobases and modified nucleobases by electrospray ionization ion trap mass spectrometry. *Rapid Communications in Mass Spectrometry*. **18:** 1686-1692.

25. Cramp, R. L.; Hudson, N. J.; and Franklin, C. E. (2010). Activity, abundance, distribution and expression of Na^+/K^+-ATPase in the salt glands of *Crocodylus porosus* following chronic saltwater acclimation. *The Journal of Experimental Biology*. **213:** 1301-1308.

26. Eicher, T.; Hauptmann, S.; and Speicher, A. (2012). Online Access of Text. The Chemistry of Heterocycles: Structures, Reactions, Synthesis, and Applications 3rd Edition. *Wiley-VCH*.

27. Nagaradja, E.; Bentabed-Ababsa, G.; Scalabrini, M.; Chevallier, F.; Philippot, S.; Fontanay, S.; Duval, R. E.; Halauko, Y. S.; Ivashkevich, O. A.; Matulis, V. E.; Roisnel, T.; and Mongin, F. (2015). Deprotometalation-iodolysis and computed CH acidity of 1,2,3- and 1,2,4-triazoles. Application to the synthesis of resveratrol analogues. *Bioorganic and Medicinal Chemistry*. **23:** 6355-6363.

28. Kharb, R.; Chander Sharma, P.; and Shahar Yar, M. (2011). Pharmacological significance of triazole scaffold. *Journal of Enzyme Inhibition and Medicinal Chemistry*. **26 (1):** 1-21.

29. Feyereisen, M.; Fitzgerald, G.; and Komornicki, A. (1993). Use of approximate integrals in ab initio theory: an application in MP2 energy calculations. *Chemical Physics Letters*. **208 (No. 5-6).**

30. Dixit, A.; Claudot, J.; Lebegue, S.; and Rocca, D. (2017). Communication: A novel implementation to compute MP2 correlation energies without basis set superposition errors and complete basis set extrapolation. *The Journal of Chemical Physics*. **146.** 211102.

31. Marta, R. A. (2009). PhD Thesis: Mass-selected infrared multiple-photon dissociation as a structural probe of gaseous ion-molecule complexes. *University of Waterloo*.

32. Frisch, M. J.; Head-Gordon, M.; and Pople, J. A. (1989). A direct MP2 gradient method. *Chemical Physics Letters*. **166 (No. 3):** 275-280.

33. Head-Gordon, M.; Pople, J. A.; and Frisch, M. J. (1988). MP2 energy evaluation by direct methods. *Chemical Physics Letters*. **153 (No. 6):** 503-506

34. Paradalos, P. M.; Shalloway, D.; Xue, G. (1994). Optimization methods for computing global minima of nonconvex potential energy functions. *Journal of Global Optimization*. **4:** 117-133.

35. Zhao, X.; and Slanina, Z. (2003). C_{98} IPR isomers: Gibbs-energy based relative stabilities. *Journal of Molecular Structure (Theochem)*. **636:** 195-201.

36. Slanina, Z.; Uhlik, F.; Shen, W.; Akasaka, T.; Lu, X.; and Adamoxicz, L. (2019). Calculations of the relative populations of Lu@C_{82} isomers. *Fullerenes, Nanotubes and Carbon Nanostructures*. **27 (No. 9):** 710-714.

37. Büchel, R. C.; Rudolph, D. A.; and Frank, I. (2020). Deterministic quantum mechanics: the role of the Maxwell-Boltzmann distribution. *The International Journal of Quantum Chemistry*. **121.**

38. Rowlinson, J. S. (2005). The Maxwell-Boltzmann distribution. *Molecular Physics*. **103 (No. 21-23):** 2821-2828.

39. Habould, H. J.; and Kumar, D. (2007). Extension of thermonuclear functions through the pathway model including Maxwell-Boltzmann and Tsallis distributions. *Astroparticle Physics*. **29:** 70-76.

40. Bruins, A. P. (1998). Mechanistic aspects of electrospray ionization. *Journal of Chromatography A*. **794:** 345-357.

