Computational Model of the Nucleophilic Acyl Substitution Pathway

Ethan M. Belknap, Dr. Justin B. Houseknecht
Department of Biochemistry, Wittenberg University, Springfield, OH, 45504
University Honors Thesis (2021)

"I affirm that my work upholds the highest standards of honesty and academic integrity at Wittenberg, and that I have neither given nor received any unauthorized assistance."

Abstract

Nucleophilic acyl substitution (NAS), specifically as it proceeds through the associative pathway, is a fundamental chemical reaction that is found as a component in many biochemical pathways. A defining feature of nucleophilic acyl substitution is the formation of a tetrahedral intermediate, which functions to stabilize the reaction complex and lower the overall energy of reaction. This formation of an energetically favorable tetrahedral intermediate is a key aspect in the discussion of NAS as an important mechanistic component of biochemical and organic reactions. However, there is evidence to suggest that given the correct reaction conditions the dissociative pathway of nucleophilic acyl substitution, in which no tetrahedral intermediate is formed, is energetically favored over the associative pathway. The potential for the dissociative pathway's favorability provides great incentive to explore this pathway by way of computational chemistry, and also grants the opportunity to model the energetics and geometry of the reaction and explore changes to them under varying conditions. In order to do so, *ab initio* calculations were utilized at every level of analysis within this research to determine the energy of the acyl system under different conditions modeled within the Wittenberg computing cluster (WARP2). By using these calculations, model systems were created for the NAS system in which correct transition state geometries and energetics were identified, and these data provide potential avenues for further development of the dissociative pathway model in the future.

Introduction

Project Goals and Significance

This project aims to develop a geometric and energetic model of the dissociative pathway of nucleophilic acyl substitution. This dissociative pathway, which is one of the pathways illustrated below in Figure 1, is an alternative route to the more common associative pathway of nucleophilic acyl substitution (also shown in Figure 1) in which a tetrahedral intermediate is formed. This tetrahedral intermediate is primarily characterized by a strong nucleophile (Nu) attack on the central carbon of an acyl compound. This causes the double bond between carbon and oxygen to break sending the electrons towards the oxygen giving it a negative formal charge. The electrons then reform the double bond between carbon and oxygen, which breaks the bond between the leaving group (LG) and carbon causing it to leave. When the reaction moves forward through the dissociative pathway, an intermediate exists wherein neither the Nu or LG are bound to the acyl compound, and a triple bond is formed between carbon and oxygen in which the oxygen possesses a positive formal charge.

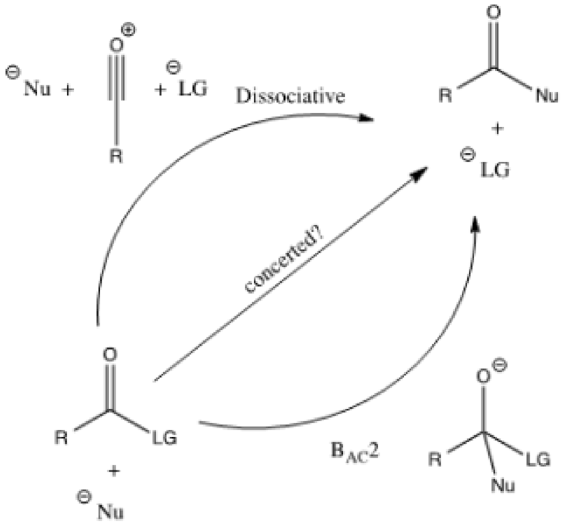

Figure 1- Associative, concerted, and dissociative pathways of nucleophilic acyl substitution.

The primary inspiration for this research is driven by the existence of studies, such as Lee, Lee, and Kim (1997), that suggest in some instances the dissociative pathway is actually energetically favorable under certain circumstances. The potential for the dissociative pathway being energetically favorable is one of the central aspects that makes this research so interesting. Should the dissociative pathway be found to be energetically favorable under consistent and reproducible circumstances, then it would provide the means for a new framework of modeling NAS reactions, and potentially other substitution reactions as well, through computational chemistry.

