Automation of Reaction Monitoring

by

Darien Yeung
BSc, Vancouver Island University, 2016

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Chemistry

Supervisory Committee

Automation of Reaction Monitoring

by

Darien Yeung
BSc, Vancouver Island University, 2016

Supervisory Committee

Dr. J. Scott McIndoe (Department of Chemistry)
Supervisor

Dr. Dennis Hore (Department of Chemistry)
Departmental Member

Abstract

Supervisory Committee
Dr. J. Scott McIndoe, Department of Chemistry
Supervisor
Dr. Dennis Hore, Department of Chemistry
Departmental Member

Automation plays an integral role in our daily lives. From transportation to agriculture, we rely on robots and programs to assist in accomplishing tasks. Chemistry is no except with the deployment of high throughput screening and the recent machine-led reaction discovery, there is increased interest to integrate artificial intelligence and robotics beyond medicinal and synthetic organic chemistry. The addition of automation to mechanistic studies can improve the method in which reactions are understood experimentally and fundamentally.

Chapter 1 introduces the basics of reaction chemistry. As we are interested in *how* the reaction occurs, for this work, there is a natural bias towards understanding kinetic behaviour. Chronograms obtained through mass spectrometry facilitate understanding of kinetics. The introduction of mass spectrometry in this chapter establishes the foundation of this technique for the subsequent experimental chemistry chapters.

Chapter 2 investigates the reduction and subsequent oxidation of titanocene, generating a complex mixture of oxidized products. During this investigation, an interesting and rare methyl abstraction event occurred that led to the deuterium label study to understand a radical-based oxo-titanium reaction. This was made possible by Pressurized Sample Infusion Electrospray Ionization Mass Spectrometry (PSI-ESI-MS) coupled with a smartphone colorimetry technique developed herein known as ColorPixel.

In Chapter 3 we explore the integration of machine learning with reaction monitoring. The attempt to classify reaction roles based on kinetic traces was done to automate the process of identifying important species in a reaction. Often there is a large amount of data from a PSI-ESI-MS experiment, but it is time-consuming to pick out the most important species. Implementing machine learning for reaction role classification can ease this process from taking three months to accomplish to one day. This chapter also outlines the development of Kendrick, an automated reaction sampler. Combined, these tools have the potential to impact reaction monitoring through robotic assistance and can speed up the process of reaction quantification through automated processing platforms to handle the streams of data.

Chapter 4 starts with the implementation of a lightweight mass spectrometry library, Spectra.ly, that is suitable for any developers using python. This platform establishes a firm foundation that can enable developers to build complex programs using simple code. This chapter also describes the collaboration project PythoMS and the development process for this framework. In addition to the framework, the chapter also describes the development of two pieces of processing software: Sinatra – a cloud-ready EDESI processing platform, and AutoMRM – a cloud-based Multiple Reaction Monitoring method development web application.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Abbreviations

Acknowledgments

In this journey through my MSc, there are many people that I owe a Thank You to. I would like to thank my supervisor, Dr. Scott McIndoe, for giving me the opportunity to work in his laboratory and taking the risk in allowing me to pursue automation for reaction monitoring. Thank you for believing in me in the past two and a half years. I would like to thank Dr. Dennis Hore for donating his autosampler to kickstart my robotics project and guiding me through the spectroscopy and 2DCOS space. I really appreciate your enthusiasm for Python, it is an amazing language to work on. Thank you, Dr. George Tzanetakis, for teaching me how to do machine learning and other statistical analyses for my project. Through MARSYAS I have learned many things in signal processing and without that opportunity, I would not have been able to implement the Flask Remote Functions.

I would like to thank Dr. Ori Granot at the Chemistry Mass Spectrometry Facility (CMSF) with all his help repairing the mass spectrometers for my projects and being so patient with every training session. Thank you for your generosity in sharing with me your knowledge. I am immensely grateful for the guidance and support from Andrew MacDonald during the construction of Kendrick in my robotics project. He taught me everything I know about circuit design for both prototyping and printed circuit boards. Thank you to Dr. Chris Barr with all your help for the NMR and taking the time to teach me the appropriate acquisitions possible for reaction monitoring. Many of the glass and mechanical parts during my MSc would not have been possible without the help from Sean Adams and from the Physics Machine Shop (Chris Secord and Jeff).

During my time as a teaching assistant, I have received amazing support from Jane Browning and Corrina Ewan, thank you so much for all your support and adding positivity to my teaching experience. Thank you to the administrative staff, Sandra Baskett and Lori Aasebo in helping me with everything, from my application to this school and figuring out how to submit travel claims. Also, I really appreciate your support for Kendrick, my robotics project. Thank you so much Sandra Carlson and Rosemary Pulez with all the help during my time at CMSF.

Thank you to all the co-workers in the McIndoe/Rosenberg/Elvira office, it was a pleasuring working with everyone there. Michelle Ting, Natalie Dean, Elena Liles, Erica Hong, thank you for being the amazing and supportive friends that you have been and I hope to see you all again when my defense is over.

Most importantly, I would like to thank my mom and brother for being so supportive over the course of my life and every time I fell down, they helped me get back up. Thank you mom! Carene Yeo, you are an amazing partner and I really have to thank you for everything and realizing my mentality of “Go Fast and Commit!”

There are so many more people I would like to acknowledge outside of UVic and thank you everyone for entrusting your knowledge to me and making this process of my MSc possible.

Dedication

For mom, brother, and Carene
"Go fast and commit"

1. CHAPTER 1: INTRODUCTION

1.1. What makes reactions go?

Starting with the mysteries of organometallic chemistry and continuing to follow my knowledge of mass spectrometry, I have ended up being in a field combining the two together to elucidate the unknowns in reaction mechanisms. Mechanisms describe the processes by which molecules can convert into different molecules. As we have learnt in first year chemistry, there are two ways to look at a reaction: thermodynamically and kinetically. Thermodynamics describes *why* a reaction occurs and kinetics addresses *how* a reaction takes place[1].

One thing we must keep in mind in our journey to learn more about chemical reactions is to understand what a reaction is. The IUPAC Gold Book[2] defines reaction as "a process that results in the interconversion of chemical species." However, there are factors that we must consider that determine whether an interconversion will occur. Factors like heat, time, concentration, and solvent are some common examples to consider when a reaction has failed to proceed. In short, these variables can be condensed to the thermodynamics and kinetics of a reaction.

The concept of thermodynamics and kinetics are essential as many fundamental aspects of reaction mechanisms get condensed back to these two concepts. An emphasis on thermodynamics and chemical kinetics is essential as many aspects of reaction mechanisms fundamentally rely on these two concepts. For instance, computational approaches to mechanisms use density functional theory to calculate the Gibbs Free Energy values for probable conformations of the reactants combined. This provides us with insights into how

the starting material becomes the product. Gibbs Free Energy calculations describe the possibility that a process is spontaneous. In any given reaction, a change in Gibbs Free Energy (ΔG) of less than zero indicate a spontaneous reaction; while values above zero are considered non-spontaneous.

Computational mechanistic studies use intrinsic reaction coordinates (IRC) to find a reaction pathway that would be best explained by thermodynamics. From reactant to product, there are many spatial transformations, known as translational motion, that occur for each atom[3–6]. Such transformation does not only occur in locked geometry whereby an atom can be closer or further away from a bound atom, like a spring: a vibrational motion[7,8]. A simple change in bond distance can change the Gibbs Free Energy[9]. From the Boltzmann distribution, lower energy processes are favored as the population molecules are low in energy[10,11]. Each transformation is considered an intermediate and understanding how a reactant turns into an intermediate or series of intermediates to the product will offer valuable insights in the optimization of reactions and application to new moieties.

1.2. Kinetics of a reaction

Kinetics gives an idea of *how* reactions work[1]. By studying the kinetics of a reaction, it is possible to speculate on the mechanism of the reaction. From an industrial point of view, learning about a reaction can lead to cost-effective practices through rational improvements in experimental design[12].

Studying the kinetics also helps us determine the type of product that gets formed from a given reaction. In a reaction with two competing products, the product that forms first at a faster rate is known as the kinetic product. This product, however, is not necessarily the

most stable. The thermodynamic product may form more slowly, but is the more stable of the two. The rate at which a reaction takes place can lead to different products being formed[13,14]. In a reaction where there are two competing products, if one of the products can be formed at a faster rate, then it will be more favourable for that product to be formed *first*[15].

In a reaction, the rate for each probable product can be determined experimentally with differing conditions to understand how the rate changes. The factor that most greatly affects a reaction rate is its concentration.

In general, given a reaction of the form:

$$\mathbf{n_a A + n_b B + \ldots \rightarrow product} \qquad \mathbf{(1.1)}$$

The rate law is defined as:

$$\mathbf{rate = k[A]^a[B]^b} \qquad \mathbf{(1.2)}$$

whereby exponents *a* and *b* are integers that can only be determined experimentally.

From equation 1.2, we can see that the rate is proportional to the rate constant, *k*, and the concentration of reactant. The higher the concentration, the faster the reaction occurs.

Now, another concept to consider is how many molecules are involved at one instance in the reaction. A unimolecular reaction requires only one molecule to occur, and an example of this would typically be a degradation-type reaction. Therefore, the rate law would look as follows:

$$\mathbf{rate} = \mathbf{k}[\mathbf{A}] \quad \mathbf{(1.3)}$$

The above-mentioned rate law (1.3) would be referred to as a *first order* rate law[1] as the overall exponent of the equation is to the power of one. Now consider a reaction involving two molecules, also known as a bimolecular process or second-order reaction. The rate of the reaction would be proportional to the product of the concentration of the two reactants[1], or to the square of the concentration of a single reactant. The rate law would be as follows:

$$\mathbf{rate} = \mathbf{k}[\mathbf{A}][\mathbf{B}] \quad \mathbf{(1.4)}$$

$$\mathbf{rate} = \mathbf{k}[\mathbf{A}]^{\mathbf{2}} \quad \mathbf{(1.5)}$$

The overall exponent of this rate law equation is now to the power of two; and therefore, the resulting rate would increase quadratically with respect to changes in concentration of the reactants. However, in a scenario where we have a bimolecular reaction, and one of the reactants has a significantly greater concentration relative to the other, the change in the more abundant species during the reaction would be minimal, to the point where we can consider this change negligible[1]. The resulting rate law would look similar to a first order rate law as we see in the following equations:

$$\mathbf{rate} = \mathbf{k}[\mathbf{A}][\mathbf{B}]$$
$$\boldsymbol{When\ [B] >> [A]}$$
$$\boldsymbol{k' = k[B]}$$
$$\mathbf{rate} = \mathbf{k}'[\mathbf{A}] \quad \mathbf{(}\text{Error! Reference source not found.}\mathbf{.6)}$$

This classification of rate law is referred to as a pseudo-first order rate law.

Understanding the different types of rate laws gives us a notion of how the reactant(s) can play significant roles in influencing the rate of a chemical reaction, but this is only just the beginning – there is so much more than just reactants and products in a reaction.

1.3. Catalysts and Intermediates

During a reaction, the transformation between reactant and product goes through an intermediary state known as the transition state[16–19]. Such a state is usually short-lived[20,21], and it can either transform into the product or revert to a reactant. Transition states are an example of an intermediate caveat being transition states cannot be isolated.

There are also chemical species that serve to form low potential energy intermediates allowing a lower energy pathway to make the desired product. Such species are referred to as catalysts and they increase the rate of reaction without themselves being consumed during the reaction.

Each of these intermediates and catalysts play different roles in a reaction. Let us now move away from the mathematical equations of reaction rates, and now consider graphically the behaviour represented by change in concentration over reaction time that can be visualized relative to the reactant and product:

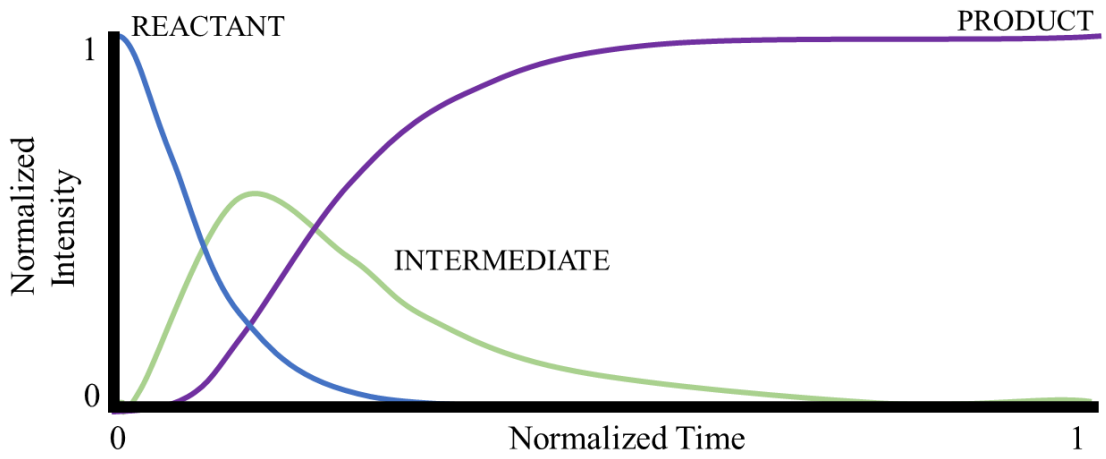


Figure 1.1: Graphical representation of the concentration change over time for Reactants, Products, and Intermediates

Today, catalysts have applications that widely span the chemical industry[22,23], from the ruthenium- based Grubbs' catalyst for production of olefins[24–26] to palladium-based tetrakistriphenylphosphine for Suzuki-Miyaura reactions to make pharmaceuticals[27–29]. These catalysts form organometallic compounds (compounds with a metal-carbon bond). Catalysts are important to organometallic chemistry as they are used extensively to perform selective transformations[22,23,30]. However, the mechanism by which they transform the reactant to a product is often not well understood. Elucidating the pathway(s) in which a reaction occurs can be beneficial to the optimization process, thereby improving the efficacy of the catalyst[23].

Organometallic catalysts used in industry are often d-block metals[23,31] whereby the valence electrons of the metal reside in the d-orbital. It is well understood that there are many varieties of organometallic complexes, but the main metals of interest for this body of work has been limited to d-block metals. The first reporting of a homogeneous transition metal catalyst[22,32] is the cobalt carbonyl complex for hydroformylation by Roelen in 1938[33]. In

1961, Heck and Breslow published their mechanistic findings for the cobalt based hydroformylation of olefins.[22,34]

1.4. Mechanistic Investigation

In trying to understand the role the cobalt and carbon monoxide (CO) played in the hydroformylation reaction, Heck and Breslow investigated the homogenous catalyst, whereby the catalyst is in the same phase as the reactant or substrate[22]. Mechanisms of such catalysts can be partially established by means of synthesizing the intermediates and understanding their behaviour through determining the product that formed from the intermediate reactions[34].

The results demonstrated that the cobalt catalyst coordinated to the olefin and was subsequently reduced by CO to form the aldehyde product[34,35]. Nevertheless, the question remains: how did they determine that the cobalt coordinated with the olefin? First, Heck and Breslow synthesized the intermediate compound – the cobalt carbonyl catalyst reacting with the olefin. Next, a change in the infrared spectrum was observed, indicating that there was a chemical change in the solution.

Infrared spectroscopy (IR) measures the vibrational modes of a chemical bond. If there are changes to the electronics of the bond, the vibrational frequency will change, and exactly the observation noted in the publication[36].

Assuming that the olefin was coordinated to the cobalt, a reductive step would eliminate the olefin with a carbonyl to yield an aldehyde; if otherwise, the reduction would not change anything. Upon addition of iodine and methanol, an aldehyde was formed,

indicating the formation of a cobalt-olefin intermediate complex. This method of investigation was extended to other speculated steps in the mechanism.

Such investigative technique is laborious and complicated as it requires the synthesis of the potential intermediate species, which sometimes can be difficult to perform. Another downside to the technique is that it uses an inductive method to interpret the chemistry of intermediate processes. For example, the formation of the cobalt-olefin complex was inferred from the change in IR spectra and the formation of the aldehyde upon reduction. But what happens during the coordination and reduction is still not known and remains difficult to determine unless there are molecular scale visualizations recording the reactions at a timescale required for one reaction event.

Although we are still not able to look precisely at a molecular reaction with the current technological advancements, our ability to make inferences about a reaction can now be done at much higher accuracy and sensitivity with modern characterization techniques used to monitor a reaction in real time. Instrumentations such as nuclear magnetic resonance (NMR)[37–43], IR[44–47] or mass spectrometry (MS)[48–53] can be used to monitor a chemical reaction in real time. By monitoring the reaction, the chemical transformation can be tracked, resulting in the ability to visualize *in-spectra* the change in intensity, and therefore, amount for each chemical species. By being able to monitor the dynamics of each species, we can perform mechanistic investigations in a much more convenient and powerful way. The results obtained are superior to the earlier method of isolating individual intermediates and inferring reaction behaviour from just the yield and identity of the product.

1.5. Mass Spectrometry

Tracking reactions in real time is helpful in understanding the kinetics and mechanistic pathways[54,55] in a given chemical reaction. Mass spectrometry is one of the methods available for characterization of chemical species. It has been used to track the change in intensity chemical species over the course of the reaction[48–53].

1.5.1. What is Mass Spectrometry

A mass spectrometer is a chemical instrument that measures the abundance of ions[56]. The initial development by JJ Thomson of the cathode ray tube led to the discovery of negatively charged particles[57], also referred to as electrons. Incidentally, when he turned his attention to understand the fundamentals of the positive counterpart, Thomson wound up constructing a precursor to the mass spectrometer, a device known as a mass spectrograph. This precursor instrument looked at ray patterns through photographic sheets; however, using photographic sheets to quantify signal was a difficult task. To combat this, an electrical device recording the intensity of the ray would be beneficial to quantify the abundance of ions as the device would have some form of voltage or current readings. Eventually, Thomson replaced the photographic sheets with a contraption known as a Faraday cup, a metal conductive cup that increases in current proportionally to the intensity of the impending ray or ion beam.

With this device, FW Aston (JJ Thomson's research assistant) worked on identifying isotopes of elements – the concept of an element having different number of neutrons was not yet well established[58,59]. In order to measure different isotopes, the instrument must be able to differentiate between masses. Using an electromagnet, the ion beam was curved,

and varying masses had different radii of curvature when exposed to the same magnetic field[36,58,60]. Thus, this presented a way to isolate the ion beam of individual masses. By varying the magnetic field, the ion beam for each mass will arrive at the Faraday cup independently, thereby making it possible to quantify the intensity for each mass[60].

Aston was able to determine the relative abundance of isotopes for an element and also the mass of each of these quantified in *m/z* (mass-to-charge ratio) units. The mass-to-charge ratio (m/z) is a ratio between the mass of the analyte divided by its charge. However, as we will see in the later section, a single mass-to-charge ratio does not always imply a single species, and this can complicate our identification of our chemical species of interest.

Understanding the origins of the mass spectrograph can assist in the understanding of the new instrumentation for the mass spectrometer. The concepts for mass spectrometry are very similar to the ideas for a mass spectrograph. Each mass spectrum plots the intensity against the mass-to-charge ratio (*m/z*). The intensity is defined as the abundance of ions in a scan[60]. Mass spectrometers possess three main components: the ionization source, the mass analyzer, and the detector.

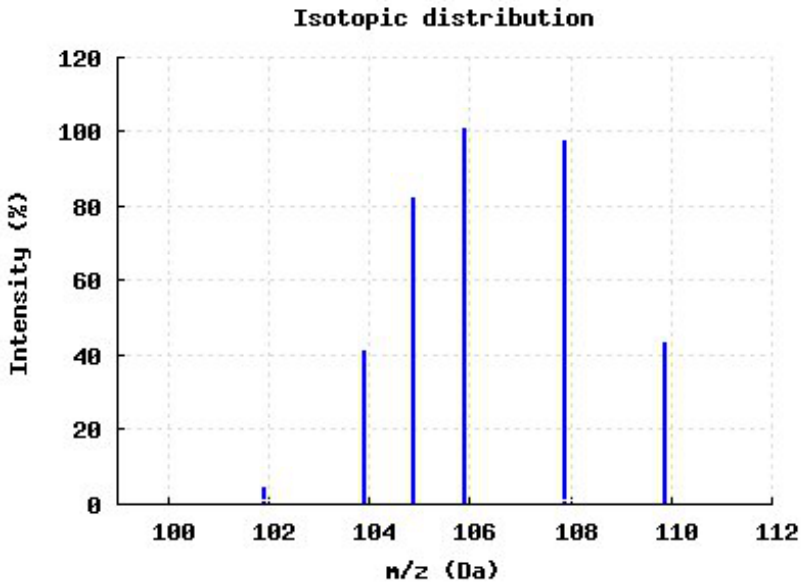


Figure 1.2: Simulated mass spectrum of palladium

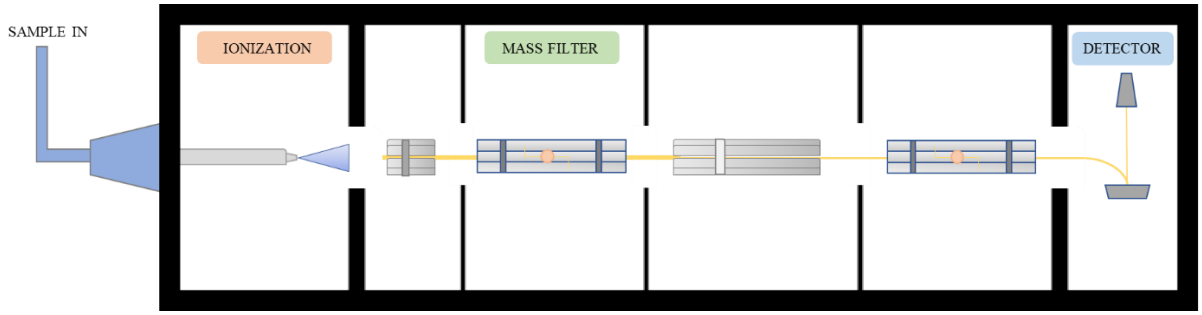


Figure 1.3: Mass spectrometer (triple quadrupole) schematic outlining the steps of analysis in MS: Ionization, Mass Filtering, Detection

To begin our analysis on the mass spectrometer, we must first introduce the chemical species of interest into the instrument. We do this at the ionization source. Typically, the chemicals or analytes are prepared in solutions as samples (an unknown solution containing species of interest) or as standards (a solution containing chemical of interest at known concentration). There are different ways in which an analyte can be detected in the spectrometer, but there are, foremost, two requirements for the analyte to be detected: it must (i) be in gas phase and (ii) be charged[60,61]. Most sample preparations are usually in

solution, or in liquid phase, and they must be vaporized to the gas phase. Once the analyte is in gas phase and is charged, it travels inside the mass spectrometer in the form of an ion beam. This transformation along with some other ionization methods will be addressed in a later section.

Following the ionization source, the ion beam moves through the second major component: the mass filter. The mass filter takes the ions in gas phase and, by using magnetic or electric fields, separate them[60,62]. Thus, this step isolates the beam to contain only one *m/z*. In an ideal world, that would mean there is only one species in the beam; but more often than not, such is not the case. As we have already defined, the mass-to-charge ratio (*m/z*) is a dimensionless ratio between the mass of the analyte divided by the charge of the analyte. A compound with double the mass and double the charge will share the same m/z. In addition, there can also be contaminants within the sample with the same approximate mass and charge, and this can result in them being identified as the target analyte as well[63,64]. However, there are methods to resolve these issues and they will be addressed in a subsequent section

1.6. High-Resolution Mass Spectrometry

Contaminants within a sample with a similar mass and charge to the chemical species of interest are difficult to discern –they would be counted (appear in the same peak) as the same species as the target analyte in a mass spectrum. One of the ways to counter this issue is to push for higher resolution in instruments. A molar mass of a target molecule can coincide with contaminants, but this is usually only at the nominal mass (mass before the decimal point)[65]. By increasing the resolution, this allows the molecule's mass to be

discerned at a higher decimal point i.e. up to the third or fourth decimal point. At this level of precision, we can eventually separate the contaminant peak from that of the target analyte.