41. Yamada, T.; and Aida, M. (2014). Fundamental frequency from classical molecular dynamics. *The Royal Society of Chemistry*. **17:** 3227-3240.

42. Simons, J. (2003). An introduction to theoretical chemistry: Chapter 6. Online Access. *Cambridge University Press*. 184-203.

43. Ho, J.; and Ertem M. Z. (2016). Calculating free energy changes in continuum solvation models. *Journal of Physical Chemistry B*. BNL-112093-2016-JA.

44. Schaefer, H. F. (1977). Text. Modern Theoretical Chemistry 3: Methods of Electronic Structure Theory. *Springer Science+Business Media New York*.

45. Feller, D.; Dixon, D. A.; and Nicholas, J. B. (2000). Binding Enthalpies for Alkali Cation-Benzene Complexes Revisited. *Journal of Physical Chemistry A*. **104:** 11414-11419.

46. Banerjee, S.; and Mazumdar, S. (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International Journal of Analytical Chemistry*. **2012 (ID: 282574).**

47. Wilm, M. S.; Mann, M. (1994). Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes*. **136:** 167-180.

48. Tobolkina, E. (2014). PhD Thesis: New analytical tools combining gel electrophoresis and mass spectrometry. *Ecole Polytechnique Federale de Lausanne*. **No. 6331.**

49. Taflin, D. C.; Ward, T. L.; and Davis, E. J. (1988). Electrified droplet fission and the Rayleigh Limit. *Journal for the American Chemical Society*. **5 (No. 2):** 378-383.

50. Kleisath, E.; Marta, R. A.; Martens, S.; Martens, J.; and McMahon, T. B. (2015). Structures and Energetics of Protonated Clusters of Methylamine with Phenylalanine Analogs, Characterized by Infrared Multiple Photon Dissociation Spectroscopy and Electronic Structure Calculations. *The Journal of Physical Chemistry A*. **119:** 6689-6702.

51. Callegari, A. (1998). PhD Thesis: Intramolecular Vibrational Energy Redistribution in Aromatic Molecules. *Princeton University*.

52. Stannard, P. R.; and Gelbart, W. M. (1981). Intramolecular Vibrational Energy Redistribution. *The Journal of Physical Chemistry*. **85:** 3592-3599.

53. Wiberg, K. B. (2004). Basis Set Effects on Calculated Geometries: 6-311++G** vs. aug-cc-pVDZ. *Journal of Computational Chemistry*. **25:** 1342-1346.

54. Chen, E. C. M.; Herder, C.; and Chen, E. S. (2006). The experimental and theoretical gas phase acidities of adenine, guanine, cytosine, uracil, thymine and halouracils. *The Journal of Molecular Structure*. **798:** 126-133.

55. Martens, S. M.; Marta, R. A.; Martens, J. K.; and McMahon, T. B. (2012). Consecutive Fragmentation Mechanisms of Protonated Ferulic Acid Probed by Infrared Multiple Photon Dissociation Spectroscopy and Electronic Structure Calculations. *American Society for Mass Spectrometry*. **23:** 1697-1706.

56. Watson, J. D.; and Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*. **171:** 737-738.

57. Singleton, M. R.; Dillingham, M. S.; and Wigley, D. B. (2007). Structure and Mechanism of Helicases and Nucleic Acid Translocases. *Annual Reviews Biochemistry*. **76:** 23-50.

58. Hush, N. S.; and Cheung, A. S. (2006). Ionization Potentials and Donor Properties of Nucleic Acid Baes and Related Compounds. *Chemical Physics Letters*. **34 (No.1):** 11-13

59. Martens, S. M. (2011). PhD Thesis: Characterization of Several Small Biologically Relevant Molecules by Infrared Multiple Photon Dissociation Spectroscopy and Electronic Structure Calculations. *University of Waterloo*.

60. Zang, I.; and Avriel, M. (1975). On Functions Whose Local Minima Are Global. *Journal of Optimization Theory and Applications*. **16 (No. 3-4):** 183-190.