The fact that essentially everything to do with this project has a foundation in computational chemistry is significant on its own. Computational chemistry provides the means to determine great amounts of information about chemical pathways in terms of energetics, effects of solvation, and geometries of reaction intermediates, among a host of other informative capabilities with relative ease. Not only this, but as David Young indicates in his book *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*, computational chemistry has come incredibly far in its availability of use since its inception. It is no longer just a set of techniques exclusive to experts in the field, but is instead now available to most anyone interested in the prospect of this type of chemistry. This research being made possible is a direct consequence of this shift in availability and education regarding computational methods, and the potential for further growth is just as exciting. Not only for the efficiency and time-saving capabilities computational chemistry offers, but for the inherent “green” qualities it possesses as well. At a time when considerations for the environmental impact of research chemistry is more important and more emphasized than ever, the importance of computational methods being able to model reactions free of waste is something that can not be understated.

The potential for nucleophilic acyl substitution modeled computationally is naturally interesting, given there are now several contributions from our research group, as well as several key contributions from previous projects focused on the same subject, that have provided the basis for the methodology and approaches that were applied throughout this research period. One of the areas in which this past research formed a basis for this current project was the conclusion that the M08HX/aug-cc-pVTZ level of theory for Gaussian calculations would be the best option for obtaining accurate, reliable results given the limited time frame. These desired results would include transition state structures which indicate geometries as the reaction progresses, IRC calculations which indicate the achievement of accurate transition states following Gaussian calculations, and energy calculations which allow for comparisons of favorability between reactions under a different parameters (for example, energetic differences between aqueous and gas phases) With that said, it is important to provide a basis for the techniques used and the theory behind them before elaborating more on the information that was ultimately gathered from this research project.

Ab initio Computations

The information regarding energetic models of NAS was gathered through computational *ab initio* calculations. *Ab initio* calculations are designed as computational means to solve a given wavefunction, and this is important given that every chemical system is said to possess a wavefunction. The wavefunction is representative of electron and nuclear positions within a given system, and when functionally combined with an operator (in this instance the Hamiltonian Operator which provides an approximation for the energy of the system) characteristics of reaction pathways can be determined. The Hamiltonian Operator accounts for a few factors within a given chemical system such as electron-electron and nucleus-nucleus repulsion forces, attractive forces between electrons and nuclei, and the kinetic energy of the electrons and nuclei. These combined parameters provide a great means to solve for the energy of simple chemical systems; however, this cannot be solved for multi-electron systems. Because of this, computational methods must be used to make approximations of the SWE (Schrodinger wave equation) for a multi-electron system.

One of these approximation methods is the Hartree-Fock approximation (Sherrill 2000), in which electrons experience implicit effects of other electrons rather than the explicit effects of specific electrons. This general approximation of electron effects means that specific electrons within a system experience other electrons as a cloud, meaning their interactive forces are indistinguishable from one another. What this method ultimately does is obtain an atomic orbital, and through the combination of atomic orbitals, molecular orbitals can be generated. These molecular orbitals are what are actually used in order to create a final approximation of a system's energy. While Hartree-Fock approximations are valuable for their ability to provide fairly accurate results for systems of reasonable size over a short amount of time, they also possess a few shortcomings. For one, these approximations consistently overestimate the effects of electron-electron repulsion within a given system. As a consequence, Hartree-Fock approximations consistently lead to an over estimation of molecular energy. The resulting high estimation of molecular energy also results in approximate bond lengths that are too short, creating a misrepresentation of molecular geometry. The nature of the geometry that results from Hartree-Fock approximations means that representations of transition states can only ever be so accurate when applying it, but as mentioned before these approximations are still useful within small enough systems. However, the shortcomings of approximations such as Hartree-Fock can be overcome through the implementation of electron-correlation methods.