The Ion Cyclotron is a one of the high-resolution methods used in mass spectrometry[66,67]. In principle, it is a non-destructive method with respect to the ion (where the analyte ion is maintained); however, in its operation it is still destructive due to the ionization method used[60]. The ion cyclotron works by means of oscillating the ion beam in a cyclic fashion – this is known as its cyclotron resonance[68,69]. Upon exciting the cyclic beam with a radio frequency (RF), the ion beam widens[70]. The widening of the beam carries the same concept as the electromagnet for JJ Thomson's mass spectrograph where the ion beam was diverging upon the activation of the electromagnet.

In application, the ion cyclotron resonance (ICR) specific RF is used to excite a specific mass which increases its radius of cyclic oscillation. When the ion is in close proximity to the electrode, it causes the metal electrodes to build up charge which results in a current, this current is known as an image current, being detected as a signal decay[60,71]. The decay can be transformed through Fourier transform (FT), resulting in the modern incarnation of the instrument we now know as FT-ICR.

One of the major flaws of the FT-ICR is that it requires large magnets to maintain the magnetic field[72]. In 1999, Makarov published the patent on a new detector ion cell, now called the Orbitrap[73]. The Orbitrap is ion trap device that traps ions in oscillation motion in a quadrologarithmic field[74]. There are two main components in the orbitrap cell, a barrel electrode and a spindle electrode[75,76]. The quadrologarithmic field is produced by creating

a DC potential between the two electrodes, V_{cell}. V_{cell} needs to be at a higher potential than the potential, V, exerted by an incoming ion at velocity *v*. The relationship between *v* and V for the incoming ion into the field is represented by:

$$\mathbf{qV = \frac{1}{2}mv^2\left(\frac{R}{r}\right)} \quad \textbf{(1.7)}$$

R is the radius of the barrel electrode, r is the radius of the spindle electrode, m is the mass of the ion, and q is the charge of the ion. When the ion enters this field is forced into an axial oscillation around the spindle due to its cylindrical geometry. To prevent the ions from colliding with the endcaps containing the trap cell, the endcaps exert a repelling potential. The ion detection step for orbitrap is done by measuring the axial oscillation of an ion in the cell. The axial motion is dependent on the mass and charge of the incoming ion and this is reflected in its frequency in oscillation. Thus when the ion packet oscillates in the cell, the image current can be measured and the result will be a time-domain transient[77]. The transient is a waveform that has periodic oscillations in intensity whose intensity will decay after each oscillation, it is also known as a free induction decay[78,79]. This decay can then converted to frequency domain by FT to obtain the mass spectrum from the unique axial oscillations which are correlated to the mass to charge ratio of the ion.

1.7. Types of Mass Filters

In general, the purpose of mass filters is to isolate the m/z in the incident ion beam. The detection of ion intensity in the resulting beam will then be representative of the isolated mass' ion count

1.7.1. Quadrupole Mass Filter

The quadrupole mass filter is a linear filter containing four cylindrical rods in a square diamond formation[60] (top-bottom and left-right) as illustrated in the figure below. The rods on opposing poles are held at the same potentials which hold a pair of rods at a negative voltage and the other pair at a positive voltage. Depending on the polarity of the ion, it will be attracted to a pair of rods and repelled by the other pair[80,81]. Holding the two pairs of cylindrical rods at a constant potential will result in the neutralization of the ion beam from collision-induced discharge at a rod. To contain the ion in quadrupole while moving towards the detector, the polarity must be oscillated at a fast rate to maintain the quadrupole motion. Therefore, the potential on the two pairs of rods are oscillated by an RF voltage at MHz frequencies[60,82]. This oscillation will induce the ion beam to be attracted to a pair of rods in the quadrupole at one moment and repelled by a pair at another moment.

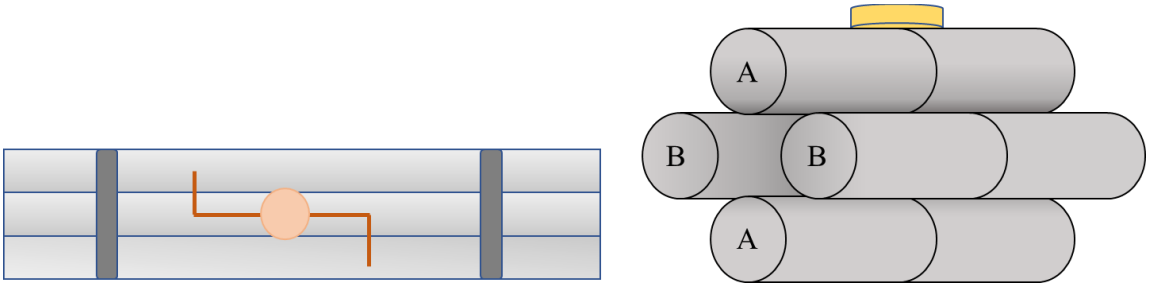


Figure 1.4: Side view and 3D rendition of a quadrupole mass filter

The ion motion in a quadrupole is shown below. The ion beam is only stable when two dictating factors are held at a specific ratio. These are (i) the DC voltage and (ii) the RF voltage on the quadrupole[83,84]. To determine the ratios, a plot of DC voltage versus RF voltage will demonstrate the possible regions of stability, as indicated by the numbered points in figure below

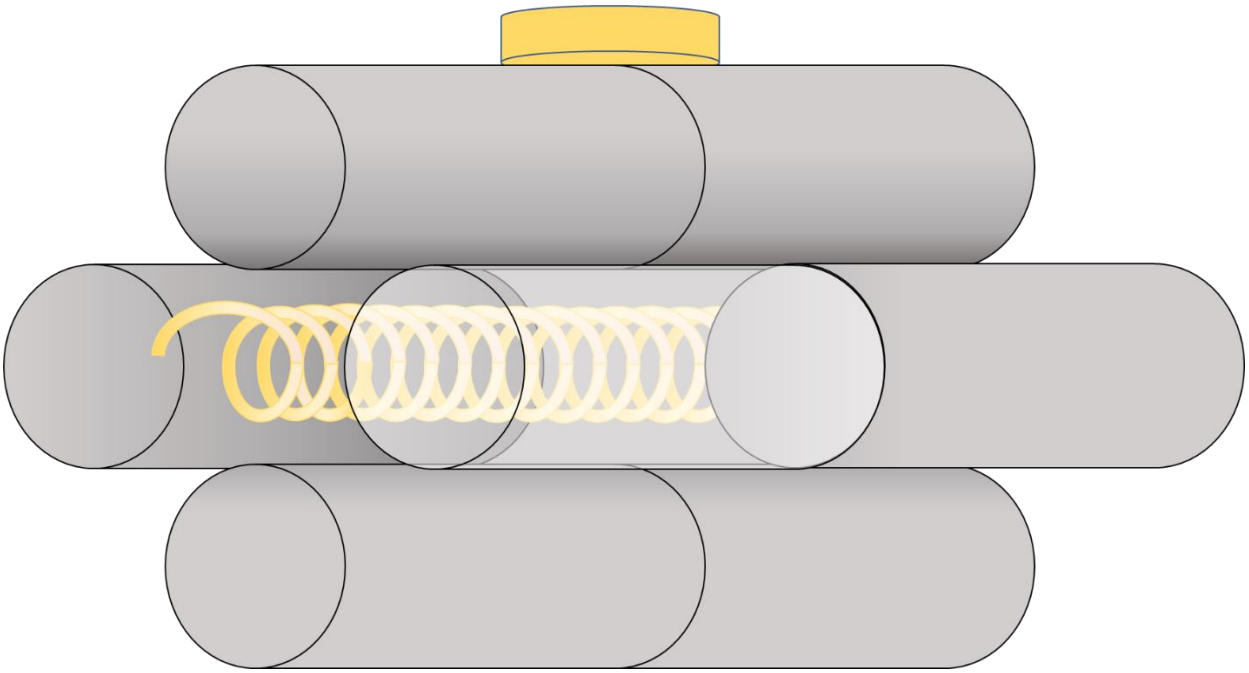


Figure 1.5: Visualization of a stable ion motion/trajectory in a quadrupole mass filter

Looking at the mass stability, shown as M1, M2 and M3, these values represent three arbitrary *m/z* that are to be isolated in the mass filter. The region specified to the variable show the potentials at which the ions of the isolated m/z will be stable. When the instrument scans through the masses, the line in which a potential ramp follows must intersect with these stable regions as demonstrated in figure above.

Quadrupoles are great mass filters when scanning through *m/z* is a necessity for the instrument. Other advantages of quadrupoles include high ion transmission, low cost of production, and low voltages needed for ion acceleration[60] . Understanding the principles of a quadrupole will be critical in our understanding of other mass filters.

1.8. Tandem Mass Spectrometry

Tandem mass spectrometry is a technique in which the instrument detects fragments of the analyte ion to identify specific compounds in a complex mixture. To obtain the analyte ion,

the compound of interest is first ionized and isolated. The ions are then subject to decomposition by high energy bombardments to facilitate fragmentation[85].

Tandem in space [86] mass spectrometry separates the tandem experiment into two separate spaces. The first region contains the analyte ion prior to fragmentation, and this is also known as a precursor ion. The second region contains the resulting fragment ion, otherwise known as the product ion[87]. This form of experiment is used in triple quadrupole mass spectrometers shown in the figure below.

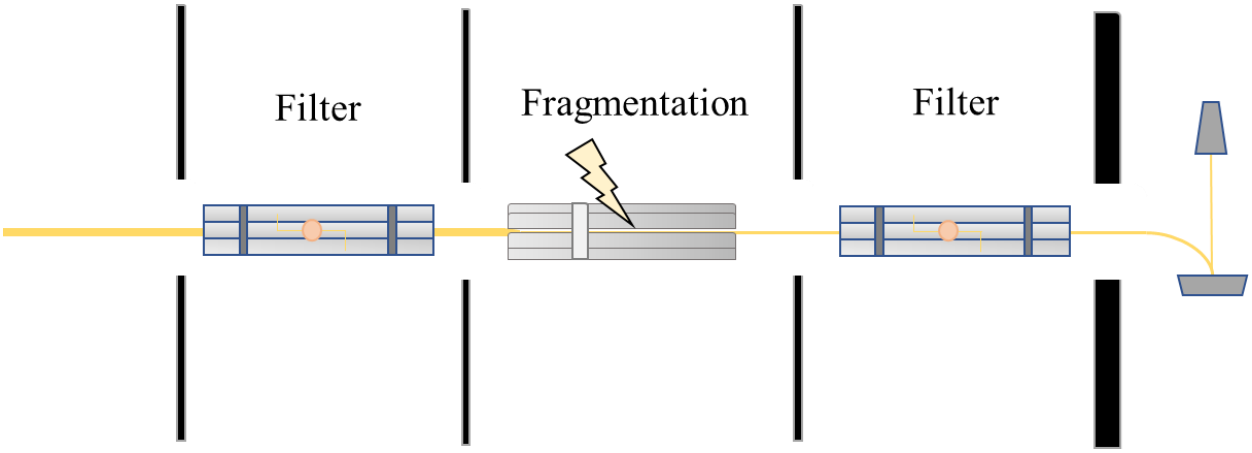


Figure 1.6: Diagram of tandem in space where filtering pre and post fragmentation happens in two locationsIonization Methods

Now that we have covered the various detectors and mass filters used in mass spectrometry, we now need to address an important question: How do the molecules ionize? In this section, ionization methods will be discussed in greater detail.

There are many methods of ionization used in mass spectrometry, however, for the purposes of this thesis, I will be focusing electrospray ionization.

1.8.1. Electrospray Ionization

Electrospray ionization (ESI) transfers preformed ions from solution into the gas phase[60,88,89]. The first step in ESI is done by imparting a high potential environment to the liquid sample, resulting in a spray, known as a Taylor cone[90], of fine droplets at the tip of an ESI capillary. The high potential generates an excess of positive ions (negative in the negative ion mode) through oxidation (reduction processes), usually of the solvent or of the capillary itself. The excess charge is on the surface of the droplets, and as the droplets are desolvated, the surface ions experience enough repulsion to be expelled into the gas phase in a process known as ion evaporation[91–93]. The process in an instrumental perspective is outlined in the figure below

Figure 1.7: Diagram of ESI Probe

1.9. Automation of Reaction Monitoring

Reaction monitoring through mass spectrometry by means of PSI-ESI-MS reveal information that are hidden by conventional techniques of NMR or IR. Mass spectrometry, having the ability to diagnose the mass to charge ratio, allow for the opportunity to extract unique ion information and interrogate the changes occurring for associated ion. Tracking an ion series, a series of ions with the same *m/z* or *m/z* transitions, reveal kinetic behaviour for the interrogated species. These kinetic behaviours are vital to the understanding of reaction mechanisms, see which species are associated with reactant, products, or intermediates.

The standard protocol in the McIndoe lab for reaction monitoring requires the use of PSI-ESI-MS acquisitions starting from baseline of a reaction mixture containing only solvent and reagents without the onset of reaction. The baseline gives an idea on what is present before the induction of the reaction. Monitoring the changes relative to the baseline offer a simplistic investigation methodology. To trigger the reaction, often catalyst is added to the reaction mixture and the reaction progress is recorded through the mass spectrometer. Setting up the reaction mixture for baseline and the subsequent reaction monitoring require typically 2 hours plus reaction time based on personal experience. This is hindering for rapid analyses of reactions or even performing reproducibility studies for a reaction. The time consuming and sensitive nature of mass spectrometry-based reaction monitoring led to the reporting of chemical observations based on an experimental set of a single acquisition.

From this experience, the design of an automated reaction monitoring platform is conceived and includes both software and hardware solutions to address this project.

Automating reaction monitoring, or even further chemical kinetics, opens new opportunities in chemical automation filling the void between high throughput screening, exploratory synthetic chemistry, and process chemistry.

Throughout this text, each chapter addresses the manifestation of mass spectrometry based reaction monitoring, from Chapter 2 describing the processes of manual acquisitions to the subsequent chapters outlining development of the automation platform through robotics, classification models supplemented by a software infrastructure sustaining the data acquisition and process workflow in reaction monitoring.

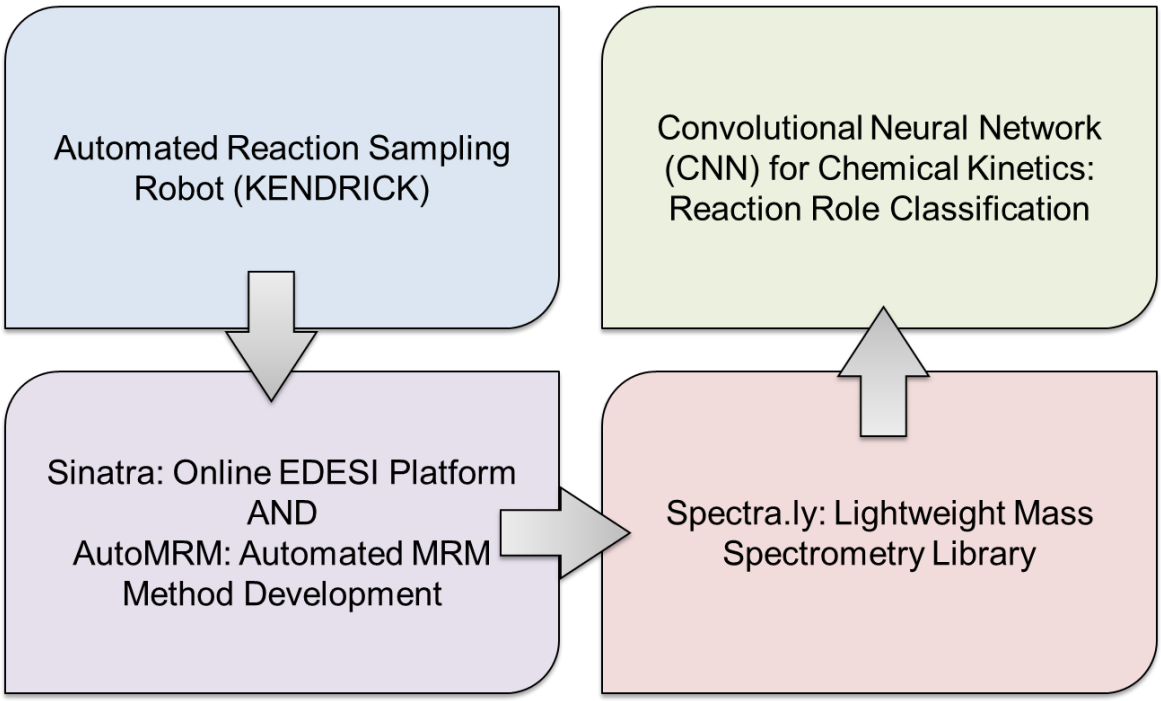


Figure 1.8: Graphical demonstration of the automation workflow for automating reaction monitoring for mass spectrometry

In Figure 1.8, the course of the automation workflow starts with hardware robotic automation in a project named KENDRICK. Robotics opens access to rapid analyses and reproducibility studies, addressing the downfalls of manual reaction monitoring. To sustain this automation workflow, the infrastructure contains three subprojects: Sinatra,

AutoMRM, and Spectra.ly. In essence, Sinatra is a software using user control automation and data processing to create a platform to perform ion characterization through tandem mass spectrometric techniques. AutoMRM takes the characterization information from Sinatra and extract key features of the compound to create Waters compliant Multiple Reaction Monitoring methods. Spectra.ly is a mass spectrometry data processing library that allows for quick and simple creation of processing method that is applicable for batch processing, packaging the data for subsequent data analyses. The Sinatra, AutoMRM, and Spectra.ly infrastructure takes data collected from the mass spectrometer for a reaction sample for creating targeted methods (using Sinatra and AutoMRM) and packaging the collected data from the experimental acquisition (using Spectra.ly) for further data analyses.

Further data analyses for this workflow involves passing the data through a reaction role classifier. Other statistical methods can be used to analyze the packaged information from Spectra.ly, however, with interest in chemical kinetics for this project, reaction role classification was an appropriate method. Taking the packaged data, the classifier can iterate through the *m/z* and identify its role in a reaction based on the kinetic information. Incorporating this to the automation workflow shown in Figure 1.8, it is possible to rapidly assess a large set of reactions with multiple replicates (through KENDRICK), create reaction tracking methods on the fly (through Sinatra, AutoMRM), package the resulting data from the experiment (through Spectra.ly), and output the reaction roles of species of interest (through Reaction Role Classification). This is powerful for a kinetic chemist as they can submit reaction samples to the automation workflow, wait for the results, and focus on interpreting the results to either create a new hypothesis for a reaction mechanism

or provide the data to support a proposed mechanism. A powerful technology as such simplifies the experimental aspects for chemical kinetics.

2. CHAPTER 2: TITANOCENE OXIDATION AND THE USE OF REAL TIME MASS SPECTROMETRY WITH COLOURIMETRIC TECHNIQUES

Results reproduced with permission from Yeung, D.; Penafiel, J.; Zijlstra, Harmen S.; McIndoe, J. Scott. “Oxidation of Titanocene(III): The Deceptive Simplicity of a Color Change” Inorg. Chem., 2018, 57, 457-461.

2.1. Real Time Mass Spectrometry

Tracking a reaction can be done in a continuous infusion fashion; in a nutshell, this means introducing the reaction sample to an instrument as the reaction occurs. Unlike the previous methods of reaction monitoring, this technique is advantageous in that it no longer requires the wait time in between injection times and column separation. The limiting factor now is lies in the instrument’s sampling time (i.e. the time interval required for signal detection). Using mass spectrometry in such a manner is known as direct infusion or real time mass spectrometry.

The only prerequisite for real time mass spectrometry is a method to introduce the sample to the instrument continuously. Pumps such as peristaltic pumps, stepper motor pump, or syringe pumps can continuously introduce liquid samples for real time monitoring. Solvent compatibility with the pumps, as well as the inability to add reagents without perturbing the infusion is an issue, particularly for air sensitive reactions.

How can air sensitive reactions be analyzed without sample preparations for reaction monitoring? In the McIndoe group, a continuous infusion method[94–101] known as pressurized sample infusion electrospray ionization MS (PSI-ESI-MS) is utilized. Much like a cannula transfer, an inert gas overpressure is exerted to a Schlenk flask to force the

reaction sample to exit the vessel through a polyether ether ketone (PEEK) tubing that is connected directly to the mass spectrometer. This form of infusion is a viable and extremely accessible method to analyze air sensitive reactions[96].

2.2. The Oxidation of Titanocene

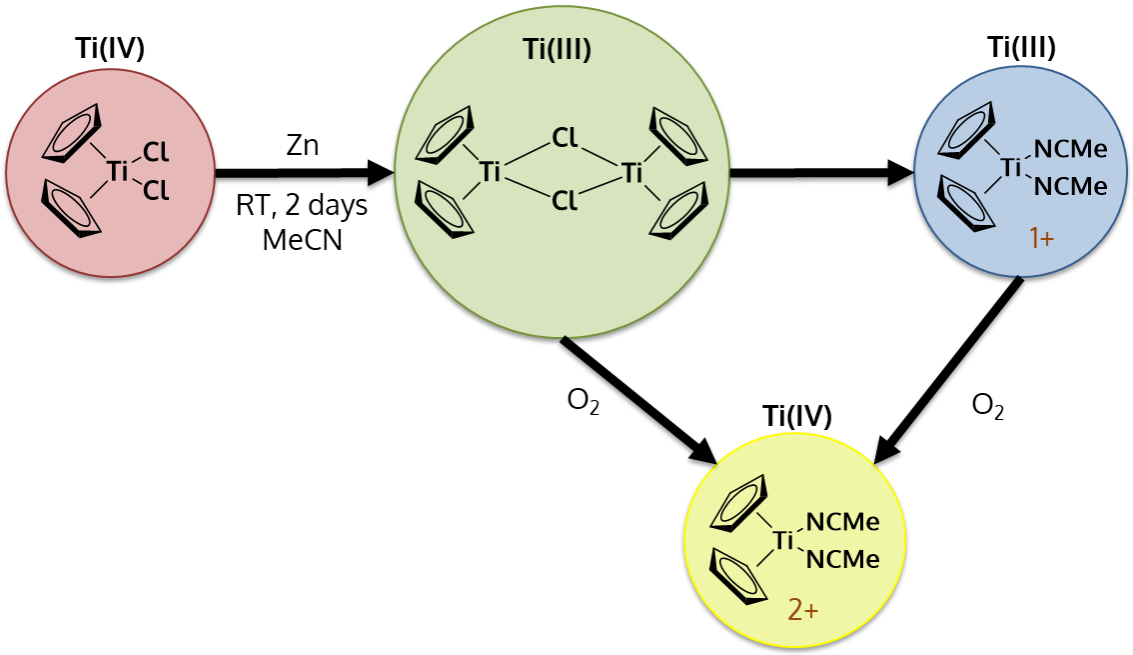
One of such air sensitive reaction is the oxidation of titanocene, which is commonly used as a glovebox indicator[102,103]. Titanocene is an organotitanium complex with two cyclopentadiene ligand bound in a η^5 conformation – the titanium is coordinated to all five carbon atoms in the cyclopentadiene ring.

A common titanocene reagent is bis(η^5-cyclopentadiene) titanium (IV) dichloride, which for the purpose of simplicity, will now be abbreviated to $Cp_2Ti(IV)Cl_2$. When $Cp_2Ti(IV)Cl_2$ is reduced with zinc, manganese, or magnesium, $Cp_2Ti(IV)Cl_2$ is converted to $Cp_2Ti(III)Cl$[104–106] The converted titanium(III) complex is reduced a radical, with the titanium bearing the lone electron[107,108]. This transition metal radical is useful for applications such as photoinitiators for light-based curing processes and also industrial syntheses[109,110] such as radical epoxide ring opening[108,109,111–118].

In the absence of coordinating solvents, $Cp_2Ti(III)Cl$ will form a dimer as $Cp_4Ti(III)(\mu\text{-}Cl)_2$[102,119]. However, in solvents such as tetrahydrofuran (THF) or acetonitrile (MeCN), solvent coordination to the metal center assists in stabilizing the complex by means of an electron count change from 14 electrons to 16 electrons. Through empirical data, a complex with 18 electrons is a stable complex and this heuristic is known as the 18 electron rule. Having 16 electrons in the titanium complex is more stable than having 14 electrons because it is closer in value to the 18 electrons as dictated by the 18 electron

rule. Therefore, any two-electron donor solvent would undergo such a process as described above[120,121].