61. Jang, Y. H.; Goddard III, W. A.; Noyes, K. T.; Sowers, L. C.; Hwang, S.; and Chung, D. S. (2002). pK_a Values of Guanine in Water: Density Functional Theory Calculations Combined with Poisson-Boltzmann Continuum-Solvation Model. *Journal of Physical Chemistry*. **107:** 344-357.

62. Camargo, A. P. M.; Baumgartel, H.; and Donner, C. (2003). Coadsorption of the DNA bases thymine and adenine at the Au(111) electrode. *Physical Chemistry, Chemical Physics*. **5:** 1657-1664.

63. Bau, J. P. T.; Anizelli, P. R.; de Santana, H.; da Costa, M. F.; and Zaia, D. A. M. (2019). An experimental and theoretical vibrational study of the interaction of cytosine and uracil with artificial seawaters: a prebiotic chemistry experiment. *Vibrational Spectroscopy*. **101:** 92-99.

64. von Sonntag, C. (2006). Online Access. Free-Radical-Induced DNA Damage and Its Repair: A Chemical Perspective. *Springer*. **Chapter 10:** 211-335.

65. Cumming, J. B.; and Kebarle, P. (1978). Summary of gas phase acidity measurements involving acids AH. Entropy changes in proton transfer reactions involving negative ions. Bond dissociation energies D(A-H) and electron affinities EA(A). *Canadian Journal of Chemistry*. **56 (No. 1):** 1-9.

66. Bruice, P. Y. (2007). Online Access. Organic Chemistry 5th Edition. *Pearson*.

67. Smith, J. R.; Kim, J. B.; Lineberger, W. C. (1997). High-resolution Threshold Photodetachment Spectroscopy of OH^-. *Physics Review A*. **55 (No. 3):** 2036.

68. Mackanos, M. A.; Simanovskii, D. M.; Contag, C. H.; Kozub, J. A.; Jansen, E. D. (2012). Comparing an optical parametric oscillator (OPO) as a viable alternative for mid-infrared tissue ablation with a free electron laser (FEL). *Lasers Medical Science*. **27:** 1213-1223.

69. Ziegler, B. E.; Marta, R. A., Martens, S. M.; Martens, J. K.; and McMahon, T. B. (2012). Structure, energetics and vibrational spectra of protonated chlortetracycline in the gas phase: An experimental and computational investigation. *International Journal of Mass Spectrometry*. **316-318:** 117-125.

70. Lange, N. A.; and Speight, J. (2005). Lange's Handbook of Chemistry, 70th Anniversary Edition. *McGraw-Hill Education*.

71. Check, C. E.; Faust, T. O.; Bailey, J. M.; Wright, B. J.; Gilbert, T. M.; and Sunderlin, L. S. (2001). Addition to Polarization and Diffuse Functions to the LANL2DZ Basis Set for P-Block Elements. *Journal of Physical Chemistry A*. **105 (No. 34):** 8111.

72. Flick, J.; Appel, H.; Ruggenthaler, M.; and Rubio A. (2017). Cavity Born-Oppenheimer Approximation for Correlated Electro-Nuclear-Photon Systems. **13:** 1616-1625.

73. Velo, G.; and Wightman, A. S. (1980). Rigorous Atomic and Molecular Physics. *Plenum Press: NATO Advanced Study Institutes Series*. 185-212

74. Salomon, O.; Reiher, M; Hess, B. A. (2002). Assertion and validation of the performance of the B3LYP(star) functional for the first transition metal row and the G2 test set. *The Journal of Chemical Physics*. **117 (No. 10):** 4729-4737.

75. Featherstone, J.; Chong, T.; Martens, J. K.; Oomens, J.; and McMahon, T. B. (2018). Inverse Sandwich Cyclopentadienyl Complexes of Sodium in the Gas Phase. *The Journal of Physical Chemistry A*. **122:** 8659-8664.

76. Orio, M.; Pantazis, D. A.; and Neese, F. (2009). Density functional theory. *Photosynthesis Review*. **102:** 443-453.