There are two types of functions that are used to model molecular orbitals. The first is a Slater-type orbital (STO), which is not preferable for the sake of this project because it is impossible to integrate the STO analytically. It is for this reason that the other type of molecular orbital, the Gaussian-type orbital (GTO), is preferred for the purposes of these calculations. The GTOs can be combined to form approximations of the SWE, and the different ways in which these GTOs are combined computationally become referred to as basis sets. Essentially, as basis sets account for more GTOs, then the approximations of the SWE they create will become more accurate. However, as the approximation is improved by the addition of more basis sets, the amount of time needed to complete a calculation will also increase. This trade-off between accuracy of approximation and time required for the calculation was the main factor considered when choosing a basis set for the *ab initio* calculations for this research. Ultimately, the basis set settled on for this research was the aug-cc-pVTZ basis set, and this will be used for all computations necessary to model the nucleophilic acyl substitution pathway energetically. This basis set belongs to a subset of Dunning's correlation consistent basis sets, and according to Gaussian.com, these correlation consistent basis sets increase computational efficiency by the removal of redundant functions. Because of this computational efficiency, the aug-cc-pVTZ seemed like the best choice for this project.

Before discussing some of the goals and questions that were in place for this project, it will be useful to outline several of the Gaussian calculations that were used in order to determine different aspects of the nucleophilic acyl substitution pathway. Firstly, QST2 and QST3 calculations were implemented with suspected approximations of transition state geometries obtained in Spartan, and these calculations yield more accurate representations of transition state geometries achieved through the reaction pathway. From these geometrical coordinates, further determinations regarding the energy of reaction can be made. In order to ensure the geometries obtained through QST2 and QST3 calculations are accurate, the coordinates must be put through intrinsic reaction coordinate (IRC) calculations. These IRC calculations run the reaction modeled by the transition state input in both the forward and reverse directions, and proper completion of IRC calculations ensures that the transition state geometry is achievable in both directions of reaction. The figure illustrated below (Figure 2) is an example of an IRC output which indicates an accurate transition state geometry for a given reaction. The hyperbolic nature of the curve is indicative of an energy maximum achieved which correlates to an accurate transition state input. As the reaction coordinate approaches 0 from both directions, energy is approaching its maximum. This can be seen in data regarding this project later, in the Results section.

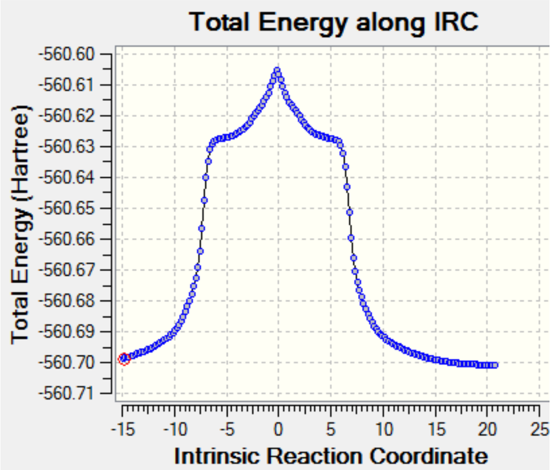

Figure 2- Example of an IRC Energy Diagram

(This image was obtained through ResearchGate.net)

Lastly, solvation effects needed to be modeled as well in order to differentiate how the nucleophilic acyl substitution pathway behaves in the gas and aqueous phases. This was accomplished through the implementation of the Polarizable Continuum Model (PCM). PCM accounts for the effects a solvent has on a given reaction essentially by making the desired solvent field self-consistent with the electrostatic potential of the solute, as stated in (Improta, *et al.* 2007). What this does is allow for the reaction to proceed while experiencing the collective implicit effects of solvent (for the purposes of this project, water) as opposed to the explicit effects of the molecules that make up the solvent. This is a time effective manner in order to determine the effects that an aqueous solution would have on a given reaction, which in this case is nucleophilic acyl substitution. With all this established, the core motivations for this project can be elaborated on.

Research Questions

There are a few questions that remained the focus for the duration of the research period. The first of these questions is do the gas-phase calculations differ significantly from the aqueous phase calculations, and if so which is more favorable? Differences that would be looked for in how these calculations perform would be based mainly around energetics and the transition state geometries achieved in each phase. Are there significant differences in the energy of the nucleophilic acyl substitution reaction between these phases that suggest favorability? Additionally, an important piece of information Gaussian calculations would ideally provide would be how well the dissociative pathway is achieved in one phase opposed to the other. Is the dissociative pathway driven more easily in the aqueous phase or gas phase, or do neither of these phases significantly affect the performance of the dissociative pathway?