In 1994, Burgmayer proposed the use of titanocene in MeCN as an indicator for a compromised inert atmosphere. He presented a scheme for the reduction of a titanocene(IV) dichloride to obtain the $Cp_4Ti(III)(\mu\text{-}Cl)_2$ dimer. This subsequently then dissociates in MeCN to become $Cp_2Ti(III)(MeCN)_2$, which is blue in color, as shown in the figure below. The blue $Cp_2Ti(III)(MeCN)_2$ changes in color to yellow upon compromise of the inert atmosphere. The mechanism for titanocene in the reduction step[115,122] or for radical reactions[111,112,123] has been well documented but the oxidation step for titanocene is not well studied. As such, this presented a great opportunity for the utilization of techniques such as PSI-ESI-MS to track the changes in species during the reaction upon oxidation. My work in this endeavour aims to elucidate and provide insight to the mechanistic behaviours of the oxidation of titanocene as shown in the figure below.



Scheme 2.1: Reaction scheme for the oxidation of titanocene (Red: Titanocene(IV) chloride, Green: Titanocene(III) chloride dimer, Blue: Titanocene(III) bisacetonitrile, Yellow: Titanocene(IV) bisacetonitrile)

2.3. Addition of Orthogonal Techniques: Colorimetry

The dramatic color change from blue to yellow in titanocene oxidation opens up another avenue for analysis of the reaction through light in the visible spectrum. This can provide secondary information to the reaction. Pairing the mass spectrometry results with changes in the visible spectrum can offer insight on rates[124–128] and more importantly, pairing the timing information of the two techniques. Between the two analyses, monitoring the correlation of events can potentially reveal the cause of the color change.

I developed a python-based program, ColorPixel, for the purpose of performing colorimetry during the reaction. First, a smartphone is used to record a video of the reaction, steadied by a tripod much like the following figure.

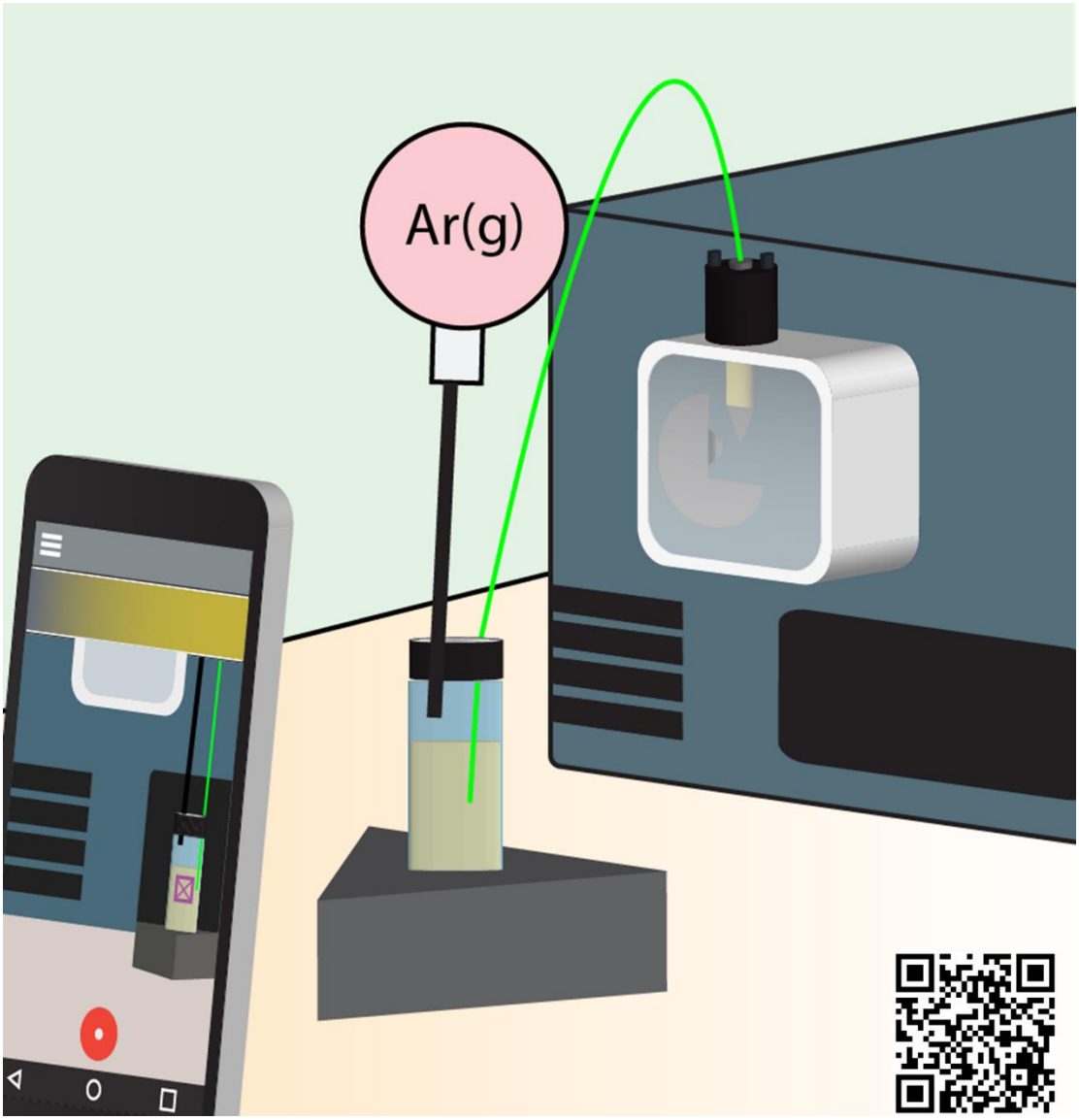


Figure 2.1: ColorPixel setup to acquire colorimetry data

Time-lapse photography will also suffice as long as the photos are stitched in post-production into a video format like MPEG4. Key is the frames or images must be steady. ColorPixel has the user choose a pixel to track the color changes and plot the color change over time as a color bar consistent with the timing information of the recording.

This technique is used in conjunction with PSI-ESI-MS in the experiments for this chapter and will be displayed with the MS reaction chronogram as shown below.

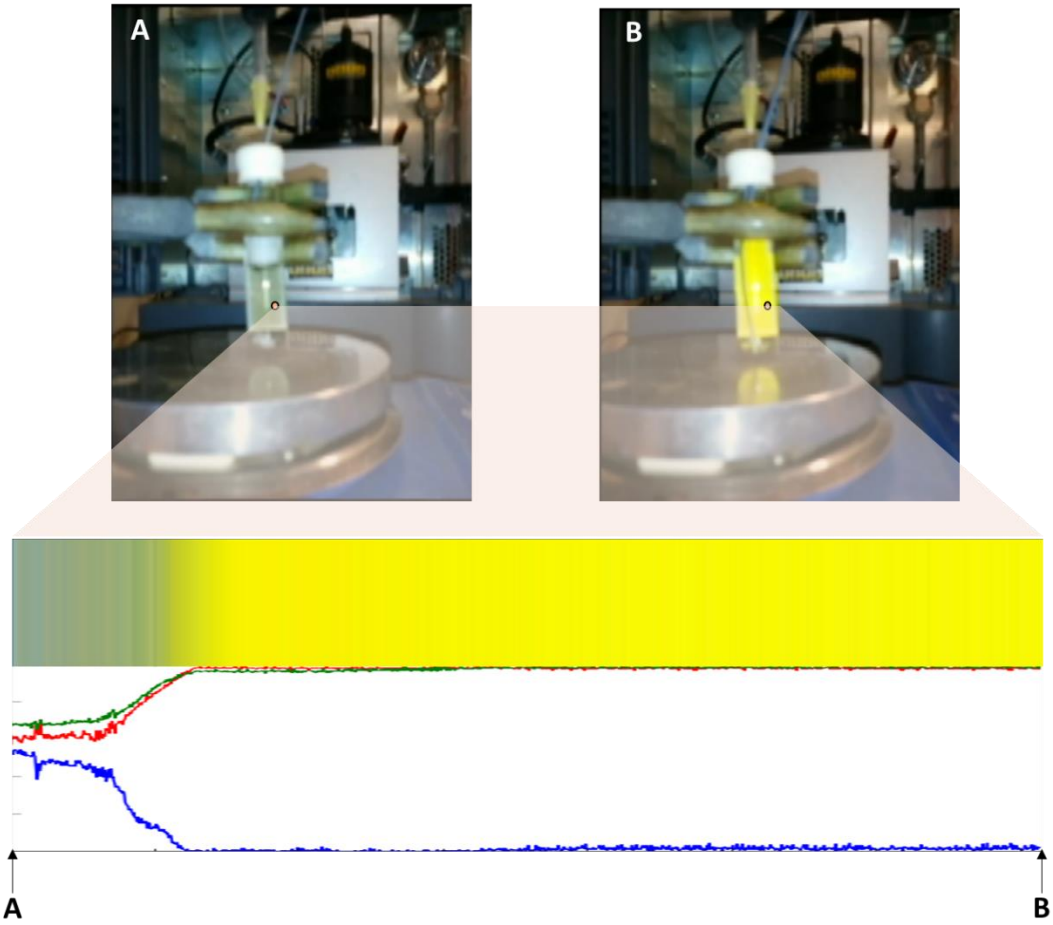


Figure 2.2: The methodology concept for constructing the colorbar through ColorPixel (A: Pre-oxidation and B: Post-oxidation)

2.4. Results and Discussion

2.4.1. Oxidation Monitoring by PSI-ESI-MS

Studying the process of titanocene oxidation is an inherently sensitive reaction, as any contamination in the inert atmosphere of the sample will compromise the analysis[99–101]. To maintain an oxygen-free environment, the reaction sample was contained in a Schlenk

flask sealed with a rubber septum. The PEEK tubing for the cannula transfer to the mass spectrometer was pierced through a septum, as depicted in the figure below.

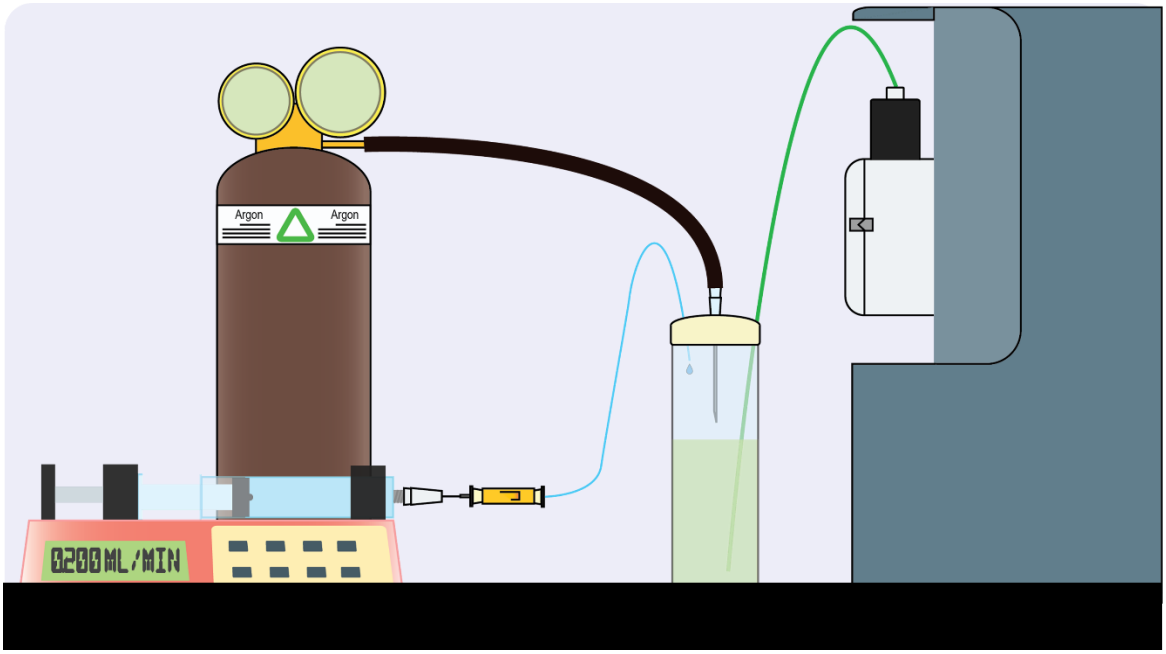


Figure 2.3: Experimental setup of PSI-ESI-MS

By continuously infusing the oxidation starting material, $Cp_2Ti(III)(MeCN)_2$, to the mass spectrometer, it was observed that the expected *m/z* 260 peak was steady in the chronogram. However, upon addition of air into the sample, the titanocene oxidation reaction occurred so quickly and let to inconclusive data due to the formation of neutrals and solid which could not be detected by ESI. This resulted in the loss of signal in the mass spectrometer. If direct introduction of air into the reaction vessel was too reactive to get any viable information, then the oxidation process must be slowed down to circumvent this problem.

A slow infusion of O_2 saturated solution of MeCN was introduced to the reaction sample, thus inducing the oxidation in a controlled manner. In doing so, the oxidation reaction traces were observed. The figure shown below demonstrates the speciation changes that occurred during the oxidation process.

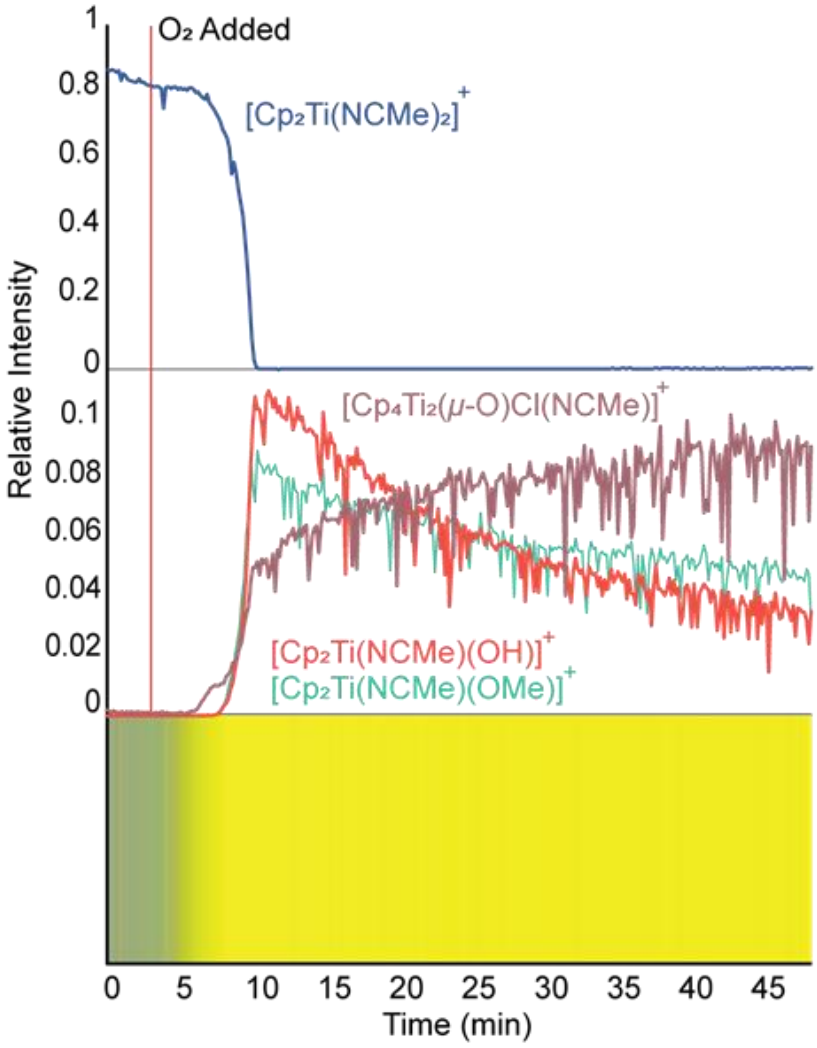


Figure 2.4: Chronogram of starting materials and products with color change profile displaying the oxidation of titanocene. (Top: Starting materials, Middle: Products, Bottom: Colorbar from ColorPixel

Based on the changes in Figure 2.4, there is a slight stagger between the mass spectrometry data and colorimetry data. The stagger can come from two possible factors: 1) there is a lag time between the reaction change to the detection of the change in the mass spectrometer and 2) the color change may come from a neutral reaction which cannot be observed by mass spectrometry.

Upon the addition of oxygen to the system, an immediate decrease in the oxidation starting material intensity was observed. Alongside that, there was an increase in signal of many other peaks in the mass spectrum. The figure only outlined three significant peaks, intensity was highest in post-oxidation spectrum or highest of products or intermediates, but there were many peaks present from the oxidation as shown in the following figure.

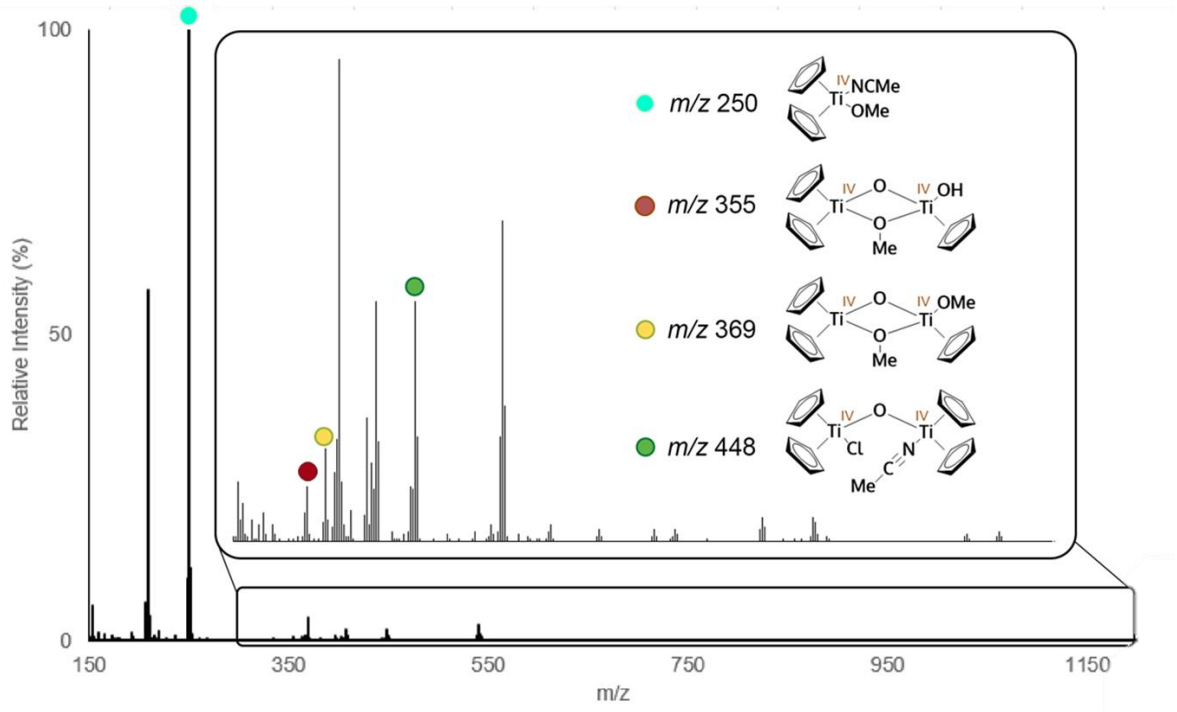


Figure 2.5: High resolution mass spectrum of the products post-oxidation outlining four major species

2.4.2. Titanium species in oxidation

The onset of oxidation generated many peaks in the mass spectrum. By assessing if the peak carries an isotope pattern, it can be determined if the peak represents a complex containing titanium. Iterating through the spectrum, it was discovered that there were, at least, twenty-three different types of titanium species present post-oxidation. Many of these species were coordinated with oxygen-containing ligands. Attempts to identify these peaks were unsuccessful using unit mass resolution mass spectrometers and tandem mass spectrometry. To increase the resolution, a high-resolution mass spectrometry instrument known as the Orbitrap was used to resolve and identify the multiple unknown species using high-accuracy m/z values. The resulting spectrum is shown below, and many unexpected possibilities for chemical formulae was elucidated.

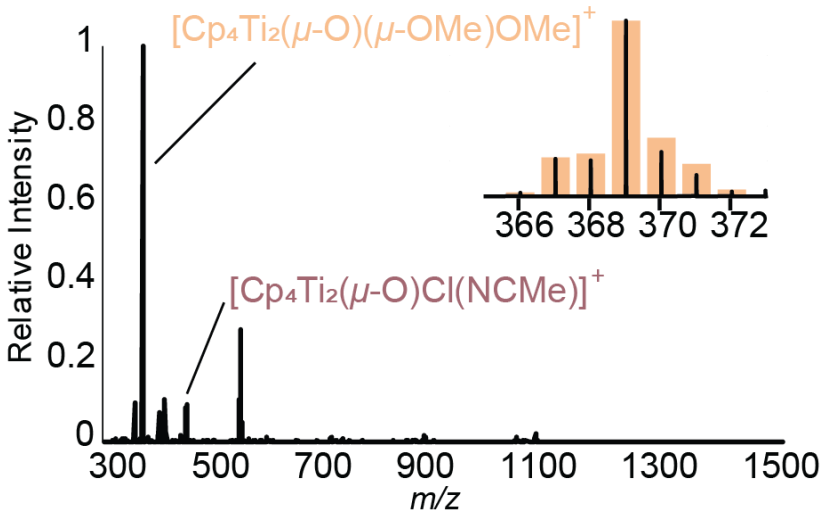


Figure 2.6: High-resolution mass spectrum for the confirmation of an oxo-titanium complex

Typically, instruments like the Orbitrap have a degree of accuracy high enough for a match to a single chemical formula[129]. Through this process of species identification, some peaks were identified; present were titanium complexes that had bridging oxygen(s), which agrees with the behaviour of the titanium oxidation[130–137], but there were also hydroxy and methoxy ligands coordinated to the titanium center(s). It was still unclear if the resulting species from the compromised inert atmosphere was caused by the oxygen in the atmosphere or by water, thus a follow-up study looking at oxidation versus hydrolysis was performed to better understand the decomposition of the titanocene indicator.

2.4.3. Oxidation versus Hydrolysis: An Investigation

To elucidate the source of the oxygen-containing titanium complexes, or oxo-titanium species, pre- and post-oxidation analyses were performed. These serve to help us figure out what is in the sample before the oxidation, and what reaction(s) occur in between to obtain a post-oxidation spectrum with not just one, but twenty-three different oxo-

titanium species. The presence of both bridging oxy and hydroxy ligands makes it exceptionally challenging to discern whether it is a combined decomposition reaction from two sources namely, oxygen and water, or a decomposition from a single source.

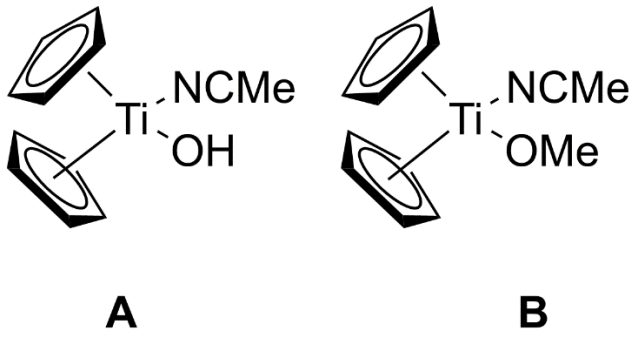


Figure 2.7: Products for the hydrolysis and oxidation of titanocene (A: Product at m/z 236 of hydrolysis $Cp_2Ti(IV)(MeCN)(OH)$ and B: Product of oxidation $Cp_2Ti(IV)(MeCN)(OMe)$ at m/z 250)

In order to isolate the two types of complexes, analyses of both oxidation versus hydrolysis of the reaction was necessitated. At the start of the infusion to the mass spectrometer, a steady *m/*z 260 signal was observed in the mass spectrum. The absence of oxidation peaks was noted. Introducing 0.1 equivalents (eq) of water in MeCN to the reaction solution, there was a small decrease in starting material in the chronogram; however, the expected hydroxy ligand was not present, only the *m/*z 260 was present.