77. Bakker, J.M.; Aleese, L. M.; Meijer, G.; von Helden, G. (2003). Fingerprint IR Spectroscopy to Probe Amino Acid Conformations in the Gas Phase. *Physical Review Letters*. **91:** 203003

78. Simon, A. (2002). PhD Thesis: Activation du méthane par deux cation métalliques de la troisième série, W+, et Ta+ : synergie entre expériences en phase gazeuse et calculs de chimie quantique. *Université Paris Sud*.

79. Kossmann, S.; Kirchner, B.; Neese, F. (2007). Performance of modern density functional theory for the prediction of hyperfine structure: meta-GGA and double hybrid functionals. *Molecular Physics*. **105 (No. 15-16):** 2049-2071.

80. Gaussian 16 (2016), Online Access, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

Appendix

Appendix A

Sample Input File for Geometric Optimization & Frequency Calculations Using A Split Basis Set:

```
%nosave
%mem=9000mb
%nproc=12
# opt freq b3lyp/gen

dU3-Na2- : Train Geometry

-1 1
C             4.14177743   6.58593551   0.61876878
H             4.09490505   6.06486109   1.57286093
C             5.09949970   7.54729396   0.43156067
C             5.16221777   8.23693632  -0.82893777
O             5.94685011   9.12054690  -1.17026114
N             4.17367549   7.79100021  -1.72635474
C             3.22342339   6.81232185  -1.48579780
O             2.40188874   6.53111385  -2.39482238
N             3.21042194   6.19654296  -0.28456490
C            -1.63574004  -3.97221392  -2.69129113
H            -0.79587627  -3.56025349  -3.24904797
C            -1.95828986  -5.29060531  -2.87410353
C            -3.06529963  -5.84342911  -2.14226155
O            -3.49641712  -6.99586701  -2.17804708
N            -3.67361651  -4.88965432  -1.30967986
C            -3.31053269  -3.55711323  -1.15585728
O            -3.98058407  -2.85842504  -0.35939197
N            -2.25890652  -3.08965670  -1.87084307
Na           -3.01402387  -1.13074539  -0.71219829
H            -4.46269031  -5.20555823  -0.76137635
H            -1.40843881  -5.92954587  -3.55022567
H             5.81267424   7.80699392   1.20079749
Na            1.58715406   4.87255066  -1.17377100
C            -1.52396262   4.09262491  -0.26752971
H            -1.87622550   5.02038993   0.17874569
C            -2.33027435   2.98597205  -0.22675809
C            -1.86589983   1.75924830  -0.81660744
O            -2.44956543   0.67804613  -0.87346942
N            -0.58101170   1.89353574  -1.37563948
C             0.19470890   3.04094148  -1.39166350
O             1.32597182   2.99828571  -1.93838373
N            -0.29042395   4.16482001  -0.82270945
H            -0.18828851   1.06738747  -1.80736904
H            -3.30719800   3.00209038   0.23492298
H             4.15912692   8.23147128  -2.63680465

C H N O o
6-311++G(d,p)
****
Na    0
S   14   1.00
 423000.0000000          0.180618D-04
  63340.0000000          0.140430D-03
  14410.0000000          0.738438D-03
   4077.0000000          0.311182D-02
   1328.0000000          0.112081D-01