It is also desirable to know which Nu/LG pairs favor the self-displacement nucleophilic acyl substitution reaction, and as a result which pairs function in the most energetically favorable manner through the dissociative pathway. For instance, is the self-displacement reaction that takes place for chloride-chloride interaction more favored along the dissociative pathway than an acyl compound where thiolate-thiolate displacement occurs, and if so what are some of the reasons for that? The hypothesis is that the two phases will differ significantly from each other, and that the gas-phase pathway will be the more energetically favorable of the two resulting in an overall lower energy of reaction. The question of which Nu/LG pair will most favor the dissociative pathway is known to be dependent on a few factors, as is discussed in Lee, Lee, and Kim (1997). These are the strength of the nucleophile as well as the ability of the leaving group to leave. Because of this, it would be expected that chlorine may work the best in the dissociative pathway. This is due to chlorine being a relatively weak nucleophile, and it also serves as a good leaving group for the sake of self-displacement reactions.

The first step in modelling any reaction pathway involves constructing models of products, reactants, and transition states. This was carried out in Spartan version 10. From here, semi empirical methods such as AM1 were used to quickly optimize geometries and create energy profiles. The coordinates of these compounds were then transferred over to WARP2, Wittenberg’s computing cluster. WARP2 was to perform Gaussian calculations using the M08HX/aug-cc-pVTZ level of theory, which in turn created more accurate approximations of reactant, product, and transition state geometries. These calculations also provide energetic data, which can be used in tandem with the geometric data to model the dissociative pathway. ChemCraft is also important in visualizing output files from Gaussian to ensure optimization steps and transition states are being properly reached.

Methods

All molecular geometries and complexes were constructed in the Spartan ‘10v1.1.0 software. Spartan ran on Windows 10 operating on an Intel Core i5 processor. Spartan was used to conduct geometry optimizations for the reactants and products of the self-displacement reactions at the AM1 level of theory. Spartan was also used in order to locate approximations of the transition states for the self-displacement reactions. This was accomplished by creating an energy profile of the reaction complex as the nucleophilic species approached the central carbon, which in turn increases the bond length between the LG and central carbon. The change in relative energy that results from the changing in bond length can be graphically evaluated in Spartan, and by locating the energy maxima along this process a reasonable approximation of the reaction’s transition state was made. The coordinates for the transition state approximation were then taken from Spartan and used for Gaussian calculations.

Gaussian computations were conducted through the 52-core WARP2 computer processing cluster with the CentOS operating system and 64 bit Intel Xeon processors, and the calculations were performed on Gaussian ‘16 Revision AO3 software which ran through the WARP2 cluster. WARP2 was then used to conduct necessary Gaussian calculations such as QST3, IRC, geometry optimizations, and frequency calculations based on the coordinates taken from Spartan. Each of these calculations were performed at the PCM/M08HX/aug-cc-pVTZ level of theory. QST3 calculations were used to determine the structure and energy of the transition states. IRC calculations are used to ensure the transition state geometries found by QST3 calculations are accurate. For the purposes of this research, PCM was used to model the implicit effects of water as a solvent for QST3, optimizations, and other Gaussian calculations.

This in turn allows for the energetic and geometric model of the dissociative pathway in water, which is a necessary step for complete comparisons across the associative and dissociative pathways in the gas phase.

Results

Transition state structures obtained from QST2 calculations were then loaded in ChemCraft so they could be visualized. Figures 3 and 4 illustrate the gas phase and aqueous phase transition states respectively. In each figure, the bond lengths of the carbon-chlorine bonds (chlorine is highlighted in green in Figure 3) are displayed in angstroms.

Frequency calculations were made as part of the QST2/3 input in Gaussian software. The transition state structure generated from this has one negative frequency value, and this negative frequency value can be examined in the ChemCraft software to determine that the frequency correlates to the correct molecular motion. In this instance, the negative frequency values correlate to the stretching and lengthening of the chlorine-carbon bond as was desired. The specific negative frequency values were -271 Hz for the aqueous phase transition state, and -168 Hz for the gas phase transition state.