At concentrations of 125 eq of water, the first formations of the expected hydroxy ligands were detected in the chronogram. However, relative to the most intense peak in the spectrum, *m/*z 260, the hydroxy complex was only present with a 0.03% abundance. Additionally, a significant loss of signal for *m/z* 260 occurred during the 125 eq hydrolysis. It was postulated that the loss of signal of *m/*z 260 and the resulting absence

of product peaks in the hydrolysis spectrum was due to the formation of neutral decomposition products. This result is consistent throughout the hydrolysis reaction, up until the hydrolysis reaction solution was sparged with oxygen to induce a marked increase in the expected hydroxy species in the chronogram shown below.

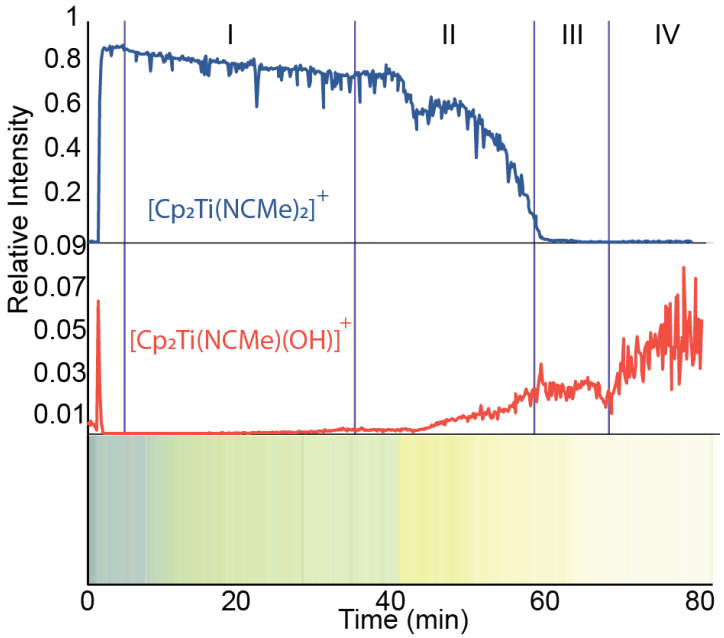


Figure 2.8: Chronogram for the hydrolysis of titanocene

From the hydrolysis experiments, it was demonstrated that the presence of hydroxy species, seen as high intensity peaks in the oxidation spectra, are not likely caused or formed by a reaction with water, but much rather by another mechanism through the involvement of oxygen. The loss of signal at *m/z* 260 can be indicative of either the formation of neutral species, or a change in electrospray ionization efficiency. The latter could be due to the addition of water to the reaction sample, thereby altering the polarity of the overall solvent composition, and leading to the change in electrospray ionization efficiency. The change in ionization efficiency is sufficient to cause a loss of signal[95];

however, further studies must be performed to properly determine the cause behind the decline in signal. In summary, hydrolysis plays a relatively significant, but minor role in the decomposition pathway in the larger scheme of the overall titanocene oxidation. From our research question of impact on titanocene in oxidation versus hydrolysis, we observed that hydrolysis plays minor role in decomposition thus we can infer from this that oxidation has the major role to forming the oxo-titanium species.

Having a better understanding of the dominant reaction in the decomposition of titanocene, there are other mysteries that are of interest other than the oxy and hydroxy titanocenes. Methoxy titanocene species were also observed in the spectrum, how are they formed during the decomposition?

2.4.4. Methyl Abstraction and deuterium labelled study

The presence of methoxy titanium species presents an interesting conundrum to resolve. There are two probabilities: Either (1) the titanocene solution was contaminated with methanol and produced a cationic methoxy titanium species[122], or (2) there was a complex methyl abstraction pathway to acquire a methyl group bound with oxygen that is coordinated to the titanium. Prior literature on methanol reactions with $Cp_2Ti(IV)Cl_2$ has shown that the complex reacts to produce $Cp_2Ti(IV)(OMe)_2$ or a cationic $[Cp_2Ti(IV)(OMe)]^{+}$[122].

In a problem with two probable answers, the principle of Occam's razor states that the solution with the less speculation is usually sounder. In other words, the simplest solution is more likely to be correct. From the two possible explanations that we have above, methanol contamination is the easier of the two to explain; however, it is also the easier

option to disprove. Reacting $Cp_2Ti(IV)Cl_2$ in acetonitrile (MeCN) with methanol will create the observed $Cp_2Ti(IV)(MeCN)(OMe)$; however, this species was not observed prior to oxidation. The results indicate that the trace methanol levels in the solvents are not sufficient to be a major peak in the mass spectrum.

This leaves the other explanation that a methyl abstraction event occurred to produce the target $Cp_2Ti(IV)(MeCN)(OMe)$ compounds. When reviewing the reaction, the only methyl source was from MeCN. Methyl abstraction is possible in the perspective SN_2 nucleophilic attack from the oxygen. However, before we rush to any conclusions, it is important to consider another factor: $Cp_2Ti(III)(MeCN)_2$ is a radical. Even though this aspect of the molecule needs to be verified by Electron Paramagnetic Resonance, prior literature on $Cp_2Ti(III)$ species indicate that it is a radical species[107,108,111,112,123]. The radical mechanism involves the termination of the titanium radical when reacted with one of the radicals on elemental oxygen. This results in a radical on the terminal end of the oxygen molecule. The other oxygen repeats the termination step with another radical titanium resulting a di-titanium species with two bridging oxygen atoms. The oxygen bridge then undergo radical initiation to form two radical oxo-titanium species with the oxygen carrying the radical electron. The oxo-titanium radical then attacks the methyl group on MeCN, resulting in a radical CN group and a methoxy titanium complex, as shown in the figure below.

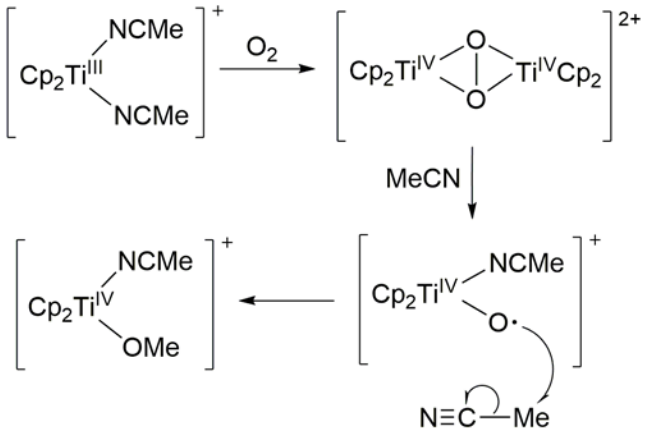


Figure 2.9: Proposed radical mechanism for the generation of methoxy titanocene species

Regardless of the mechanism, it must first be determined if the methyl abstraction is the explanation by which the methoxy group was formed. A deuterium-labelled study was designed to determine the source of the initial methyl group. This was accomplished by performing all previous procedures for the titanocene oxidation in deuterated MeCN (d_3-MeCN). This included the synthesis of the oxidation starting material $Cp_2Ti(III)(MeCN\text{-}d_3)_2$.

Prior to oxidation, the new oxidation starting material $Cp_2Ti(III)(MeCN\text{-}d_3)_2$ was identified at *m/*z 263 and this peak held steady in the chronogram. Upon addition of oxygen-saturated MeCN to the deuterated reaction sample, the formation of the $Cp_2Ti(III)(MeCN\text{-}d_3)(MeCN)$ occurred first, followed by the completed replacement of MeCN- d_3 with protonated MeCN. After complete ligand replacement, the methoxy formation began. In the two methoxy traces shown in the reaction chronogram below, one had three deuterium atoms and the other was completely protonated. Methyl groups will be triply deuterated thus indicating a methyl group on the methoxy titanium complex

is deuterated, but it could be on either the MeCN or on the methoxy. Upon fragmentation, a peak presenting (OMe-d_3) was present thus indicating that the methyl group on the methoxy was indeed abstracted from MeCN.

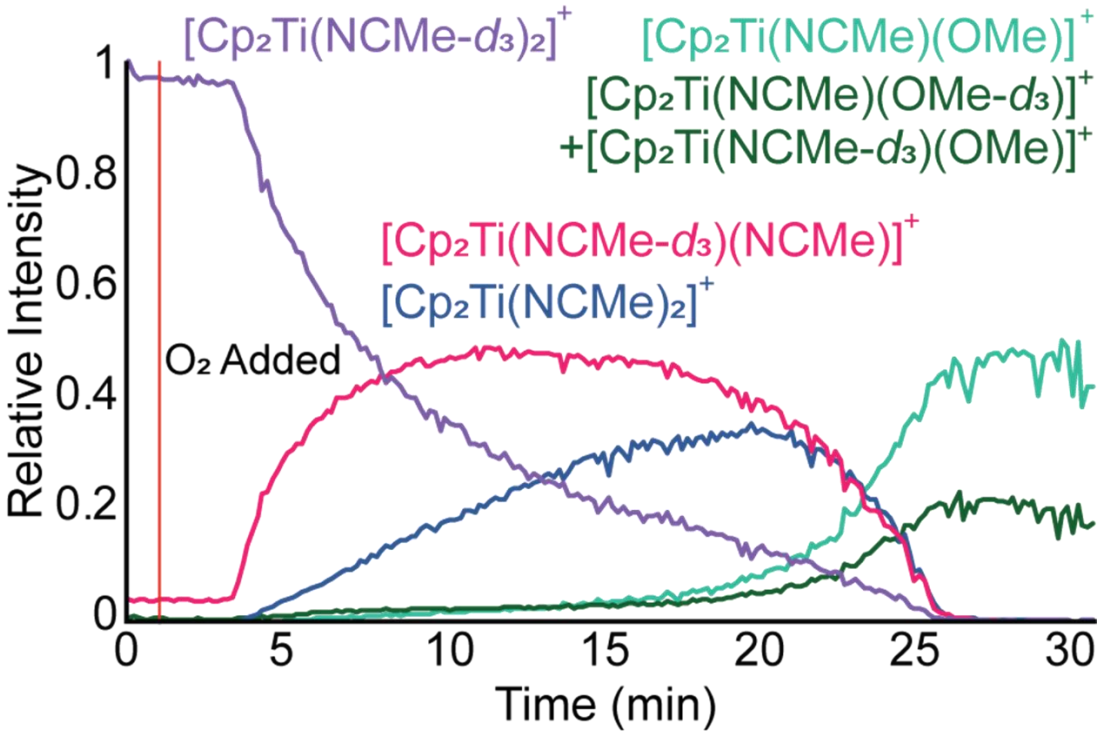


Figure 2.10: Deuterium labelled study for the oxidation of titanocene

In summary, the deuterium-labelled study provided two main insightful results: firstly, it helped determine the possibility of a methyl abstraction over a simple methanol contamination, and secondly, it showed us the compound from which the methyl group was acquired from.

2.5. Experimental

2.5.1. General considerations

All experiments were carried out under inert atmospheric conditions using standard Schlenk techniques and glove box techniques. Acetonitrile (reagent grade, Caledon Laboratories) was dried over CaH_2 and distilled prior to use. Acetonitrile-d_3 (Sigma Aldrich) was used without further purification. Cp_2TiCl_2 (Sigma Aldrich) and zinc dust (325 mesh, Anachemia) were used without further purification. Glass fiber syringe filters (0.45 m, Whatman) were dried prior to use.

The colorimetric chronograph was constructed using my own in-house built software ColorPixel, whereby time-lapse photography of the reaction captured via a smartphone was processed to get the final chronograph.

All ESI-MS spectra were recorded using a Waters Acquity Triple Quadrupole Detector equipped with a Z-Spray electrospray ionization source. The capillary voltage was held at 3.1 kV, cone voltage at 5.0 V, and extraction cone at 1.0 V. The following settings were used to obtain optimal desolvation conditions: desolvation gas flow rate 300 L/hr, cone gas flow rate 150 L/hr, source temperature 85℃, desolvation temperature 180℃. The detector gain was set to an optimal voltage of 470 V. Scan time was set to 1 s, with an inter-scan time of 0.1 s. MS/MS experiments were performed with a collision energy between 2-50 V with an argon collision gas flow rate of 0.1 mL/hr.

2.5.2. Preparation of solutions for analysis

$Cp_2Ti(III)(MeCN)_2$ was prepared using a modified literature synthesis in which Cp_2TiCl_2 (32 mg, 0.13 mmol) was dissolved in 60 mL of MeCN. To the solution was added zinc

dust (2 g, 30.6 mmol) and the resulting solution was stirred for 2 days at room temperature. Filtration of the solution enables a final $Cp_2Ti(III)(MeCN)_2$ concentration of 5 mM that was used directly in the ESI-MS experiments without significant decomposition.

Upon infusion of the synthesized $Cp_2Ti(III)(MeCN)_2$ as the oxidation starting material, a prominent peak at *m/z* 260 was present indicating the successful formation of $Cp_2Ti(III)(MeCN)_2$. To ensure that this species was the target titanocene complex, a comparison of the experimental data with the theoretical isotope pattern was made and the two patterns matched, as shown in the figure below.

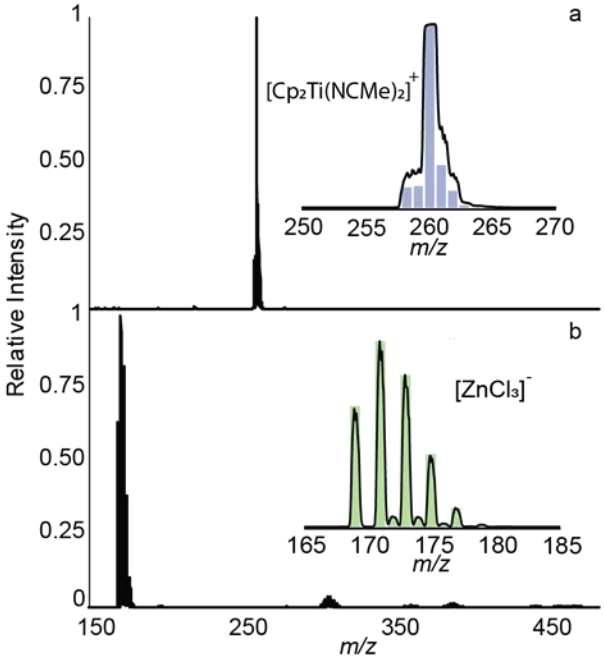


Figure 2.11: Isotopic confirmation of the cationic titanocene(III) bisacetonitrile (a) with its anionic counterion zinc(II) chloride (b)

2.5.3. Reaction PSI-ESI-MS Details

2.5.4. Oxidation

In a typical experiment 4 mL of a 5 mM stock solution of $[Cp_2Ti(III)(MeCN)_2]^+$ was taken and filtered through a glass fiber syringe filter into a 8 mL sample vial enclosed with a septum. This system was connected to the ESI-MS and used for the initial data collection. In a separate vial, oxygen was sparged into 10 mL of acetonitrile for 2 minutes to ensure saturation. After data collection had started, this solution was infused into the vial via a syringe pump at 0.200 mL/min to start the oxidation. The speciation was monitored until no more changes were observed (20 min).

2.5.5. Hydrolysis

In a typical experiment 4 mL of a 5 mM stock solution of $[Cp_2Ti(III)(MeCN)_2]^+$ was taken and filtered through a glass fiber syringe filter into a 8 mL sample vial enclosed with a septum. This system was connected to the ESI-MS and used for the initial data collection. Separately different solutions of water in acetonitrile were prepared volumetrically and introduced to the main solution via a stepper motor pump at 0.200 mL/min. Where applicable oxygen was sparged into the main solution to induce oxidation.

2.6. Conclusion

Studying the oxidation of titanocene has led to many new-found knowledge about how the radical titanocene(III) behaves under the conditions of a compromised atmosphere. Starting with the initial confirmation, through mass spectrometry, the incorporation of solvent into the titanium(III) complex supports the findings by Nugent and

RajanBabu[120,121]. The oxidation behaviour of the titanocene radical to the oxo-titanocene(IV) complexes have been elucidated through PSI-ESI- MS and the reaction chronograph outlines the reaction progress as the oxidation occur pair with the color change information through ColorPixel.

Through high resolution mass spectrometry, many oxo-titanium species were identified. Unfortunately, not every species was able to be determined, thus other methods for the characterization of titanium complexes would assist in this further endeavour in fully understanding the roles of the low intensity peaks. Oxidation and hydrolysis comparisons have helped determine the properties of titanocene as a glovebox indicator, and suggesting that the indicator is more sensitive to the presence of oxygen than water.

The deuterium-labelled study explores the possibility of a methyl abstraction and determining the source of the methyl group by locating deuterium labels in the mass spectrum. This study has confirmed the methyl abstraction pathway for the formation of methoxy titanium species under the oxidation conditions and fragmentation studies have determined the methyl group was abstracted from acetonitrile. The decomposition of titanocene has offered many new insights on oxo-titanium chemistry.

3. CHAPTER 3: MASS SPECTROMETRY SOFTWARE

3.1. Introduction to Software Development

There has been a marked increase in the use of technology in all domains of human endeavours to solve complex problems. Chemists often talk about toolboxes for reactions: especially for medicinal, computational, and physical chemistry[138–141]. Being a field that focuses on application-driven technological advances, the use of software is a significant part of the industry. The utility of these software tools transcends various fields of research and problem solving – from programmable logic controllers for large scale production of chemicals, to machine learning algorithms for the determination of drug compounds optimized to be potent and safe for clinical trials[142–146].

The chemical industry and academia use data analytics to handle the sheer amount of data collected from data-driven experiments. There is an old English proverb that goes, "Measure twice, cut once." This makes literal and figurative sense in that one should always be careful and thorough in planning before taking any action. The same applies in industry: in this analogy, measuring refers to the simulations or models that are obtained from the data collected, and the cutting refers to the execution of the experiments derived from results in these data models.

The foundation to data processing lies in programming and software development. The development of a stable stack, concept of inter-software workflow, indicated by an efficient and robust software pipeline. This allows for scalability in resource consumption and ease of development for other programmers using the methods[147,148].

ProteoWizard has been a common resource for processing raw MS files. Since its first publication in 2012, there has been 505 citations attributed to the use of the software in some form for processing MS files[149–152]. However, despite its extensive usage, there are still very few available converters like the ProteoWizard MSConvert that are agnostic to proprietary file types. Outputting the MS data as a text file removes the difficulty of the parsing binary files in an exploratory manner by removing the need to interact with binary files from the start.

3.2. PythoMS

Prior software projects for MS data processing in the McIndoe group include a recent publication "PythoMS: A Python framework to simplify and assist in the processing and interpretation of mass spectrometric data"[153]. The development of this project initially started as a single script file for processing full scan and tandem mass spectrometry, MS/MS, data. Functions of this software included generating spectra and chronograms from raw experimental data, along with video animations of reaction progress and isotope pattern overlays for characterizing a target compound of interest.

The script was then expanded to a framework by re-factoring the code to accommodate the development of scripts that can use the same functions and methods such as extracting chronogram information for individual species. However, the performance of the framework began to require more resources for processing data; mainly, more memory and computational power were needed. Users started running into issues such as low memory and insufficient computational power to perform the tasks in a timely

fashion. A run time ranging anywhere from 20 minutes to 2 hours at maximum load on a single core was a major drawback for users. Other limitations faced included difficulties with the installation of the framework along with having a consistent code base amongst all installments, due to the lack of an updating mechanism for new bug fixes for the software. Users reported the advantage of automated file conversion and processing for figure development, but often gave negative feedback about resources, installation, bugs, and unfriendly interface.

There was an imminent need to address these issues, and thus the goal of this project was to develop a library to handle MS files. The main objective in doing so was to make all MS software conform to a single standard, regardless of language and MS file format. This would attempt to resolve the issues of resources, installation, and bugs. The library was designed for developers to construct MS code, and provide a consistent methodology in accessing and processing data.

3.3. Spectra.ly: A Lightweight Mass Spectrometry Data Processing Framework

Spectra.ly was developed to be lightweight in size and resources for processing MS data. With a total size of just 23 kB in native python, this makes the library easily accessible to all developers regardless of their storage resource limitations for development. Built with an object-oriented approach, each file was instantiated (i.e. initiated as a class) to be a Run class that consists Spectrum, Function, and Chronogram classes. Each of these classes were instantiated and on start would parse through the file and acquire

information related to the class. Each class has a long list of getters and setters – a consistent means to access the information for the external scripts.

The architecture of this library was designed with the intent to have a server-based access of methods and functions, and the information will be procured as API requests, the Remote function mentioned in Chapter 5. The surge of client-based JavaScript software[154–156] has the processing take place on the client/user hardware. Many front-end focused flavours of JavaScript leverage the resources of the client's hardware, to minimize computational strain on the server. Applications using Angular, Meteor, or even Electron use client hardware[157–159]. Such a paradigm shift is most suited for light computational tasks, such as to-do lists, online shopping, words translation, and animations. However, when the computational tasks at hand become too taxing, it would improve client-side performance by offloading the taxing processes to the server. The caveat for server-side processing is the availability of sufficient server resources. Many webserver architectures provide minimal resources to send the JavaScript application and have the product render using client hardware. Migrating such servers to facilitate server-side processing would be difficult due to the resource bottleneck.

Accessing the functions of Spectra.ly can either occur locally as a Python library or can be done as API calls. Local Python access would be done by a simple "#include" and a Run object can be instantiated by providing the experiment MS file and the EXP file containing the MS parameters. The constructor of the Run object will provide instances

of Spectrum, Function, and Chronogram objects which can be accessed through the Run object.

To demonstrate the simplicity of Spectra.ly, an analysis of a carboxylic acids was done. Using barium acetate, the carboxylic acids (CA) formed a Ba-CA adduct in spray and is introduced to the mass spectrometer[160,161]. Upon fragmentation of any Ba-CA adduct, two diagnostic peaks appear: $[Ba\text{-}H]^+$ and $[BaOH]^+$. A precursor scan locking on the product ion BaOH will yield a spectrum outlining all the peaks that formed the Ba-CA adduct. The data is then processed through Spectra.ly and below shows the resultant figure of the experiment and the code necessary to construct the plot.