    478.6000000          0.352828D-01
    186.2000000          0.959897D-01
     76.9200000          0.213735D+00
     33.3200000          0.348688D+00
     15.0000000          0.324566D+00
      6.8690000          0.112633D+00
      2.6830000          0.706797D-02
      1.1090000          0.598010D-03
      0.0601500         -0.530870D-05
S   14   1.00
 423000.0000000          -0.440653D-05
  63340.0000000          -0.343443D-04
  14410.0000000          -0.180114D-03
```

```
   4077.0000000          -0.763900D-03
   1328.0000000          -0.275248D-02
    478.6000000          -0.886016D-02
    186.2000000          -0.247939D-01
     76.9200000          -0.605995D-01
     33.3200000          -0.116446D+00
     15.0000000          -0.162437D+00
      6.8690000          -0.438891D-01
      2.6830000           0.337917D+00
      1.1090000           0.561347D+00
      0.0601500           0.406754D-02
S  14   1.00
 423000.0000000           0.663019D-06
  63340.0000000           0.515769D-05
  14410.0000000           0.271250D-04
   4077.0000000           0.114635D-03
   1328.0000000           0.415118D-03
    478.6000000           0.132978D-02
    186.2000000           0.375595D-02
     76.9200000           0.914025D-02
     33.3200000           0.179859D-01
     15.0000000           0.251477D-01
      6.8690000           0.763522D-02
      2.6830000          -0.614589D-01
      1.1090000          -0.115721D+00
      0.0601500           0.626406D+00
S  1  1.00
      0.4540000           1.0000000
S  1  1.00
      0.0238200           1.0000000
S  1  1.00
      4.1890000           1.0000000
S  1  1.00
      0.6260000           1.0000000
S  1  1.00
      0.0066500           1.0000000
P  8  1.00
    243.3000000           0.224392D-02
     57.3900000           0.173997D-01
     18.1000000           0.774125D-01
      6.5750000           0.219102D+00
      2.5210000           0.378522D+00
      0.9607000           0.394902D+00
      0.3512000           0.160424D+00
      0.0982700           0.233311D-02
P  8  1.00
    243.3000000          -0.222401D-03
     57.3900000          -0.174277D-02
     18.1000000          -0.775456D-02
      6.5750000          -0.225187D-01
      2.5210000          -0.384330D-01
      0.9607000          -0.450177D-01
      0.3512000          -0.192132D-01
      0.0982700           0.182697D+00
P  1  1.00
      0.0373400           1.0000000
P  1  1.00
      0.0150000           1.0000000
P  1  1.00
      1.5690000           1.0000000
P  1  1.00
      0.5120000           1.0000000
P  1  1.00
      0.0070000           1.0000000
D  1  1.00
      0.1367000           1.0000000
D  1  1.00
      0.0636000           1.0000000
D  1  1.00
      5.4040000           1.0000000
D  1  1.00
      1.5300000           1.0000000
D  1  1.00
      0.0223000           1.0000000
F  1  1.00
      0.1397000           1.0000000
F  1  1.00
      3.4650000           1.0000000
F  1  1.00
      0.0714000           1.0000000
****
```