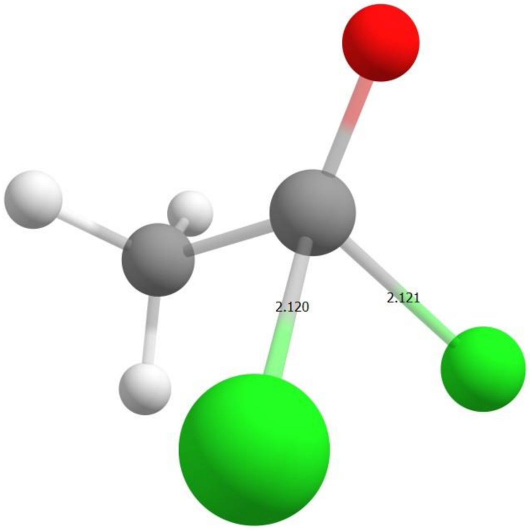

Figure 3- Transition State Structure for the Gas Phase Concerted Pathway

Frequency: -168 Hz

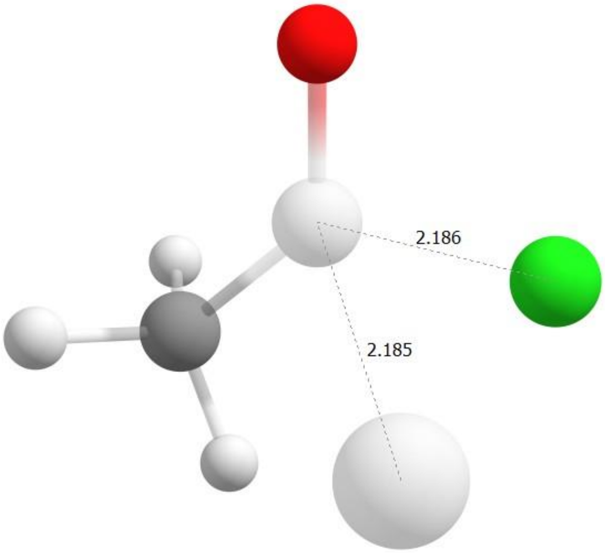

Figure 4- Transition State Structure for the AQ Phase Concerted Pathway

Frequency: -271 Hz

Note: Only one chlorine is highlighted in green. The other is the atom with a bond length listed.

IRC calculations serve as the means to ensure the correct transition state has been reached for a given reaction complex and set of Gaussian QST2 and QST3 calculations. This is a necessary step in the process of modeling a reaction pathway computationally, otherwise the transition state reached through Gaussian QST2/3 calculations may be a misrepresentation of the reaction pathway. IRC calculations demonstrate accurate transition states through parabolic curves, which reflect the reaching of a given transition state in both the forward and reverse directions of reaction. Figures 5 and 6 feature IRC data for the chloride-chloride self-displacement reaction for the aqueous and gas phases respectively. This data was modeled by graphing relative energy of the structure (kJ/mol) as a function of the change in the carbon-chlorine bond length (angstroms).