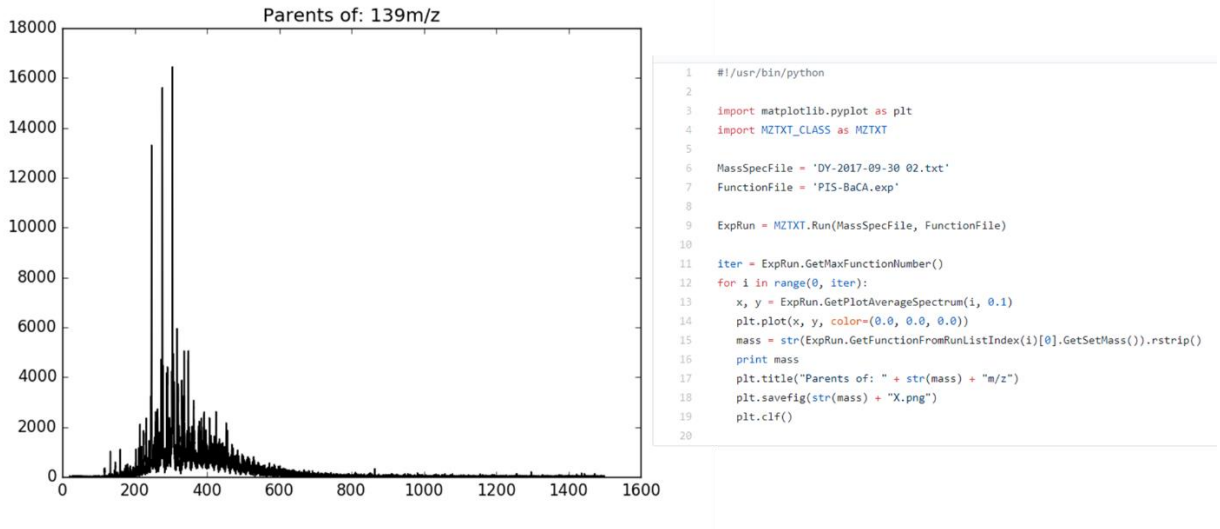
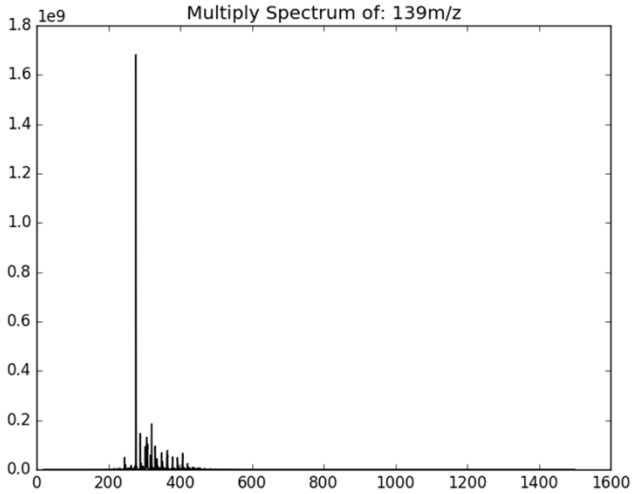


Figure 3.1: Demonstration of the Precursor Ion Scan processing script and the resulting Precursor Ion Scan Spectrum of a Naphthenic acid mixture

The processing methods are not only restricted to the scan modes of a mass spectrometer. Continuing with the Ba-CA adduct experiment, it was found that noise had a profound impact on the full scan spectrum. Averaging the spectrum yielded unpromising results in handling noise. I developed a processing technique for mass spectrometry to enhance the high intensity signals in a spectrum. Taking the product of all the scans, the peaks with

high intensity will get higher with each scan product, whereas noise peaks as they are smaller in intensity will experience a less dramatic enhancement. The final result amplifies the desired peak information and minimizing the noise in the spectrum. Spectra.ly was used to construct a proof of concept for the signal processing methodology and shown below is the code necessary to construct the plot along with the process spectrum.




```
#!/usr/bin/python

import matplotlib.pyplot as plt
import MZTXT_CLASS as MZTXT
import numpy as np

#Multiply Spectrum is used to discern common peaks between multiple spectrums

MassSpecFile = 'DY-2017-09-30 02.txt'
FunctionFile = 'PIS-BaCA.exp'

ExpRun = MZTXT.Run(MassSpecFile, FunctionFile)
lenRun = ExpRun.GetMZDeltaIndices(0, 0.1)
arrX = []
arrY = np.ones(lenRun)
iter = ExpRun.GetMaxFunctionNumber()
for i in range(0, iter):
    x, y = ExpRun.GetPlotAverageSpectrum(i, 0.1)
    arrY = np.multiply(arrY, np.array(y))
    arrX = x
    mass = str(ExpRun.GetFunctionFromRunListIndex(i)[0].GetSetMass()).rstrip()
    print mass

plt.plot(arrX, arrY, color=(0.0, 0.0, 0.0))
plt.title("Multiply Spectrum of: " + str(mass) + "m/z")
plt.savefig("MultiplyX.png")
plt.clf()
```

Figure 3.2: Demonstration of the Multiply Spectrum function and the resulting Multiply Spectrum of the Precursor Ion Scan spectrum

3.4. Class Information

The Run class is used to manage the information for one RAW file. It first determines the amount of scan modes and acquisitions that are present in each experiment with supplementary information provided by the experimental EXP file. This allows the software to accurately calculate factors pertinent to the data processing directly from the RAW and EXP file combined. Traditional methods rely only on the RAW file to provide

the information required EXP files also contain insight on how the data was acquired and other instrumental parameters that are hard to obtain otherwise.

In this class, functions to obtain information of the overall experiment are made available for script developers. Such functions include: obtaining the reconstructed single ion reaction, finding the *m/z* resolution, binning *m/z* to appropriate bin for resolution reduction.

Reconstructed single ion reaction (RSIR)[162–164] is the use of experimental data and isolating the signal change of a single *m/z*. The resulting chronogram from this processing is considered an RSIR. RSIR is used to monitor the changes of species in a reaction is used in Chapter 3 to monitor the oxidation of titanocene.

However, relying on EXP can be considered naive as the expectation of the user having an EXP file can vary depending on the instrument manufacturer. The RAW files accessible in this body of work are produced by a Waters instrument which can output experimental parameter files as EXP.

3.4.1. Sinatra: Automated EDESI for Selective Quantification

Sinatra is a program that automates the construction of energy dependent electrospray ionization (EDESI) experiments and renders the results into a fragmentation contour map. To provide context, EDESI is a method to study the fragmentation of an analyte species by leveraging the power of tandem mass spectrometry. Tandem *in space* mass spectrometry, as mentioned before, is a powerful technique for triple quadrupole MS^{86}. In

triple quadrupole MS, the fragmentation for the tandem MS/MS experiments are done by collision induced dissociation (CID)[165–167]. CID is done by introducing argon to the collision cell and allowing the inert gas to collide with the accelerated analyte ion. Due to the energy of the collision, the energy is sufficient to induce fragmentation of the precursor ion to product ions[168–170].

Finding the optimal collision energy is challenging and can only be determined by obtaining the fragmentation pattern at each collision energy. Typical methods include an accelerated collision energy ramp at 5 V intervals from 0 V to 50 V. While this method would be sufficient for routine analytical quantification work, it loses structurally relevant information about the analyte. The results from EDESI can not only assist in improving signal quality by traversing into higher energy regimes, it also can reveal structural information, such as ligands or functional groups, throughout the fragmentation process mapped into a contour plot.

Prior work on EDESI involved the manual operation of collision energy ramp[171–175] and the use of Origin scripts were difficult to use[176]. For many applications, less is more; in the case of software such as mobile applications, the installation process is simple and will enforce the behaviour of downloading apps as it can be done in one click and its ready to use[177–179]. When users encounter unknown software packages that are required for the user to install prior to have a functioning script, then it is more difficult for the end user to manage. Web applications alleviates many issues with end user involvement and the application is available for use without installation. Many web applications require

only user subscription or an account creation. Software like TinkerCAD are more accessible due to the simplicity in starting the software[180].

To improve the accessibility of EDESI, two things are of focus: first is to automate the experimentation aspect by having software controls for the collision energy for the experiment and the second objective to improve is the accessibility in constructing such plots by improving the start up experience and support simple intuitive controls.

Sinatra is a facelift of the original EDESI processing software to combat the accessibility issues surrounding the prior iterations. The intuitive nature of using Sinatra with a GUI and also a command line interface, processing EDESI contour plots is simplified where users do not need to perform manual operations to get the desired output. Built with the intent of automation, it is also can be headless to allow the streamlined work flow from Figure 1.8.

To solve the experiment automation, AutoHotKey (AHK) was used to control the MS software and more specifically the slider for controlling collision energy[180]. The function of this script is to increment the slider by one collision energy per second or the duration of time set by the user. This will allow the analyte ion to be analyzed through EDESI automatically. Even though the collision energy is collected by the output Waters RAW file and obtained through ProteoWizard[181], Sinatra offers an additional fail safe by tracking the collision energy ramp into a text file.

Improving the accessibility of the software, the controls, and parameters for the software are available in two methods. The first method is through a Graphical User Interface (GUI) Tkinter[182] interface that has a pop-up menu asking for the relevant parameters for the user to input. The second method is through command-line switches such that it can be available for programmers who desires to run a batch script to process multiple EDESI plots simultaneously.

The delivery of the software can be improved through a packaging step and constructing an installation file MSI file allowing the user to download and install the MSI file. This method will allow for automatic installation of required packages, Python 2.7.x/3.x, and ProteoWizard for MSConvert.

3.4.1.1. Real World Application: Protein and Sugars

Having explained the software, there are two real world examples of samples that are used to test Sinatra.

To demonstrate the power of Sinatra, automated protocols available from the software are performed on a Waters Acquity TQD running MassLynx. Two real world samples of a Cbx7 inhibitor peptide and sucrose were used for this demonstration.

3.4.1.2. Peptide – Cbx7 Inhibitor

Peptides are amino acids connected by amide bonds and under mass spectrometric conditions fragment in a regular pattern[183,184]. The mass loss between the fragments will

give insight into what amino acid residues are in the peptide[185,186]. This would be useful in a shotgun proteomics[187–189] in identifying the digested peptides. Current methods of SWATH DIA require hardware to support to do exploratory proteomics[190,191]; however, using Sinatra, older instruments are able to sequential exploration through EDESI of each ion without hardware modifications.

As a proof of concept, the Cbx7 inhibitor CA-1-160 from "Structure-Activity Relationships of Cbx7 Inhibitors, Including Selectivity Studies against Other Cbx Proteins" by Simhadri et al[192].

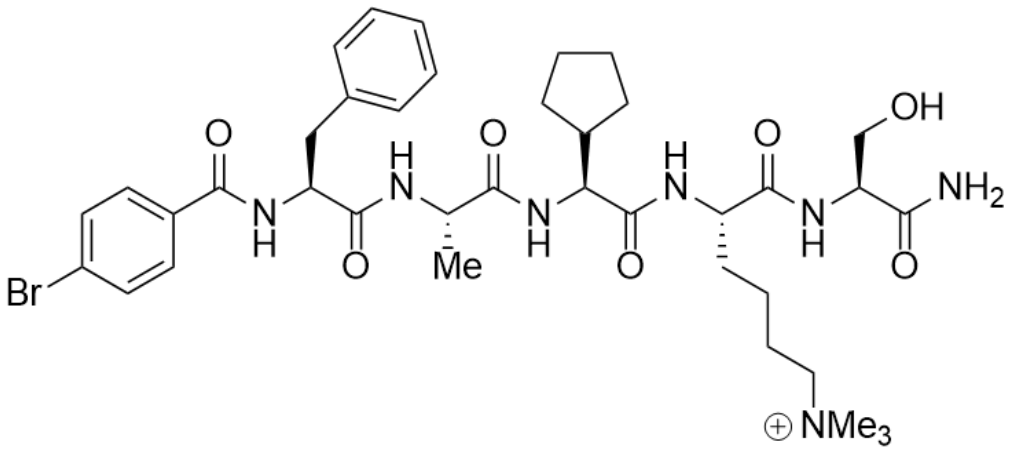


Figure 3.3: Structure of peptide-based inhibitor for Cbx7

Sinatra was used to analyze this sample and confirm the identity of the CA-1-160 peptide. From the fragmentation contour plot, the individual amino acid residues can be determined by the mass difference. Upon comparing the residue to the mass transition, the following residues were found.

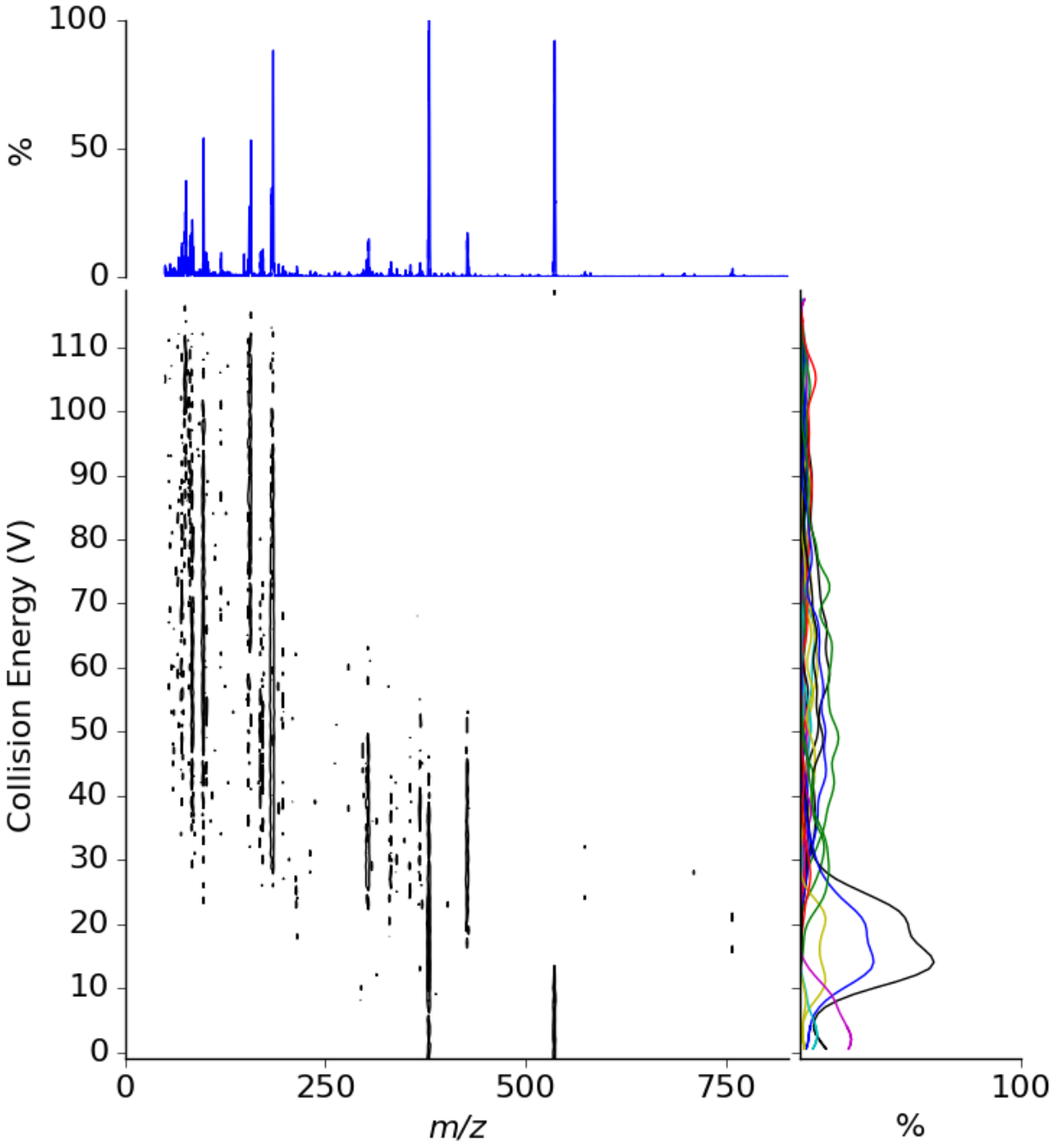


Figure 3.4: EDESI contour map for the fragmentation of the Cbx7 inhibitor

3.4.1.3. Oligosaccharide - Sucrose

Sugars, or rather polysaccharides are of interest when paired with MS in glycomics. The strive to understand the structural aspects and activities it is a common post translational modification[193]. Mass spectrometry is a suitable technique for biologically relevant polysaccharides as they are difficult to characterize due to their low abundance in a biological system[194–197]. Eight percent of post translational modifications are glycosylations and this modification is significant for protein folding and cell to cell adhesion[198]. Work by Yamagaki et al presented an interest in the structural fragmentation of sugar through tandem MS experiments[194]. My work in this section attempts to assist in the expansion of understanding mass spectrometric parameterization in glycomics. Using sucrose as the model oligosaccharide, Sinatra was used to obtain the EDESI information shown below.

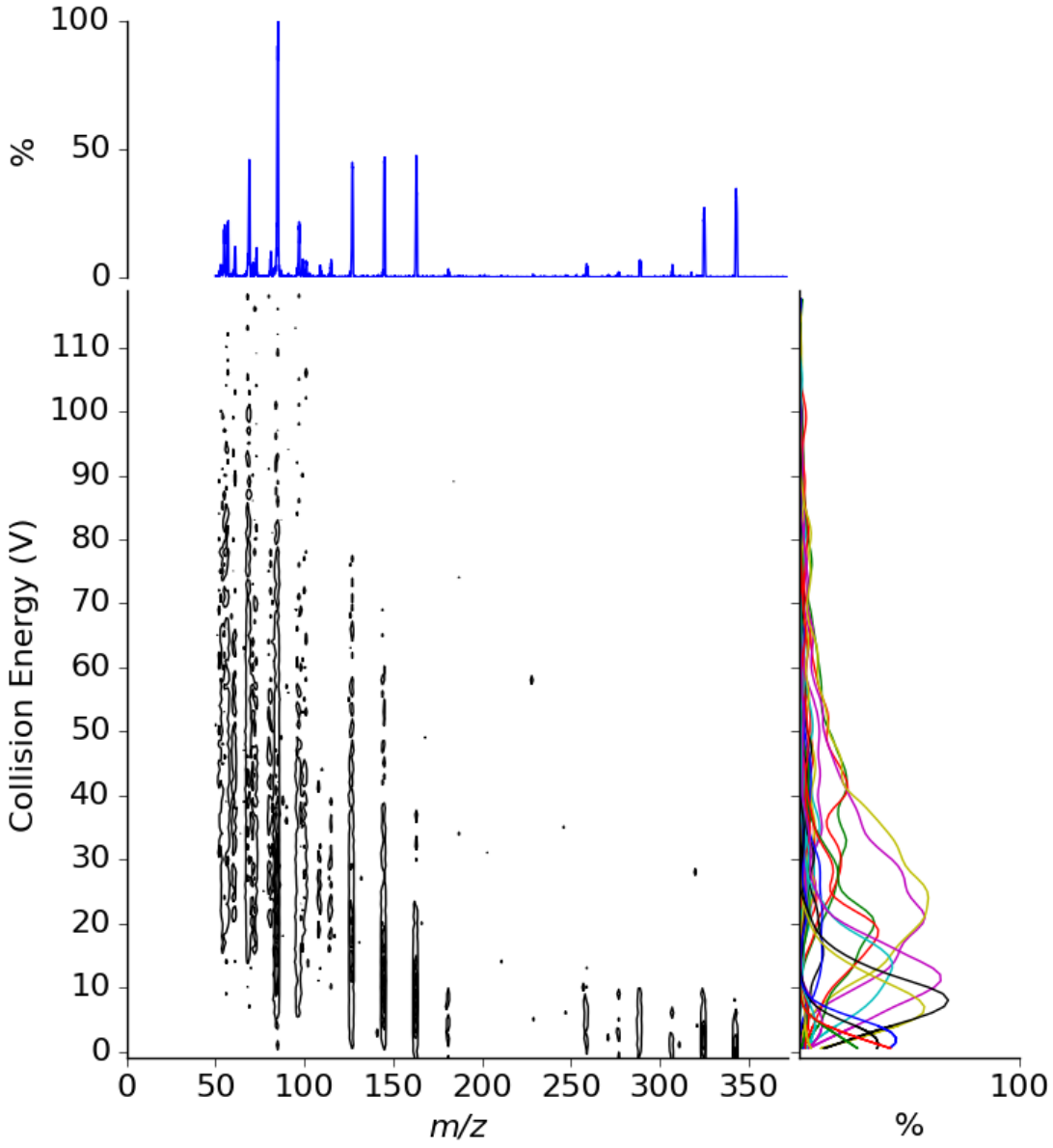


Figure 3.5: EDESI contour map for the fragmentation of sucrose

As can be seen the species fragment in range 0 – 80 V and any higher than 80 V has no further effect on the fragmentation. The *m/z* gap between each fragment correspond the fragment dissociated and is consistent with fragmentation of sucrose[194].

3.4.1.4. Organometallic Cluster – Methylalumoxane

Sinatra can also be suitable for inorganic complexes or clusters[199,200]. EDESI started as a method to determine the amount of carbonyl ligands in a metal cluster. Similarly, EDESI can be used to understand ligand identity and behaviour for systems like methylalumoxane (MAO). MAO is a main group organometallic cluster of aluminum, oxygen, and methyl groups. MAO is used as a co-catalyst during the Ziegler-Natta process for olefin polymerization[201–203].

Below is an EDESI plot for a MAO (16, 6) cluster where there are 16 MeAlO and 6 Me_3Al ligands in this macromolecule.

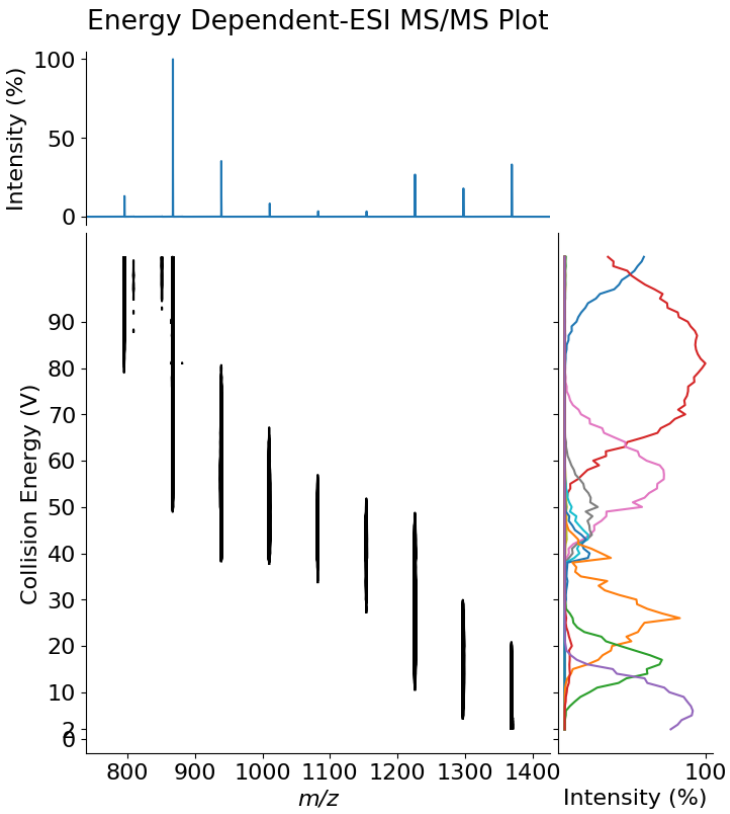


Figure 3.6: EDESI contour map for the fragmentation of the (16,6) MAO cluster. Note: (16,6) is the ratio of (MeAlO:Me_3Al)

3.4.2. AutoMRM: Automation of MRM Method Development

Multiple reaction monitoring (MRM) is an ion selective scan mode for triple quadrupole instruments. The method isolates the precursor ion in the first quadrupole then isolate a resulting product ion that is specific to the analyte in the third quadrupole. From the two isolation events, the filtered ion beam after the third quadrupole will only contain product ions that were once the target precursor ion, not a contaminant. It can be assumed the resulting ion beam contain only ion that represent the analyte. This assists in removing overlapping signals and contaminants as the mass spectrometer is only monitoring one analyte as seen in the figure below[204–206].

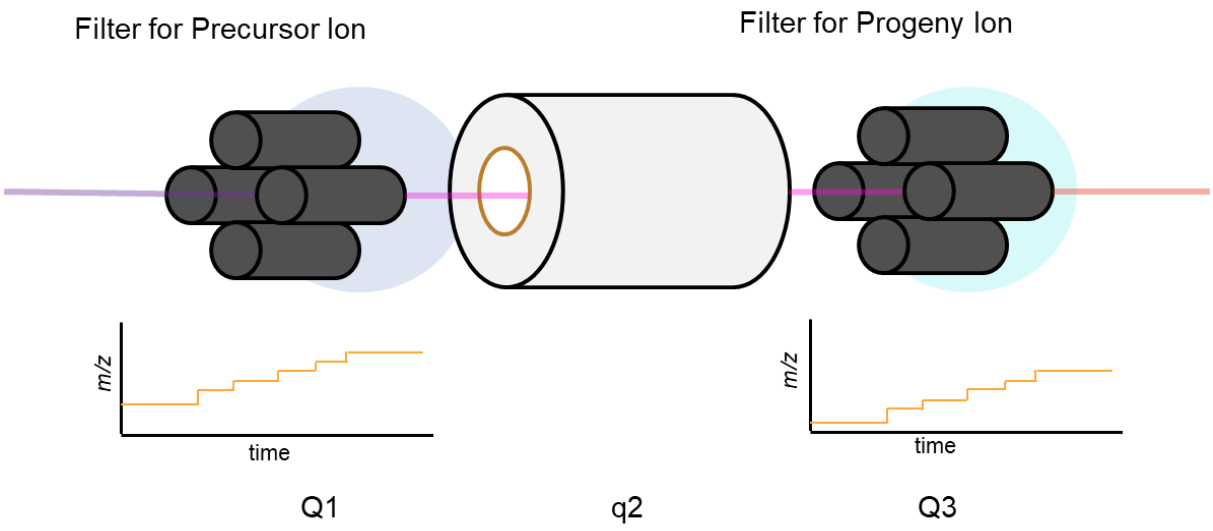


Figure 3.7: Visual demonstration of Multiple Reaction Monitoring and the graph showing the *m/z* isolated for each quadrupole

Traditionally, MRM methods are made manually. Technicians would infuse the sample into the mass spectrometer and observe the peaks in the spectrum that appear post-infusion. Upon obtaining a list of peaks, the technicians then sequentially isolate the

peaks and for each peak increase the collision energy until fragmentation starts. The technician would then log the fragment ion and the collision energy during this process while continuing to find other fragments at higher collision energies. This procedure would then be repeated for the list of peaks found in the initial spectrum. The tasks are repetitive and are well suited for automation.