Appendix B

Sample Input File for Single Point Energy Calculations Using Split A Basis Set:

```
%nosave
%mem=9000mb
%nproc=8
# mp2=full/gen

Title Card Required

-1 1
C              7.03184000   -0.71286400    1.59429000
H              6.54825600   -1.06136400    2.50509400
C              8.39880500   -0.60357200    1.56531500
C              9.03981000   -0.14382500    0.36255600
O             10.24429200    0.01287500    0.16171800
N              8.11514800    0.13977400   -0.65906000
C              6.73475300    0.01351300   -0.58316700
O              6.03301900    0.30437900   -1.57797400
N              6.18610000   -0.42536200    0.57961500
C             -7.30347200   -1.73178900   -0.07953600
H             -6.82835900   -2.70228700   -0.20728500
C             -8.66918500   -1.66144900   -0.00211400
C             -9.29869800   -0.37789900    0.16695400
O            -10.50014000   -0.13350100    0.25459500
N             -8.36188200    0.67170100    0.23080500
C             -6.98447600    0.55158800    0.14751900
O             -6.26743600    1.57782300    0.22036800
N             -6.44637500   -0.68533400   -0.01193300
Na            -4.33727000    0.37171800   -0.00524100
H             -8.73242200    1.60569200    0.34938400
H             -9.29495300   -2.54021300   -0.06374200
H              9.01427900   -0.85227800    2.41808800
Na             4.05519600   -0.21056300   -0.48331300
C              1.03422800    1.81734000    0.20603300
H              1.51692900    2.76027400    0.45423500
C             -0.34398600    1.74392800    0.23493300
C             -0.96316400    0.50745200   -0.09194900
O             -2.18982700    0.25845400   -0.11826900
N             -0.06585500   -0.50261100   -0.40919000
C              1.32513600   -0.38728100   -0.42680200
O              2.01683600   -1.37414900   -0.73069800
N              1.87146100    0.81702300   -0.10599600
H             -0.44917800   -1.40730300   -0.65146700
H             -0.95981800    2.59234600    0.49553500
H              8.49439100    0.46819600   -1.53753100

C H N O o
6-311++G(d,p)
****
Na    0
S   14   1.00
 423000.0000000          0.180618D-04
  63340.0000000          0.140430D-03
  14410.0000000          0.738438D-03
   4077.0000000          0.311182D-02
   1328.0000000          0.112081D-01
    478.6000000          0.352828D-01
    186.2000000          0.959897D-01
     76.9200000          0.213735D+00
     33.3200000          0.348688D+00
     15.0000000          0.324566D+00
      6.8690000          0.112633D+00
      2.6830000          0.706797D-02
      1.1090000          0.598010D-03
      0.0601500         -0.530870D-05
S   14   1.00
 423000.0000000         -0.440653D-05
  63340.0000000         -0.343443D-04
  14410.0000000         -0.180114D-03
   4077.0000000         -0.763900D-03
   1328.0000000         -0.275248D-02
    478.6000000         -0.886016D-02
    186.2000000         -0.247939D-01
     76.9200000         -0.605995D-01
     33.3200000         -0.116446D+00
     15.0000000         -0.162437D+00
      6.8690000         -0.438891D-01
```

```
     2.6830000           0.337917D+00
     1.1090000           0.561347D+00
     0.0601500           0.406754D-02
S  14   1.00
 423000.0000000           0.663019D-06
  63340.0000000           0.515769D-05
  14410.0000000           0.271250D-04
   4077.0000000           0.114635D-03
   1328.0000000           0.415118D-03
    478.6000000           0.132978D-02
    186.2000000           0.375595D-02
     76.9200000           0.914025D-02
     33.3200000           0.179859D-01
     15.0000000           0.251477D-01
      6.8690000           0.763522D-02
      2.6830000          -0.614589D-01
      1.1090000          -0.115721D+00
      0.0601500           0.626406D+00
S  1   1.00
      0.4540000           1.0000000
S  1   1.00
      0.0238200           1.0000000
S  1   1.00
      4.1890000           1.0000000
S  1   1.00
      0.6260000           1.0000000
S  1   1.00
      0.0066500           1.0000000
P  8   1.00
    243.3000000           0.224392D-02
     57.3900000           0.173997D-01
     18.1000000           0.774125D-01
      6.5750000           0.219102D+00
      2.5210000           0.378522D+00
      0.9607000           0.394902D+00
      0.3512000           0.160424D+00
      0.0982700           0.233311D-02
P  8   1.00
    243.3000000          -0.222401D-03
     57.3900000          -0.174277D-02
     18.1000000          -0.775456D-02
      6.5750000          -0.225187D-01
      2.5210000          -0.384330D-01
      0.9607000          -0.450177D-01
      0.3512000          -0.192132D-01
      0.0982700           0.182697D+00
P  1   1.00
      0.0373400           1.0000000
P  1   1.00
      0.0150000           1.0000000
P  1   1.00
      1.5690000           1.0000000
P  1   1.00
      0.5120000           1.0000000
P  1   1.00
      0.0070000           1.0000000
D  1   1.00
      0.1367000           1.0000000
D  1   1.00
      0.0636000           1.0000000
D  1   1.00
      5.4040000           1.0000000
D  1   1.00
      1.5300000           1.0000000
D  1   1.00
      0.0223000           1.0000000
F  1   1.00
      0.1397000           1.0000000
F  1   1.00
      3.4650000           1.0000000
F  1   1.00
      0.0714000           1.0000000
****
```