Figure 5- Aqueous Phase IRC for the Concerted Pathway of Cl-Cl Self-displacement

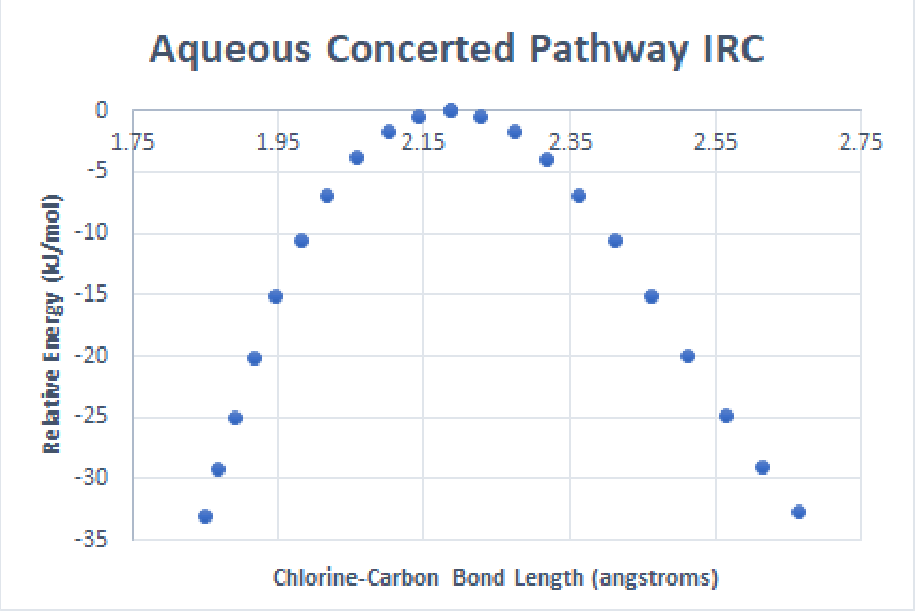

Figure 6- Gas Phase IRC for the Concerted Pathway of Cl-Cl Self-displacement

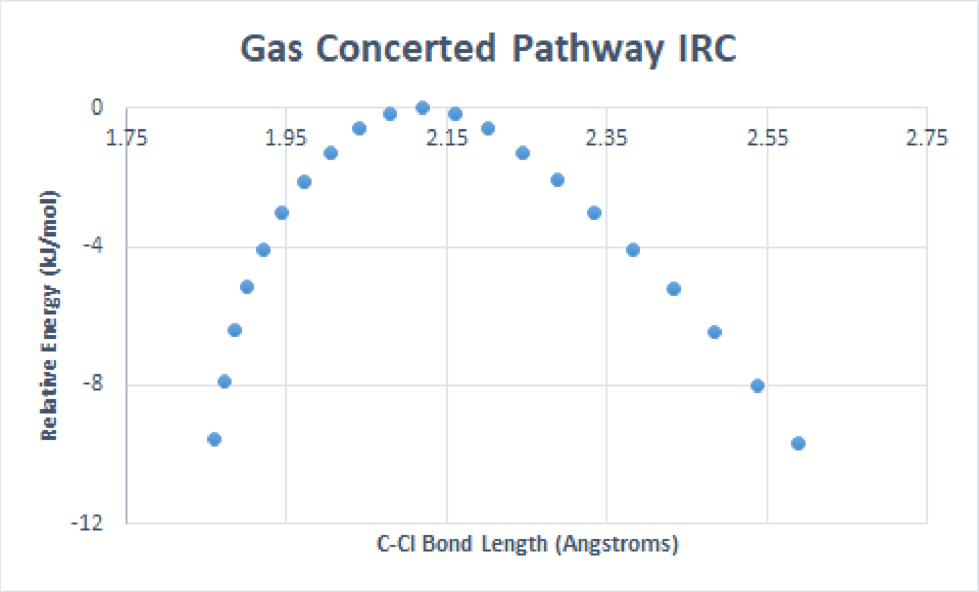

Table 1 includes geometric data for the transition state structures of the chloride-chloride self-displacement reaction in both the gas and aqueous phases. These structures were obtained through Gaussian QST2 calculations at the M08HX/aug-cc-pVTZ and PCM/M08HX/aug-cc-pVTZ levels of theory. The M08HX/aug-cc-pVTZ method was utilized in the case of the gas phase, and the PCM/M08HX/aug-cc-pVTZ method was used in the case of the aqueous phase. This data was used to determine whether or not there is a significant difference between the gas and aqueous phase when modeling NAS, particularly as it pertains to the transition state geometries reached in each and their overall energies of reaction.