Automating mass spectrometry software is not an easy task. Most mass spectrometry software are closed source and difficult to find access points to control the mass spectrometer. GUI automation was the simplest method employed to control the software of the mass spectrometer by emulating mouse and keyboard commands. AHK was a GUI automation language[207] that was designed to testing UI functions to ensure buttons and sliders were behaving according to specification. Leveraging the abilities of AHK, functions were made as an interface between code and UI controls.

To minimize the amount of content required for installation on the client side, only two AHK files and a lightweight python client are required. The two AHK files consists a library of functions "MSAuto_Class.ahk" and the other is the active script "AutoMRM.ahk". The script initiates the sequence by starting a full scan acquisition for 10 seconds to survey the peaks present in the sample and the result Waters RAW file is translated to text by ProteoWizard[181]. The text file is sent to a processing server hosted in Amazon Web Services (AWS) on a Lightsail 1GB RAM instance.

After the server is done processing, it returns a list of peaks to monitor. The client-side script the initiates the second step and sequentially iterate through the peaks to determine the fragmentation pattern. Fragmentation was determined by EDESI. The EDESI request returns a fragmentation map of collision energy versus and *m/z*. For this proof of concept, the highest intensity fragment from EDESI contour plot was chosen and the collision energy is recorded as information for the MRM transition. This concludes the flow of information of one MRM method for one analyte ion and this repeats until the peak list obtained from the full scan is completed.

When all MRM methods are completed, the information is compiled into a Waters EXP file ready for experimentation. This file can then be used for quantification using a triple quadrupole mass spectrometer and if applicable, to a liquid chromatography setup. This will allow selective quantification to be done in a fast and consistent manner.

The flow information is a requires a constant communication between the client and the server. The server is where all the processing take place; from determining the peak list in the full spectrum to constructing the EDESI map to find the optimal fragment. The processing steps are abstracted to minimize the resources needed on client-side operations, servers carry the processing load with its higher performance components. This paradigm remains unchanged when the AWS instance is migrated to a local system. The workflow remains the same and the only change is the server IP on the client software. This implementation of AutoMRM is designed to integrate seamlessly into any workflow and will work on top of the MS control software.

On the processing side of the server, there are 4 available Application Programming Interfaces (API) that can be accessed from the client software: "/PeakList" - for returning the peak list for fragmentation, "/FragSearch" - for finding a suitable MRM fragment through EDESI, "/InitPage" - for starting the EXP file, and "/MRMInfo" - for returning MRM method blocks for the EXP file.

3.4.2.1. PeakList

Input: RAW file in text format by ProteoWizard.
Output: String output containing peak list delimited by commas.

The PeakList endpoint requires the input of the RAW file converted into a text file via ProteoWizard. From that, PeakList averages all spectra obtained in the ten second time frame and selects peaks that are relevant to the spectra. The peak selection is done by a package known as peakutils which is the main dependency for this function.

3.4.2.2. FragSearch

Input: RAW file in text format by ProteoWizard AND Collision Energy text file made automatically by client side AutoMRM script
Output: Optimal fragment ion (in this case max signal) and collision energy in string format delimited by commas.

FragSearch takes in the text file and pairs the information with the collision energy. It then follows the same procedure as Sinatra and generates a 2D contour array of intensities. From that, it picks the value with highest intensity and will find the index the entry is located in. Using the entry value as the key to determine the collision energy, the server then returns the collision energy and *m/z* value delimited by commas.

3.4.2.3. InitPage

Input: N/A

Output: Preamble for MRM EXP file

InitPage returns the starting parameters that are standard amongst MRM EXP file for Waters instruments. This allows for the subsequent parameters to be appended after this index and combined with other MRM information blocks to complete the method EXP file.

3.4.2.4. MRMInfo

Input: JSON wrapped data containing

- INDEX - Index number of current MRM transition
- M1 - Precursor Ion of current MRM transition
- M2 - Product Ion of current MRM transition
- CollE - Collision Energy of current MRM transition

Output: String containing parameters for current MRM transition

MRMInfo returns the parameters for the current MRM transition and is appended after the MRM EXP preamble from InitPage's output. This allows for modular approach to constructing an MRM EXP file and is scalable to however many analytes present in the spectrum.

3.4.2.5. Real World Application - Pharmaceuticals

Now that the program is implemented, testing the application on real world problems would be a valuable demonstration on its potential. Pharmaceuticals are good candidates for preliminary testing as they have a well documented ingredients list and it is a sufficiently complex matrix (i.e. filler, color dye, and polymers) to test AutoMRM. Two drugs are tested for this: first is Reactine, a single active compound in the presence of matrix and the second is cold medication from Bufferin.

3.4.2.6. Reactine

Upon infusion of the Reactine solution, only one peak was present in the spectrum at *m/z* 389. AutoMRM detected a peak at *m/z* 389 and found many peaks in the noise of the spectra. This is normal for many peak picking algorithms and performance can be improved by altering different variables such as the peak threshold. The spectrum for the infusion can be seen below.

The high molecular weight signals in Figure 3.8 maybe attributed to electronic noise or chemical noise in the form of contamination. The secondary emission detectors in the mass spectrometer multiply the signal of an ion colliding to the detector. A stray ion in

the vacuum colliding with the detector at the wrong time can trigger a signal response result in electronic noise for the mass spectrum. Contamination from prior experiments can influence the results of the current acquisition. In the McIndoe lab, species from palladium catalyzed Suzuki cross coupling can create ions in the region above *m/z* 1000 which can attribute to the signals observed in the high molecular weight region.

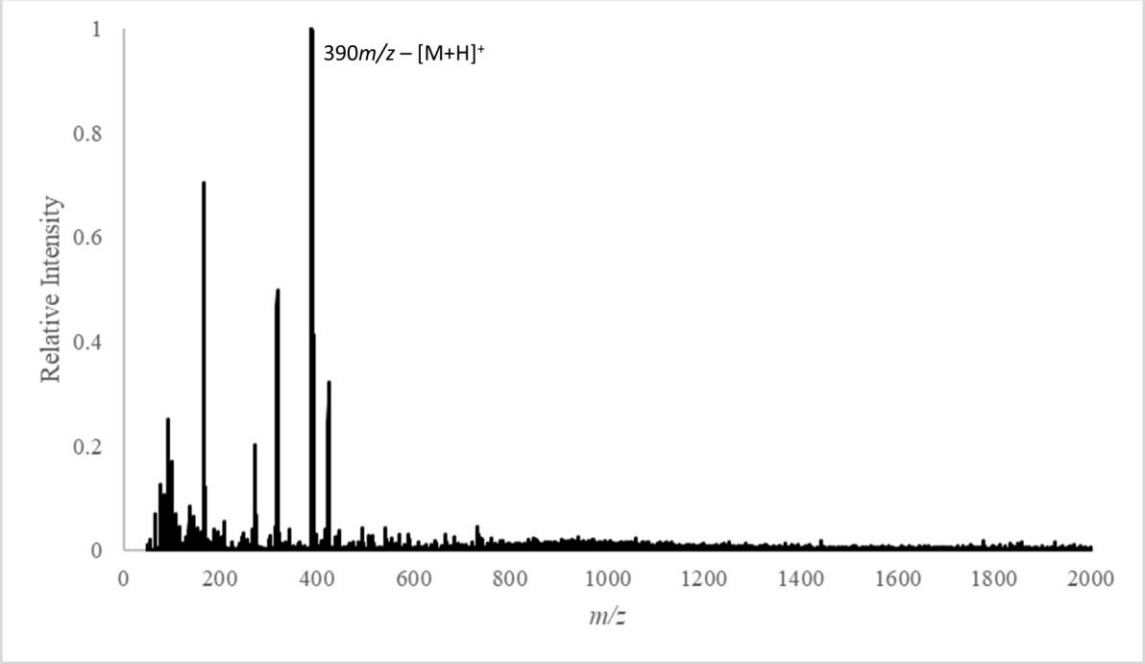


Figure 3.8: Mass spectrum for the infusion of Reactine

3.4.2.7. Cold Medication

In a more complex system such as cold medication, it can be used as a benchmark for AutoMRM to see how well it performs. There are two things that can be assessed for this software: 1) peak picking and 2) precursor and product mass for transitions. The infusion of the sample resulted in four major peaks at: *m/z* 161, 267, 307, 418. The first attempt at locating the peaks with the peak picking algorithm returned: *m/z* 5, 116, 267, 373, 567,

721, 872, 1024, 1174, 1326, 1477, 1628, 1779, 1928 at a peak distance of at least one hundred. All the target peaks were missed. The peak picking should improve when the peak distance is set lower than one hundred amu, such a twenty amu. Upon changing the peak distance, the peak list expanded to have 59 entries; this time it included the four peaks of interest. There are trade offs to improving peak picking but may result in increased false positive peaks.

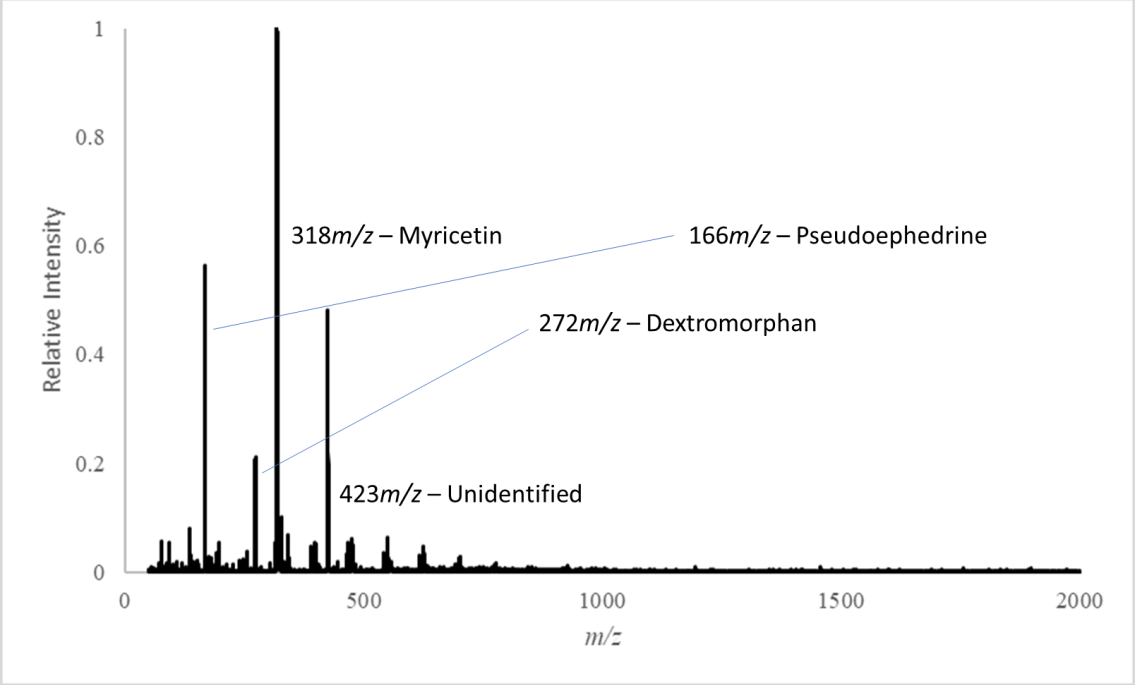


Figure 3.9: Mass Spectrum for the infusion of Cold Medication

3.5. Experimental

Both Sinatra and AutoMRM used the Waters Acquity TQD mass spectrometer. The sample was introduced by syringe pump (NE-1000, New Era Pump Systems, New York) through PEEK tubing (Internal Diameter: 0.010”, Blue). The capillary voltage was set to 3.1kV and the desolvation and cone gas was set to 200L/h and 100L/h respectively. The

source temperature was set to 85°C and desolvation temperature was 185°C. Collision energy for the EDESI experiment was conducted from 0-120V.

3.5.1. Sinatra

The peptide (CA-1-160) was prepared in methanol by creating a 1000 ppm stock which was diluted to 10 ppm for EDESI analyses. These concentrations are to serve as a guide to minimize the possibility of contaminating the mass spectrometer rather than as basis for quantification.

For the oligosaccharide project, sucrose was obtained as a Rogers white sugar sachet and was dissolved in methanol following the same procedure as the peptide resulting in a final concentration at 10PPM.

Methylalumoxane was prepared following the procedure of "Additive and Aging Effects on Methylalumoxane Oligomers" by Zijlstra et al[208]. MAO (10 wt % in toluene) was obtained from Albemarle and was stored in the glovebox freezer when shipped to the laboratory. OMTS was used to charge the MAO to obtain the EDESI spectrum and OMTS (98%) was obtained from Sigma Aldrich. The solvent used was THF and fluorobenzene which was dried and degassed following the referenced procedure. To analyze MAO for EDESI, the referenced procedure was followed to prepare the sample at ~600PPB for analysis.

3.5.2. AutoMRM

Samples of Reactine were prepared from pulverized Reactine Regular Strength tablets containing 5mg of cetirizine hydrochloride. The 10mg of the powder was weighed out and dissolved in 10mL of methanol to make a 1000PPM stock of the Reactine tablet. The stock was subsequently diluted to 10PPM for AutoMRM experiments.

The same procedure was performed on the Bufferin Cold medication where the sample was pulverized and 10mg of the powder was dissolved in 10mL of methanol to constitute the 1000PPM stock. The stock was diluted to 10PPM for AutoMRM experiments.

3.6. Conclusion

Automating methodologies through hardware and software has yield three successful proof-of-concept projects including the development of Kendrick, Sinatra, and AutoMRM. Kendrick was a hardware automation project that focused on building an automated reaction sampler that can facilitate exploratory mechanistic investigations. An autosampler was refitted with custom circuitry for the photointerruptor system and open-sourced hardware (Raspberry Pi B and Synthetos TinyG). Hardware limits as photointerruptors were installed on the sampler and will prevent motor damage from occurring during the movement along the axes. Simple dispensing and aspirating experiments were done on liquids and was successful in accurate deliveries. Future work on Kendrick will involve the integration of a homing sequence to allow for reproducible positioning for the sampling rod.

Sinatra was a software development that aimed to improve the workflow for users and developers using EDESI for analytical measurements. EDESI is a useful to characterize analytes in a sample through their fragmentation pattern upon increasing collision energy. The mass loss between the product ion can reveal information about ligands or functional group present in the target compound. Through Sinatra, this process of characterization can be done in a simple manner through user-friendly interfaces and the ability to migrate as a web application.

Finally, the last automation in this chapter is AutoMRM, a software that automates the technician's workflow in developing MRM methods. Traditional methods for MRM construction is laborious and AutoMRM alleviates this issue by combining the versatility of AHK scripts, lightweight Python clients, and a remote function-based webserver API. The main focus was to make the analyses simple where the analysis involves the infusion of the sample and running the program to obtain the resulting MRM method. AHK scripts are meant to control the mass spectrometry software through GUI automation techniques and will be able to control the acquisitions and collision energy for the analyses. The Python client acts as the intermediary between the AHK script and the server and the client shuttles data to the server for intensive processing. The server then returns the processes data for the client to compile for the final experimental file.

Three difficult tasks in analytical chemistry of exploratory mechanistic investigation, user-friendly EDESI, and MRM method development was automated through techniques explained in this chapter. The content of this chapter encourages the further incorporation of automation throughout chemistry to ease the duties for a chemist.

3.7. Conclusion

In conclusion, the development of mass spectrometry software cannot lose sight of current programming development initiatives. The importance of accommodating users and other developers in the accessibility of the program and code respectively carries over into such instrument specific software. PythoMS opened up the opportunity for the community to have access to a list of methods for chemical reactions and catalysis, but it is not limited in its applications unlike certain proteomics software. The impression that software in chemistry is a small field that is willing to accept software that only works once. Carrying the impression “no one will ever use this software” to justify a niche subpar software negatively impacts the production of quality software for chemistry. Thus, initiatives are made through PythoMS and Spectra.ly. To address the accessibility issue for developers and maintainability in PythoMS, Spectra.ly was made. It improves the ability to write mass spectrometry related code and requires less time to learn and process. It consumes less resources natively, even remotely when using server-side processing with remote functions. Spectra.ly now opens a platform for instrument software and script developers to build user accessible systems and mitigate the negative experiences associated with prior software.

4. CHAPTER 4: REACTION ROLE CLASSIFICATION WITH MACHINE LEARNING

4.1. Introduction

As seen from the results of the oxidation of titanocene in chapter 2, the different species identified in the reaction were assigned to be reactant(s), intermediate(s) and/or product(s). The method of identification of these compounds was by inspection of the reaction behaviour in the chronogram, and by chemical intuition from understanding reaction dynamics. The reaction behaviour for specific reaction roles are outlined in the ACS Catalysis paper figure[209], shown below.

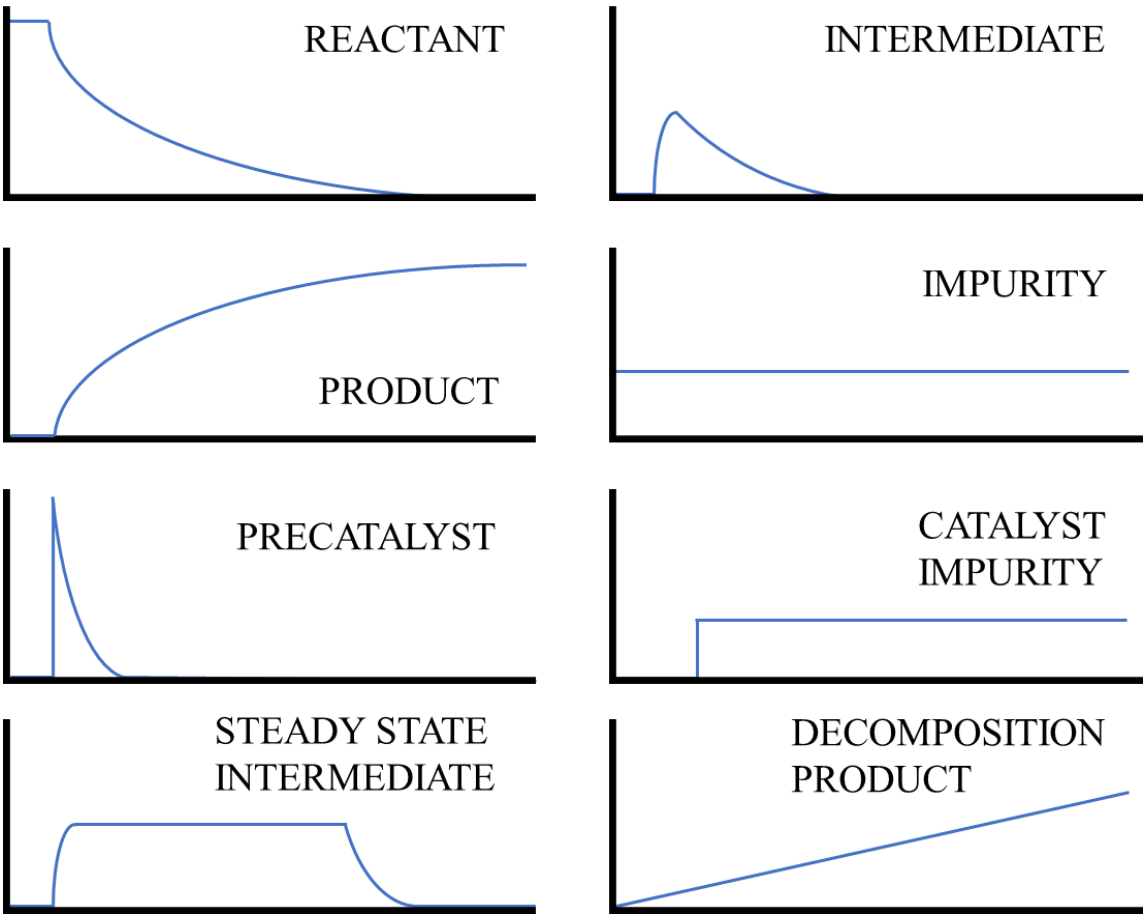


Figure 4.1: Reaction chronograms for different reaction roles

If there were a computational method by which reaction roles could be discerned, reaction analyses could be done in a more efficient manner, and we could routinely assign low-intensity species in the reaction chronogram.

Combining the fields of reaction role classification with machine learning can seem like a complicated subject matter. However, behind all the complex mathematics, it is describing a means by which chemical species can be labelled as a reactant, product or any other form of role the species plays in any given reaction by constructing models to classify how these roles are unique. The mathematics behind machine learning is daunting, but the equations are only a way to describe a relationship to the reader.

While investigating reaction mechanisms we realized that there was a lack of concise metrics to discern what a molecule is doing in the reaction. A profile that starts with nothing and ends with something is consistent with a product speciation. The way they are assigned is no different than saying an apple is an apple because it looks like an apple. How do we know an apple is an apple? Is it because it is red and juicy? What metrics can be provided to train an individual what an apple is? Teachers in kindergarten teach what an object is by showing that object and telling the children its name. “Apple, A-P-P-L-E, Apple” then teacher proceed to pull another object out of the box and say “this is NOT an apple. No! It is an orange, O-R-A-N-G-E.” If the teacher chooses a more difficult route “this is also an apple, but it’s green, a granny smith apple. Apple do not have to be red, A-P-P-L-E, apple.”

When the teacher said “apple”, it was meant as a label to the raised object. This is a way to associate the label with the object. When the orange came up, it was an example of

something that was not the target label “apple”. The contrast is needed in case the children assume everything is an apple or ignoring the object is the apple and thinking the teacher is saying apple for no reason. The granny smith apple was to show apples are not only red, but it can also be green. The students from this has learned: not everything is an apple, apple can be red or green, and orange is not an apple. Once again, “apple” can be replaced with “reactant”, “everything” with “*m/z”*, and “orange” with “noise”; therefore, “not every *m/z* is a reactant, reactants can have various *m/z*, and noise is not a reactant.” Though that may have been hard to understand through words, it would be easily understood by an algorithm.

4.2. Mathematical Analysis of Reaction Roles

Initial endeavours to construct a metric to discern between reaction roles, solving for analytic expressions was done to find differences between roles. This is possible because the chronograms are functions, the intensity over time satisfies the vertical line test. To demonstrate, reactants have the following chronogram.