Table 1- Geometric Data from the Gas and Aqueous Phases

Structure	C-Cl BL (Nu) (Å)	C-Cl BL (LG) (Å)	O-C BL (Å)	a ˚O-C-Nu	a˚ O-C-LG	a˚ O-C-C
Acetyl Cl (AQ)	-	1.795	1.170	-	118.67	128.33
Acetyl Cl (Gas)	-	1.744	1.225	-	120.79	125.62
TS (AQ)	2.186	2.185	1.171	110.61	110.60	131.65
TS (Gas)	2.121	2.120	1.183	112.48	112.51	127.62

Discussion

Data regarding IRC calculations for both the gas and aqueous phases indicate that the correct transition state was found for each. The curves seen in both Figures 5 and 6 is what is desired when looking at IRC results. This shows that at a given point along the carbon-chloride bond distance there exists a point where relative energy is at a maximum for both phases, which indicates the correct transition state structures have been located for the gas and aqueous phases. The transition structures reached for the gas and aqueous phases both have nucleophile and leaving group bond lengths are equal distance away from the central carbon, as shown in Figures 3 and 4. This indicates that these transition states are reflective of the concerted pathway of NAS. This was not the original pathway of interest when it came to modeling the NAS reaction; however, data based on the concerted pathway still possesses important insights into the overall reaction. For instance, data shown in Table 2 indicate that whether or not the chloride-chloride self-displacement reaction proceeds through the gas phase or aqueous phase does not have a significant effect on the resulting transition state structures. The carbon-chlorine bond lengths do not differ significantly across the two phases (2.185 Å for the aqueous phase and 2.120 Å for the gas phase), and the bond angles present in each also do not experience a significant change (110.60° for the aqueous phase O-C-Cl bond angle compared to the 112.51° angle present in the gas phase, and 131.65° for the aqueous phase O-C-C bond angle compared to the gas phase angle of 127.62°). This geometric data does support the idea that the NAS concerted pathway is not significantly altered based on whether it takes place in the gas phase or aqueous phase.

Unfortunately, MP2 calculations requiring far more time than anticipated meant one of the areas of interest addressed at the beginning of the research period could not be fully realized. MP2 effects on the energetics of a carbon-fluorine self-displacement (through the associative pathway) had previously been outlined, but energies of activation had not been established with MP2 for other self-displacement Nu/LG pairings. Therefore, the goal was to establish the potential differences with other self-displacements such as thiolate-thiolate and cyanide-cyanide using MP2 calculations to determine if the MP2 effects were consistent. Because MP2 calculations could not complete in a reasonable time frame, these potential differences in MP2 effects remain to be determined. Though these MP2 calculations could not be reconciled, there is likely a benefit to be gained from the energetic and geometric determinations produced through the M08HX/aug-cc-pVTZ method.

Another one of the main ideas present from the beginning of this project was the desire for an energetic model of the dissociative pathway. There were efforts made to model this aspect of the pathway using chloride-chloride self-displacement, but unfortunately no such model was able to be produced. This aspect of the research proved to be difficult because the geometry for a dissociative progression of nucleophilic acyl substitution requires precise geometry inputs as well as trial and error in order to obtain suitable inputs for Gaussian calculations.

When looking at some areas of the research process that presented challenges, a major issue that recurred throughout the duration of this research project was reaching the correct transition states for the dissociative pathway when using the chlorine-chlorine Nu/LG pair. Obviously this is such an integral aspect to the research being successful, so the repeated nature of this issue was unfortunate. Part of this problem was a product of the long time period often required for Gaussian calculations to complete. For instance, mod redundant calculations used to lengthen the bond between a thiolate leaving group and a central carbon ended up requiring three days to complete. This on its own is not a desirable time stipulation to be working with, but the issue is also compounded by the chances of calculations failing halfway through completion. In referencing the same mod redundant calculation from before, it was found that the molecule was undergoing rotational motion which caused the calculation to fail. This could not be determined until the calculation had already completed, so the necessary adjustments could only be made after the fact, thus requiring more time for the corrected calculation to begin again. This is the type of dynamic that made significant progress difficult to achieve.

Future Work

One of the main goals this research aimed to accomplish was unfortunately not able to be reached. The dissociative pathway of nucleophilic acyl substitution proved more difficult to create a model for than was originally anticipated, but hopefully future work on the subject will be able to provide a functional model. Much of the work and effort that went into this research was centered around collecting reproducible and accurate data for the associative pathway, so the hope is that by building on that solid foundation the creation of a functional dissociative model will be made that much easier. Because the dissociative pathway model was not produced, the work that went into modeling aspects of the associative pathway, such as its function in an aqueous environment versus a gaseous medium, became the greater focus.