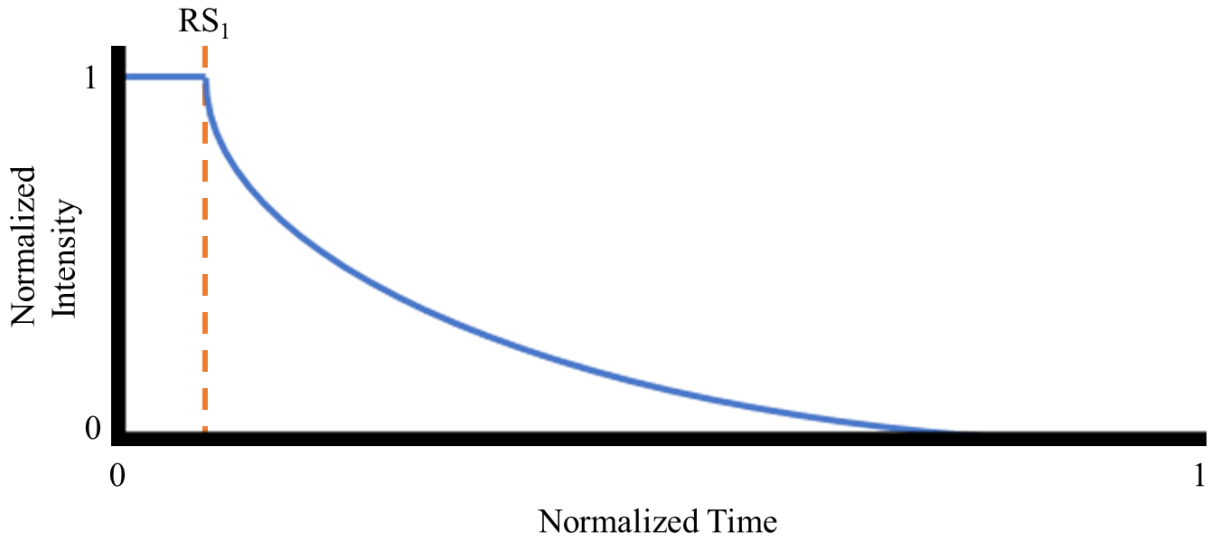


Figure 4.2: Reaction chronogram of a reactant labelled with the Reaction State times

Prior to the start of the reaction, the concentration of the reactant is expected to be maintained thus should be a horizontal line in the chronogram. Therefore, before reaction starts, $y = constant$. After the reaction starts, the chronogram signal starts to decay at a rate similar to $\frac{1}{x}$ behaviour[210,211] We can then represent y after reaction start as $y = k * \left(\frac{1}{x}\right)$ where k is a real number coefficient. The reactant function can then be presented as:

$$\textbf{intensity} = \begin{cases} \mathbf{k}\left(\frac{\mathbf{1}}{\mathbf{t}}\right), & \mathbf{t} > \textbf{Reaction start} \\ \mathbf{k}, & \mathbf{t} < \textbf{Reaction start} \end{cases} \quad \textbf{(4.1)}$$

This would be a closed form representation of a reactant, but this is only assuming if the reaction proceeded in first or pseudo-first order kinetics. Higher order kinetics may involve solutions with $\frac{1}{x^n}$.

How about products? If reactants are reacted away at $\frac{1}{x}$ rate, then product would gain at $-\frac{1}{x}$ rate. Reactions may exhibit such form of kinetics as seen in Suzuki-Miyaura cross coupling[212–214], but how about other types of reaction roles?

Intermediate species have the following chronogram behaviour.

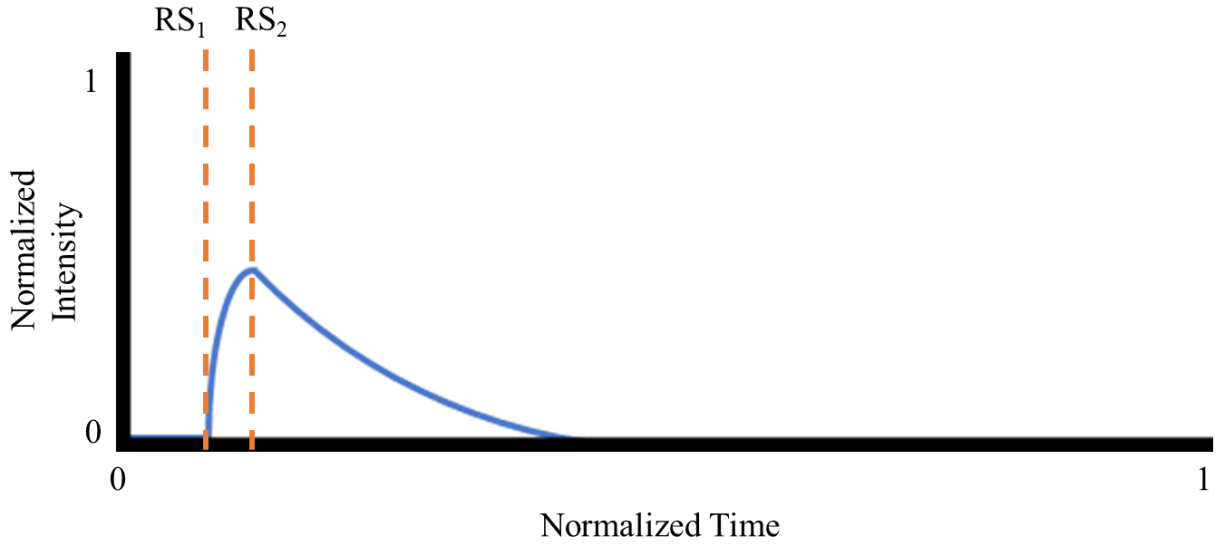


Figure 4.3: Reaction chronogram of an intermediate labelled with the Reaction State times

Intermediates are products turned into reactants for a subsequent step. Therefore, intermediates can be represented as follows:

$$\mathbf{intensity} = \mathbf{k}\begin{Bmatrix} \frac{\mathbf{1}}{\mathbf{t}}, & \mathbf{t} > \mathbf{RS_2} \\ -\frac{\mathbf{1}}{\mathbf{t}}, & \mathbf{RS_2} < \mathbf{t} < \mathbf{RS_1} \\ \mathbf{0}, & \mathbf{t} < \mathbf{RS_1} \end{Bmatrix} \quad \textbf{(4.2)}$$

RS_1: Start of Reaction

RS_2: Start of Intermediate Reaction to form Product

Steady state intermediates are similar to intermediates but have an added caveat of an induction period after the intermediate is formed until the intermediate reaction start to form the product.

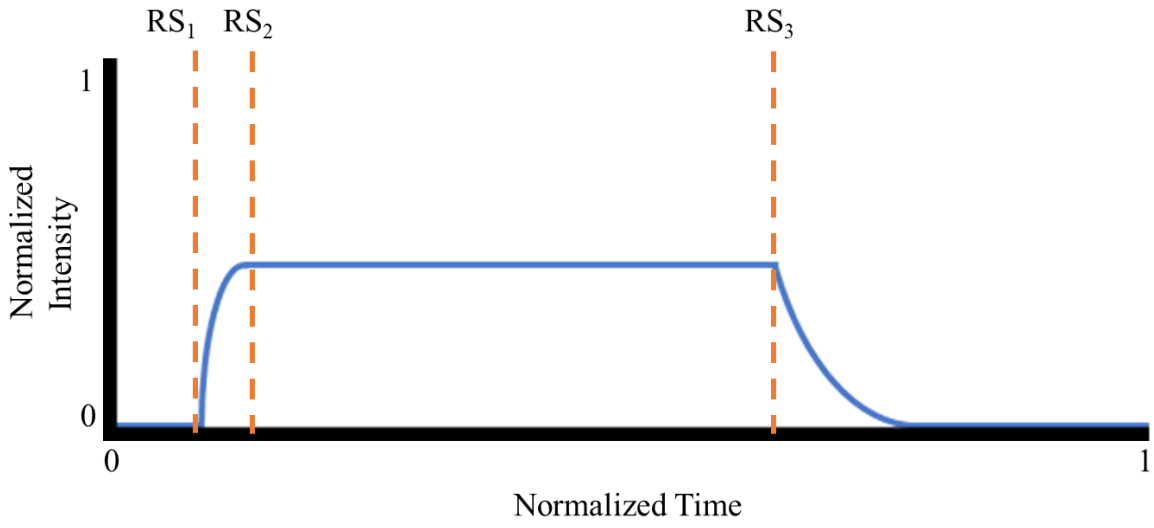


Figure 4.4: Reaction chronogram of a steady state intermediate labelled with the Reaction State times

Therefore,

$$\mathbf{intensity} = \mathbf{k}\begin{Bmatrix} -\frac{\mathbf{1}}{\mathbf{t}}, & \mathbf{RS_3} < \mathbf{t} \\ \mathbf{1}, & \mathbf{RS_2} < \mathbf{t} < \mathbf{RS_3} \\ \frac{\mathbf{1}}{\mathbf{t}}, & \mathbf{RS_1} < \mathbf{t} < \mathbf{RS_2} \\ \mathbf{0}, & \mathbf{t} < \mathbf{RS_1} \end{Bmatrix} \quad \textbf{(4.3)}$$

RS_1: Start of Reaction

RS_2: Start of Induction Period

RS_3: Start of Intermediate Reaction to form Product

Next reactions roles to elucidate are catalyst associated compounds: precatalyst, catalyst decomposition, and catalyst impurity. Precatalyst, in organometallic catalysis, is the compound that contains the catalytic metal but is inactive due to the ligand sets limiting its reactivity. The precatalysts ligands are exchanged with free target ligands in solution to form the reactive complex[215,216]. The reaction chronogram would look similar to a reactant except before catalyst addition the intensity would be at zero. Therefore,

$$\mathbf{intensity} = \mathbf{k}\begin{Bmatrix} \frac{\mathbf{1}}{\mathbf{t}}, \mathbf{t} > \mathbf{RS_{cat}} \\ \mathbf{0}, \mathbf{t} < \mathbf{RS_{cat}} \end{Bmatrix} \quad \textbf{(4.4)}$$

RS_{cat}: Injection of catalyst

Catalyst decomposition would behave like product. Decomposition products are products of a decomposition reaction. It would either appear to have a zeroth order decay and behave like $y = x$ or it would have a first order decay and behave like a $-\frac{1}{x}$ product.

$$\mathbf{intensity} = \mathbf{k}\begin{cases}\mathbf{t\ or} -\dfrac{\mathbf{1}}{\mathbf{t}}, & \mathbf{t} > \mathbf{RS_{cat}} \\ \mathbf{0}, & \mathbf{t} < \mathbf{RS_{cat}}\end{cases} \tag{4.5}$$

RS_{cat}: Injection of catalyst

Catalyst impurity is present in the chronogram since the time the catalyst was injected at a constant concentration. Therefore,

$$\mathbf{intensity} = \mathbf{k}\begin{cases}\mathbf{1}, \mathbf{t} > \mathbf{RS_{cat}} \\ \mathbf{0}, \mathbf{t} < \mathbf{RS_{cat}}\end{cases} \tag{4.6}$$

RS_{cat}: Injection of catalyst

Similarly, reaction impurity is at a constant concentration, but it would be present since the start of the experiment.

$$\mathbf{intensity} = \mathbf{k} \tag{4.7}$$

Therefore, all 8 classes of reaction roles have a mathematical behaviour they adhere to. In order to concisely define the differences to a processing program, the signal over time is required in addition to different reaction state (RS) times RS_1, RS_2, ..., RS_n, RS_{cat}.

Determining the fit of the chronogram to a $\frac{1}{x}$ trend is important. Even though the trend for each reaction role is now understood, being able to fit the data to the trend is essential to matching the reaction role. Matching the data directly to the characteristic function can be done by adjusting the parameters and coefficients for a characteristic function to

minimize the edit distance, which is computationally expensive. through methods like gradient descent[217–219] or other greedy algorithms[220–222].

Although the characteristic functions can lead to successful classification, having a way to execute this classification whilst avoiding an optimization problem is difficult to attain. Future work consists of attempting to use polyfit on $\frac{1}{x}$ functions and depending on the results, progress to reaction data and attempt to match $\frac{1}{x}$ values to determine naively if the predicted role is product or reactant. This may have difficulties with constant terms due to noise and signal drift.

4.3. Machine Learning

The next avenue in pursuing reaction role classification is perusing possibilities for using clustering algorithms to classify reaction roles. Clustering algorithms are generally well-performed for classification problems[223–226]. Principle Component Analysis (PCA) was a viable candidate for clustering as it can reduce data dimensionality and reveal principle components with the largest variance[227–230]; therefore, it will be possible to explore the underlying contributing factors for classifying reaction roles. However, clustering algorithms need single vectors to cluster in the principle component space[231–234]. This is consistent for other clustering method such as k-nearest neighbour[235,236].

A popular classification algorithm for recent Kaggle data sets is random forests[237,238]. Random forests belong in a class of classifiers known as ensemble learning[239–241] and consists of many individually trained decision trees whose predicted output would be part of a representative average of all the predicted outputs. Ensemble learning leveraging

many models for one classification task to improve accuracy[242–246] and minimize overfitting. It relies on the notion that an average of many data points represents a data set better than one individual point. Through decision trees, exploratory learning of where the partitions exist set may give insight on unknown parameters amongst the reaction roles. However, ensemble methods require large datasets[246] to be accurate and there is a limited supply of reaction-based chronogram for each reaction role. This would make it difficult to expand the data density and the data would be unlabelled.

Neural networks are feasible methods to attempt training classification model reaction roles. There are 4 main categories of neural networks: 1) General purpose neural network, 2) neural network for images - Convolutional Neural Networks, 3) neural network for sequential data Recurrent Neural Networks, and 4) neural network for unsupervised learning - Deep Boltzmann Machine[247].

After encountering numerous unfeasible algorithms for the project's objectives and the available dataset, a paper by Koch et al titled "Siamese Neural Network for One-shot Image Recognition" became a superb fit for the classifier requirements[248]. One-shot training is ideal given the small dataset size based on the limited labelled data. One-shot learning is a set of machine learning algorithms that are effective for small datasets as small changes in the training data is inferred information for the network[249].

The Siamese network, as the name suggests, pairs two convolutional neural networks (CNN) together to train for purpose of identifying differences between the two networks. Originally introduced by Bromley and LeCun in 1993 for signature verification and in the same way it functions by extracting features through an artificial neural network and

obtaining the difference between the two networks[250]. In 2006, Li et al published a neural network for object detection requiring a minimalistic dataset and this model of learning was coined as one-shot learning earning over a thousand citations[251]. The neural network published by Koch combines one-shot learning with the power of differentiation from Siamese networks.

The design and implementation of this project's convolutional neural network was based on the methodology outlined by Koch et al[248]. This project's Siamese network follow a traditional triple-pass convolute and pool network with the exception that Siamese have two of the identical networks paired in tandem whereby both extract features of their respective input image. The resulting features are then passed through a distance layer through a sigmoid activation to discern the difference between the two outputs. Finally, when the distance is determined, the results are then compared to result in a match or not a match. From that, template reaction role images can be compared with the input. The class with the high probability match is predicted to be the correct label for the chronogram.

This method is powerful for identifying if two images are similar. For instance, the images of handwriting in the MNIST database can be compared against each other[252]. If an image of a 6 is compared against an image of a 7, it will identify the two as different, a low probability of a match. However, if the 6 was matched to a new image of 6 then both images are identified as the same; thus, the new image of 6 is labelled as a 6 among the categories of 0-9.

Extrapolating this idea to chronograms can assist in teaching the neural network to identify the label for a new unknown chronogram rather than 0-9 in handwriting. This method only require images of the chronograms, thus would be scalable to analytical methods other than mass spectrometry.

4.4. The Dataset

For any machine learning model, the data is the most important aspect[253,254]. Mass spectrometry is a data dense methodology with a large span of *m/z* values in each acquisition. Assuming 5 experiments with each having good results (defined by the absence of machine faults and human error during the analysis) and have an acquisition span of *m/z* 50-2000, there is 1500 *m/z* data points available. If the acquisitions have an average time span of 20 minutes at a scan rate of one scan per second, there is 60 x 20 data points per *m/z* in each experiment. Thus, one experiment would have 1500 x 60 x 20 data points. However, consider that the individual data point per scan is not useful individually, then it can be considered that there are 1500 data points per experiment and 5 experiments would yield 5 x 1500 data points. Analyzing and labeling 7500 data points is a laborious task and is difficult to outsource as there is a need for chemical knowledge and consistency. Understanding how to identify noise versus a low intensity intermediate with a slight $log(x)$ characteristic requires a trained eye.

Bearing in mind the one-shot approach for this project, the data density can be reduced to a simpler set. Taking data from titanocene oxidation, a few candidate species were chosen to be processed and labeled. Even though one-shot learning has a low data requirement, having more data can enhance the accuracy[253]. Recounting the mathematical

behaviour of the reaction role, an attempt was made to construct synthetic data with equations from section 4.2. However, the task of finding a good fit of coefficients and constants resulted in an unfruitful attempt at an optimization issue. An optimization issue is a problem that involves finding a best match to the query; however, it is difficult to find a definitive answer and can only optimize to approximate the answer[255,256].

Taking inspiration from the concept of synthetic datasets[257,258], there is a potential to create synthetic data for the training Siamese networks given the shape of the reaction roles are well known[248]. Comparing the reaction role profiles to experimental data, it was found that there are specific points in which the chronogram curve changes, corroborating with the RS findings in section 4.2. These points are analogous to control points in Beziér curves. Thus, a partial set in the training and testing datasets are synthesized by fitting the control points and Beziér curves to match the experimental trace. The control points can now translate to different x, y coordinates through Adobe Illustrator.

The experimental dataset was processed via Python and Matplotlib[259] while leveraging ProteoWizard[181] to convert mass spectrometry RAW files to human-readable text file. A lightweight python framework, Spectra.ly, described in chapter 6 was used to process the data. The chronograms are normalized to the total ion current to determine the true signal changes independent of environmental parameters and then normalized to the maximum intensity in the chronogram such that training is fixed on the shape of the change rather than the intensity of the change. Thus, all chronograms are on a 0 to 1 scale. The scales are subsequently removed leaving only the axes line and the chronogram trace to

eliminate distractions in the training set. The synthetic datasets are made based on these processed experimental sets.

4.5. Training the model

The first attempt at replicating the Koch one-shot model involved the construction of the Siamese network and then training the network with the dataset constructed using the technique outlined in section 4.4. The training and testing data was done by splitting the dataset in a 70:30 ratio. Initial testing of the methodology was done by training the network with three classes of reactions roles as a proof of concept to classifier for reactants, products, and intermediates.

Training was done on Amazon Web Service on a c4.2xlarge instance with 8 virtual cores, 34 elastic compute units, and 16 GB of RAM. CPU training was chosen over GPU due to pricing as c4.2xlarge (CPU-based compute instance) was 0.34/hour and p3.2xlarge (GPU-based computer instance) is 3.06/hour in the current time of writing this text. Training was finished in one day; however, a significant portion of metrics were not obtained such as epoch to convergence, accuracy over time. and so on[260,261]. Due to insufficient funding, this was the only attempt available. Now having the weighted model for a three-class reaction role classifier, the metrics were assessed to see how the training faired. The accuracy of the model was 31.2% after 100 iterations of training on three classes with a small batch size of 3. The accuracy was tested over 250 iterations to determine the class feature vectors. There are a few alarming issues on this attempt. 100 iterations are quite low for a one shot learning problem, prior work has used 900000 iterations[248,262]. It would be better to define the training over epochs to know exactly how

many times the dataset has been traversed and is standard to convey number of epochs for training as opposed to iterations. The 900000 iterations could be half an epoch, an epoch, or 10 epochs, without the proper metrics, it is difficult to understand the performance. Next problem is the batch size. The batch size has been shown to influence the accuracy of the model due to altering the course of the convergence[263]. Typical batch sizes are between 32 and 512[264–266]. Larger batch sizes tend to have sharper convergence meanwhile smaller batch sizes have flatter convergence and improved accuracy. This phenomenon has not been elucidated but has been demonstrated by Kesker et al[267]. The batch size of 3 may have been too small to provide any convergence. Finally, the accuracy check with 250 iterations had more iterations than training and should have been a sign to increase the training iterations. Further work on training the model would advise the use of a suitable a batch size within the standard range and train over approximately over 10-25 epochs to monitor the loss over time. It is probable that the training set was too small and the model was incorrectly configured leading to network overfitting the training data. This indicator should question which of the two L1 or L2 regularizers would be better to minimize overfitting.

4.6. Experimental

The dataset of 60 images were split at 10:5 ratio by training:validation at 66%:33%. Images were processed by python software Spectra.ly and synthetic data was made by Adobe Illustrator.

The network is as follows:

Parameters	Values
Input shape	[105,105,1]
Layer 1	64 layer @ 10x10 filter
Layer 2	Maxpool 64 layer @ 2x2 pool
Layer 3	128 layer @ 7x7 filter
Layer 4	Maxpool 64 layer @ 2x2 pool
Layer 5	128 layer @ 4x4 filter
Layer 6	Maxpool 64 layer @ 2x2 pool
Layer 7	256 layer @ 4x4 filter

Table 4.1: Outline of the Layer in each of the two CNN of the Siamese Network

The distance layer is computed to calculate difference in tensor values of each network in the Siamese network. L2 regularizer was used for the model following[262,268] the methodology in Koch et al. The output prediction layer is a single output layer with sigmoid activation for binary output; therefore, binary cross entropy was used with Adam optimization for the output. Finally, pickle was used to package and load the image data for training and validation.

4.7. Conclusion

An attempt to create a machine learning model to determine the reaction role of a chronogram trace was performed in this section of work. Initial analyses of chronogram behaviour through mathematics has offered insight on reaction behaviour but is unable to use the characteristics to determine the reaction role. This task an optimization issue of matching the characteristic equation to the experimental data.

Other statistical methods have been attempted such as PCA or Ensemble machine learning methods; however, PCA require single point vectors and ensemble methods need

large datasets. Neural network is promising as it can learn from a dataset where the vectors represent the entire chronogram as an image. CNNs are great for image recognition thus was the desired network type to train on. Koch et al Siamese network was used to train the model to identify differences in two images and by comparing the template reaction role images to experimental images, the reaction role with the highest percentage of match is the role in which the experimental trace belongs to.

Due to human error in configuring the network, the resulting accuracy yielded 31.2%. Improvements can be made by revisiting the basic understanding of network parameters and standards by which networks are built upon and can be done if there are sufficient resources to perform the training.

4.8. Automation and Control

The recent surge of machine learning papers in chemistry for navigating reaction space, how are there sufficient data to supplement training for the algorithms?[144,269–273] Automation plays a key role in obtaining a large amount of experimental data. Pharmaceutical companies such as Bristol-Myers Squibb and GlaxoSmithKline[274–276] employ a division of automated tools for synthetic and analytical purposes. Recent work on automated reaction exploration with machine learning and robotics reveal promising future for chemistry as automation may navigate a large breadth of reactions through high-throughput screening in shorter time frame than humans[271,277–280], a method in which many reactions are run concurrently (typically 96[281]) to determine pharmaceutical efficacy or surveying properties of compounds.

These concepts of automation from machine learning to robotics are equally applicable to mechanistic investigation of a reaction. Both facets of automation for software automation (applications, statistics and machine learning) and hardware automation (microcontrollers and robotics) are to be investigated in this chapter.

5. CHAPTER 5: Robotics for Mechanistic Investigation

Originally, this project was developed to integrate a robot built by a Victoria start-up company, North Robotics, for automated mechanistic investigations coupled with mass spectrometry. The start of the project was to identify the tasks that are suitable for automation which are: capping/uncapping, addition of reagents, and dispensing/aspirating reaction samples or liquid reaction quenching solutions. The tasks were design to analyze the SMCC as the initial study given the prior literature published from the McIndoe group[48]. Vials would be suitable vessels for offline analyses such as GCMS, LCMS or NMR. The ability to analyze catalysis at reaction scale in conjunction with orthogonal techniques would offer valuable information about the reaction bulk-phase. When studying the reaction by mass spectrometry, insights on the mechanistic pathways can be scrutinized due to the difference in reaction conditions from benchtop syntheses. The scale of reaction designed for online mass spectrometry techniques, like PSI-ESI-MS, are in within part-per-million (ppm) to part-per-billion (ppb) region[50,99] whereas benchtop reaction could be on the order of 1M, approximately greater than 10000 ppm depending on molar mass.

The goal of this project is to offer a mechanistic investigation setup that would serve as a proof of concept of a automated reaction exploration or interrogation platform. The combination of robotics, on-line mass spectrometry, and orthogonal techniques would bring this concept to reality. However, after 18 months North Robotics had not provided the robot so I decided to build a robotic setup to facilitate the liquid handling procedures

and following my required specifications: capping/uncapping, addition of reagents, and dispensing/aspirating reaction solutions.

5.1. Kendrick: Automated Reaction Sampler

The automated reaction sampler (“Kendrick”) was initially an autosampler for an atomic absorption spectrometer that was graciously donated by a committee member, Dr. Dennis Hore. At the start, I had to determine if the circuit boards present in the autosampler was suitable for reuse. Upon consultation with the electronics technician, Andrew MacDonald, the original printed circuit boards (PCB) were determined to be proprietary and would be difficult to understand the design from the intentional obfuscation of the circuitry. The proprietary PCBs were removed, and the motors controlled by the PCBs were isolated for future use with new aftermarket PCBs. The manufacturer’s power supply remained in the setup as there were available specification sheets to solve the pin layout.