With both the gas and aqueous phases being used to model this pathway, the aim is also to be able to make better comparisons based on the differences in energetics and geometries of these two phases. While conclusions could be drawn from this data as is, the hope is that further efforts regarding this research are able to benefit from these models and ultimately draw more conclusions, while also asking further questions about the information these models provide. Another future goal is to model the self-displacement reaction using different Nu/LG pairs other than chloride-chloride reactions.

A pretty significant departure in the intended trajectory of this work came with the shift away from MP2 methodology in favor of the M08HX/aug-cc-pVTZ basis set. This choice was made mainly based on the time required for the Gaussian calculations to complete for each method. From the outset, MP2 calculations would take far longer to complete on the chloride-chloride self-displacement system than calculations for the M08HX/aug-cc-pVTZ method would require for the same system. The differences in time required for completion were so pronounced that MP2 calculations would often take several days longer than the M08HX/aug-cc-pVTZ technique; where MP2 might take 48 hours to complete, M08HX/aug-cc-pVTZ would finish in an afternoon). This occurrence was considerably strange because MP2 calculations were demonstrated in the past to require far less time to reach completion than the calculations needed over the course of this research.

Developing models of the associative pathway that do not utilize the same displacement pairing as chloride-chloride would elaborate the general understanding of the energetics of NAS, and would provide insight into what types of specific interactions NAS would most likely be favorable within. Additionally, efforts to utilize different Nu/LG pairs for self-displacement would likely be beneficial to the general understanding of the dissociative pathway, but it should be acknowledged that this would follow the establishment of a dissociative model. This would include furthering work centered around cyanide-cyanide self-displacement through the dissociative pathway, or another Nu/LG pairing of interest outside of chlorine-chlorine self-displacement . It would also be worthwhile to explore the thiolate-thiolate self-displacement reaction, but presently there has been little progress made into this reaction. These alternative Nu/LG pairs include thiolate and cyanide, but options outside of these should also be considered.

It should also be mentioned that even though this research addresses the differences in geometries and energetics produced across the gas and aqueous phases, solvent polarity is still an important consideration that carries great potential to affect the NAS pathway. The models represented in this paper were created using water as the simulated environment for the aqueous phase, so the potential information that could be taken away from modeling chlorine-chlorine self-displacement in other media is exciting. This could be done using other polar solvents such as ethanol or acetic acid. It should be said that exploring the effects of a non-polar solvent like acetone would not be as productive, given that non-polar environments would not be conducive to the nucleophilic acyl substitution reaction.

Additionally, questions still remain about low energies of activation produced in a previous study for the thiolate-thiolate self-displacement (based on bottom of the well calculations). Furthering the use Of MP2 or M08HX/aug-cc-pVTZ calculations may help provide answers for why this activation energy is so low for this associative pathway.

References

Lee, I.; Lee, D.; Kim, C. K. Theoretical Studies of Structural Effects on the Mechanism of Acyl

Transfer Reactions1. The Journal of Physical Chemistry A 1997, 101 (5), 879–885.

Radisky, E., Lee, J., Lu, C., & Koshland, D. (2006, May 02). Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. Retrieved March 31, 2021, from https://www.pnas.org/content/103/18/6835

Libretexts. (2020, September 23). 21.2 nucleophilic acyl SUBSTITUTION REACTIONS. Retrieved March 31, 2021, from https://chem.libretexts.org/Courses/Athabasca_University/Chemistry_360%3A_Organic_Chemistry_II/Chapter_21%3A_Carboxylic_Acid_Derivatives%3A_Nucleophilic_Acyl_Substitution_Reactions/21.02_Nucleophilic_Acyl_Substitution_Reactions

R Improta, G. Scalmani, M. J. Frisch, and V. Barone, "Toward effective and reliable fluorescence energies in solution by a new State Specific Polarizable Continuum Model Time Dependent Density Functional Theory Approach," *J. Chem. Phys.*, **127** (2007) 074504

Basis sets. (n.d.). Retrieved April 12, 2021, from https://gaussian.com/basissets/

Young, D. C. (2011). *Computational chemistry: A practical guide for applying techniques to real world problems*. Oxford: Wiley-Blackwell.