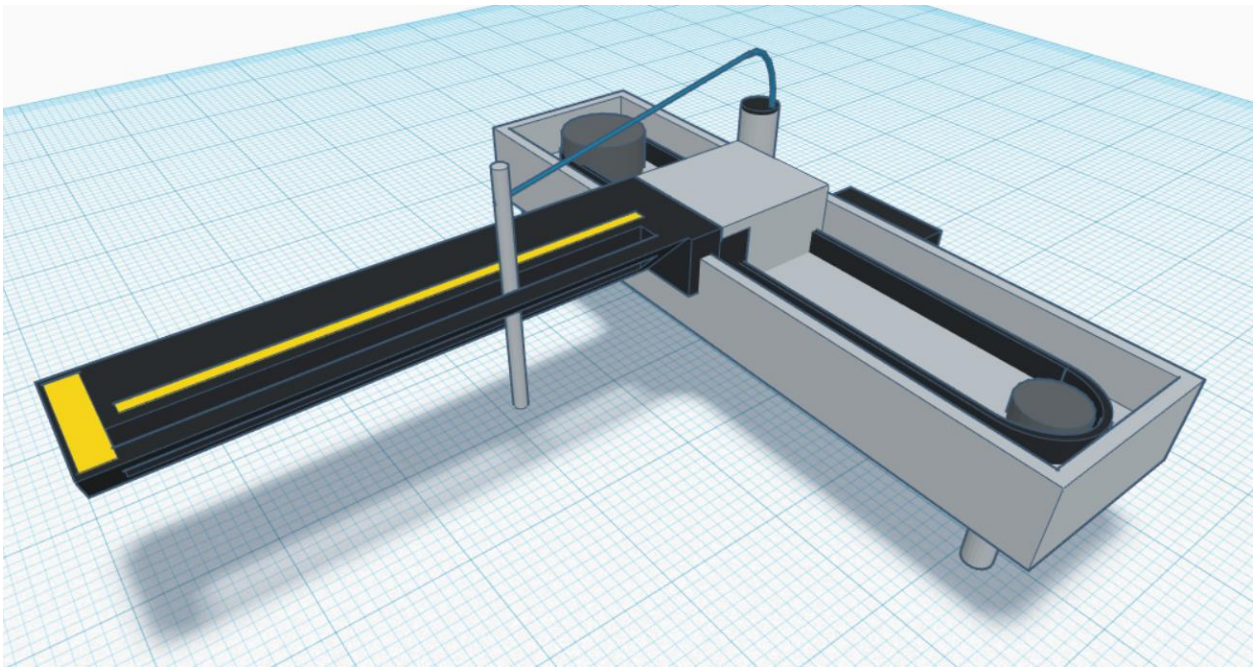


Figure 5.1: Graphical representation for the 3-axis motor setup for Kendrick

After removing the PCBs, the autosampler contained three stepper motors. Each of the motors controlled an axis in the Euclidean space. A controller would be needed to move the motors in a precise manner and combine motions between motors to achieve positional accuracy. The travel, movement along an axis[282], appeared to be similar to a 3-axis computer numerical control (CNC) machine, often used for machining complex parts. Based on this similarity, a small-scale CNC control board, made by Synthetos[283], was used to control the automated reaction sampler. The board was called TinyG and uses G-code, an industry standard for CNC maneuvers, to control the travel along the x, y, and z axes.

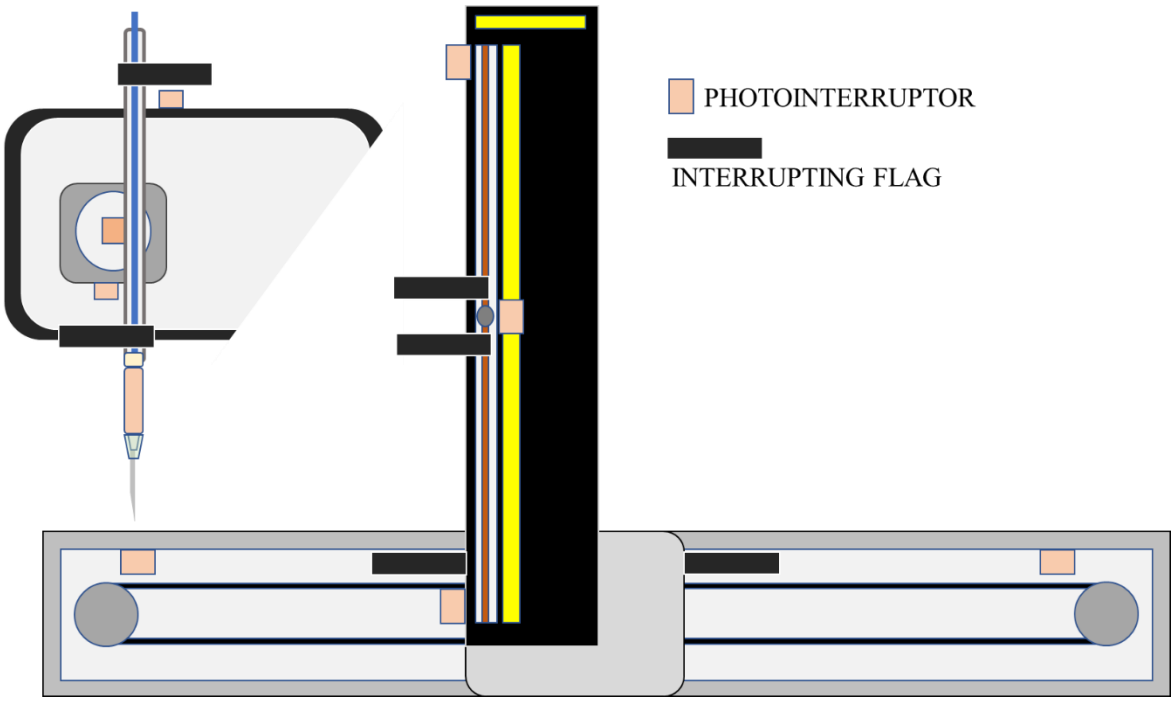


Figure 5.2: The placement location of the photointerruptors used at the minimum and maximum axes location of Kendrick

There are limits to how much travel is available for the sampling unit. If a G-code instruction forces the motor to move past the minimum and maximum of the axis, a physical limit in movement where the sampler can not move past, it would attempt to collide and push against limiter nut. Exceeding the minimum and maximum of the axis would damage the motor over time. Such errors can lead to the motor burning out and no longer functioning[284]. To prevent collisional accidents, limit switches can be triggered prior to the collision and halt the sampler immediately. TinyG contains an input ports specifically for limits switches attached to the extremes of the axes. The switches would send a signal to the board in the event of an axis limit breach and halt the motors. The active high photointerruptor switch sends 5 V to the limit switch input port of the TinyG to establish an accident is about to happen. A photointerruptor is a small device that

detect an optical obstruction between an anode and cathode and on the sampling rod is a metal flag that obstructs the photointerruptor, as shown in the figure below, to trigger the switch. In order for the photointerruptors to be in the correct position, mounts are required to ensure that the switches are maintained in the same space and the obstruction occurs reliably. 3D printed mounts were fabricated to support this concept and is demonstrated in the figure below.

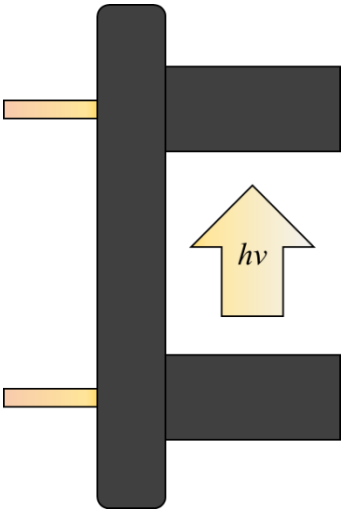


Figure 5.3: Diagram of a photointerruptor used for Kendrick

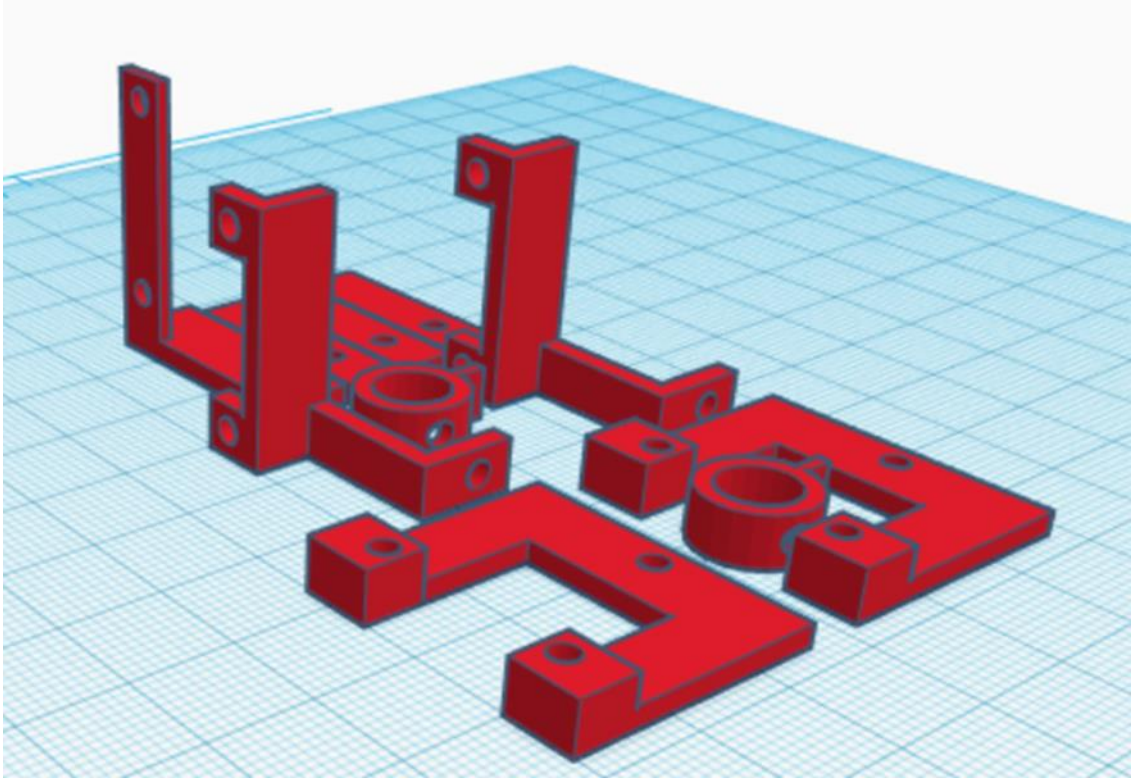


Figure 5.4: Photointerruptor mounts modeled in TinkerCAD for 3D Printing

The next step in the development was the consolidation of the signals from all the photointerruptors for diagnostic purposes. Initially all the photointerruptors are soldered on a signal bus that would relay the information to the TinyG. This design was successful and is also made into a PCB design ready for large scale fabrication for further replication of this sampling robot.

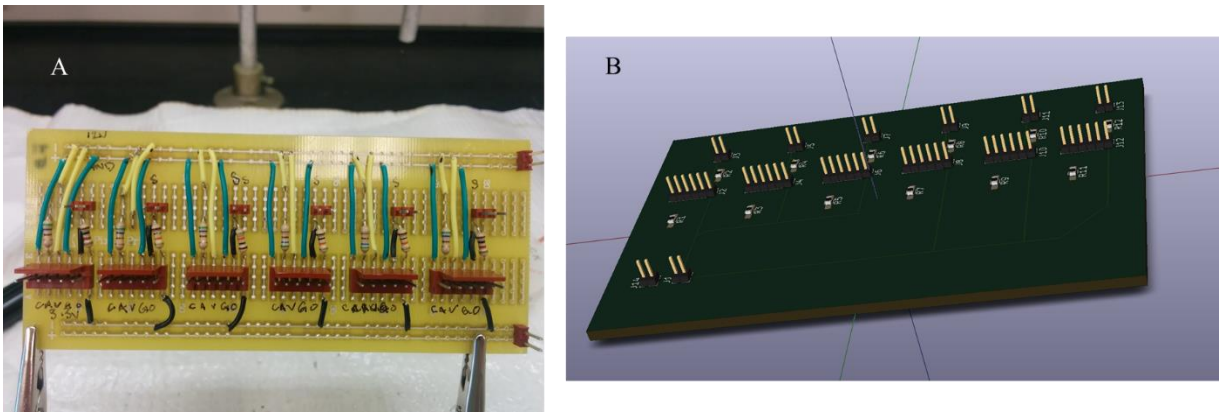


Figure 5.5: The photointerruptor board used to consolidate the signals from the limits to the input of TinyG. A: is the prototype for this PCB, B: the rendering of PCB for production of this photointerruptor bus board

Due to the modular design of the aftermarket parts from this development, the overall design of the robot is accessible to amateur engineers as the designs have abstracted most of the laborious tasks and alleviated the need to design a sampling robot from scratch. All of the parts are designed for 3D printing or 2-layer PCB fabrication thus are easily obtainable and can produced as a kit.

The control flow of Kendrick is as follows: a board-scale computer Raspberry Pi 3B acts as the controller of the robot, including the control of TinyG, and the TinyG board controls the movement along the Euclidean space. Kendrick can be accessed remotely through Secured Shell (SSH).[285] A Python adapter code translates movement to G-code[286].

The liquid dispensing aspect of the robot is located at the sampling rod show in **Error! Reference source not found.**. The sampling rod is hollow and the liquid flow through

PEEK tubing tapped with a Luer-lok fitting. Luer-lok allow the use of common disposable needles of any gauge. To engage in the motion to aspirate and dispense liquid, the VICI M6 stepper motor pump is used. Future work on attaining replicable movement is required before reaction sampling is possible as currently there is no reference point for the sampler to find its absolute position in Euclidean space. It is advised to add a homing sequence to the sampler to find a x=0, y=0, and z=0 location as a reference for all the axes.

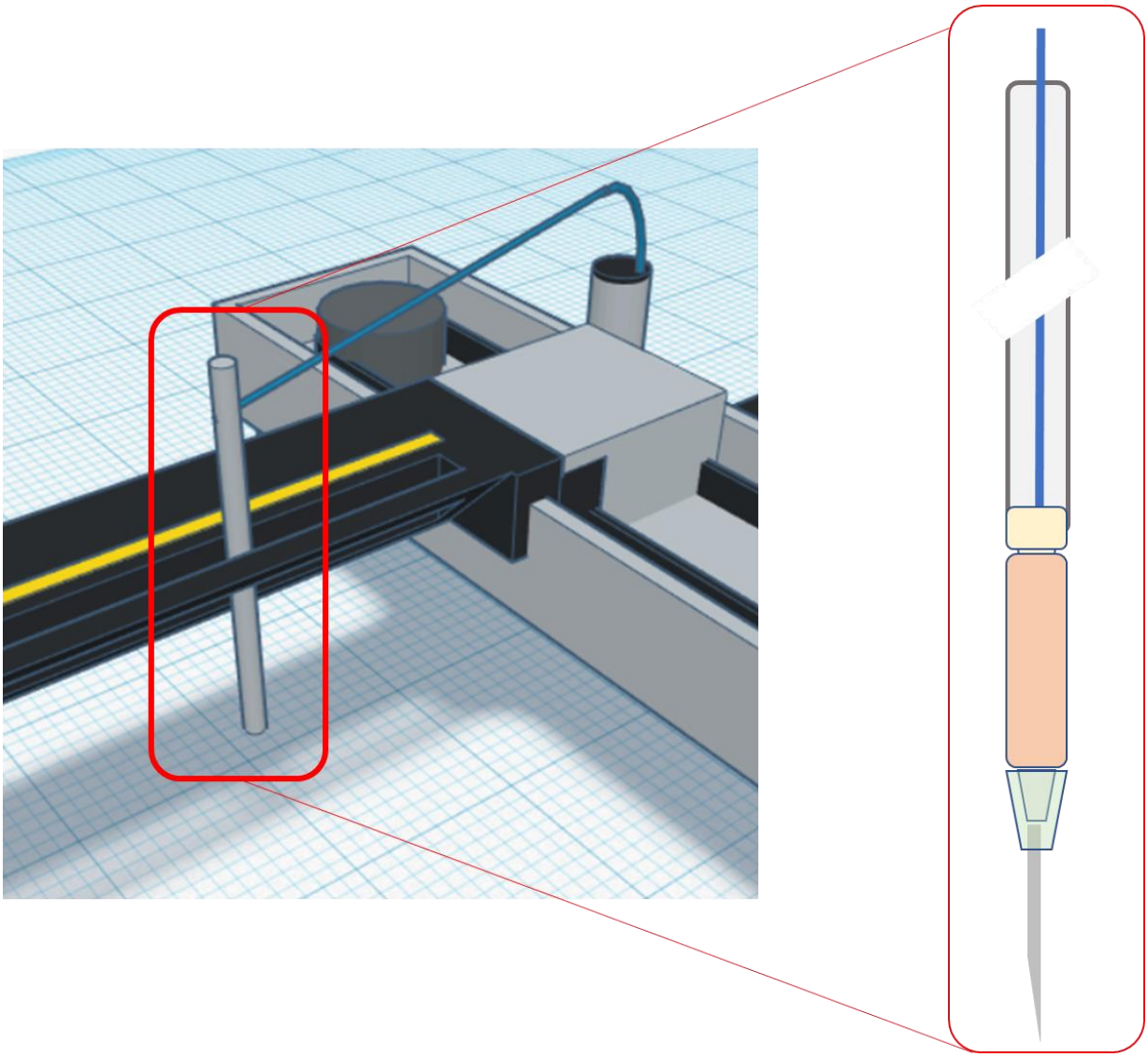


Figure 5.6: Detailed expansion of the sampling rod used for Kendrick outlining the conversion of PEEK to Luer for the integration of disposable needles

KENDRICK as a custom built system has a competitive edge against current commercial solutions. The advantage of being able to incorporate a needle to exert an inert atmosphere into a reaction vial allows for air-sensitive chemistry to be done on a benchtop environment. Current commercial systems require the robot to be placed in a glovebox to perform air-sensitive chemistry. For synthetic purposes, a glovebox containing the robotic system would be sufficient as the subsequent analyses can be

brought out through the antechamber and submitted for analysis. However, when the samples are submitted, often the integrity of the inert of atmosphere is not maintained In the synthetic instances, this is not a concern as the final product is the target of the analyses. KENDRICK has the potential to mitigate this issue and maintain the inert atmosphere upon sampling for offline analyses like NMR or IR or for online processes like PSI-ESI-MS. For continuous flow methods, the inert atmosphere injection can be continuous as well resulting in the required overpressure to perform a cannula transfer to the analytical instrument of choice. Having such added flexibility, the home-made system (KENDRICK) is of research interest to assist in the automation outlined in Figure 1.8.

6. CHAPTER 6: CONCLUSION AND FUTURE STUDIES

6.1. The Beginning and the End

My research aims incorporate automation and machine learning to the study of catalytic reaction mechanisms. The recurring theme of accessibility in content and resulting research is important as the diversity of background from the reader warrants the need to introduce unfamiliar topics.

In Chapter 1, I introduced some basic and fundamental concepts in chemistry. This served as the foundation upon which subsequent theories and ideas of this dissertation built on. The intent is to bring the reader up to speed with the necessary knowledge about chemistry and mass Spectrometry, which served as the central analytical tool in this work. The involvement of reaction-based chemistry was most significant in the Chapter 2 and 3 and more specific details were elaborated throughout the chapter.

Chapter 2 used real time mass spectrometric analysis to study the oxidation of reduced titanocene. The decomposition pathways of this reactive species was the focus of this project. The preparation of this catalyst used an acetonitrile-stabilized variant, commonly used as an indicator for oxygen-contaminated atmospheric conditions. An oxidation setup was constructed to facilitate the analysis by pressurized sample infusion-electrospray ionization mass spectrometry (PSI-ESI-MS) under inert conditions. The results indicated a methyl abstraction due to the formation of a methoxy ligand, in the absence of a methoxy moiety prior to the oxidation. An isotopically-labelling study was performed to determine and confirm that a methyl abstraction event occurred by labelling the only

methyl source in the reaction, the solvent acetonitrile. Upon oxidation, it was determined that a methyl abstraction did indeed occur by obtaining the methyl group from acetonitrile. This project has demonstrated that it is possible to study a radical air-sensitive reaction to elucidate underlying mechanistic questions.

Chapter 3 describes the development of mass spectrometry-related software. This body of work ranges from the collaboration in PythoMS, to developing independently Spectra.ly, an object-oriented library, that allows for accessible code for developers. In addition, mass spectrometric fragmentation software developments were made called Sinatra, a cloud-ready energy-dependent electrospray ionization (EDESI) platform. Sinatra strives to improve the processing workflow by mitigating the difficulties present with running code. By packaging Sinatra as a potential web application, it allows users and developers a convenient way to access a processing tool. AutoMRM is the third development in this chapter on automation, and its purpose is to remove a tedious task in analytical mass spectrometry. It provides an automated means of constructing methodologies that can quantify unique chemical compounds through multiple reaction monitoring (MRM). It has been demonstrated to be successful for Waters MassLynx; future developments should include expansion to other MS software from different MS systems.

Chapter 4 ventures into the realm of computer science, where the project was to create a reaction role classifier. Starting with the introduction of concepts and fundamental theories associated with techniques used in the project, this provides the reader with context about the purpose and methodology of the research. Using numerous

mathematical and statistical instruments have led to insights on what is needed to construct such a classifier with minimal data. Neural networks, specifically a one-shot based Siamese convolutional neural network was constructed. The network was trained using both experimental and synthetic data to 'learn' to classify the different reaction intermediates and products. In spite of the poor results in this case, many lessons were learned in this project to aid similar future endeavours. Adjusting the parameters and understanding their purpose to would help improve accuracy and minimize overfitting. An automation platform was built in an attempt to develop an automated reaction sampler. Circuit boards were personally designed and constructed alongside many 3D printed parts for the construction of the automated reaction sampler (Kendrick). Kendrick was able to perform aspiration and dispensing motions along the planes of a three-dimensional Cartesian coordinate system. Some limitations to the current model include difficulties obtaining highly accurate positioning, and this can be improved with future iterations.

Bibliography

(1) Oxtoby, D. W.; Gillis, H. P.; Butler, L. J. *Principles of Modern Chemistry*; Cengage learning, 2015.

(2) *IUPAC Compendium of Chemical Terminology*; Nič, M., Jirát, J., Košata, B., Jenkins, A., McNaught, A., Eds.; IUPAC: Research Triagle Park, NC, 2009.

(3) Wang, H.; Peslherbe, G. H.; Hase, W. L. Trajectory Studies of SN2 Nucleophilic Substitution. 4. Intramolecular and Unimolecular Dynamics of the Cl----CH3Br and ClCH3---Br- Complexes. *J. Am. Chem. Soc.* **1994**, *116* (21), 9644–9651.

(4) Graul, S. T.; Bowers, M. T. The Nonstatistical Dissociation Dynamics of Chloride(Bromomethane) Cl-(CH3Br): Evidence for Vibrational Excitation in the Products of Gas-Phase SN2 Reactions. *J. Am. Chem. Soc.* **1991**, *113* (25), 9696–9697.

(5) Graul, S. T.; Bowers, M. T. Vibrational Excitation in Products of Nucleophilic Substitution: The Dissociation of Metastable X-(CH3Y) in the Gas Phase. *J. Am. Chem. Soc.* **1994**, *116* (9), 3875–3883.

(6) Bernstein, R. B. (Richard B. *Atom-Molecule Collision Theory : A Guide for the Experimentalist*; Plenum Press, 1979.

(7) Teleman, O.; Jönsson, B.; Engström, S. A Molecular Dynamics Simulation of a Water Model with Intramolecular Degrees of Freedom. *Mol. Phys.* **1987**, *60* (1), 193–203.

(8) Wilson, E. B.; Decius, J. C.; Cross, P. C.; Sundheim, B. R. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; Courier Corporation, 1955; Vol. 102.

(9) Wang, L.; Hermans, J. Change of Bond Length in Free-energy Simulations: Algorithmic Improvements, but When Is It Necessary? *J. Chem. Phys.* **1994**, *100* (12), 9129–9139.

(10) Camuffo, D.; Camuffo, D. Consequences of the Maxwell–Boltzmann Distribution. *Microclim. Cult. Herit.* **2014**, 347–366.

(11) Sharp, K.; Matschinsky, F.; Johnson, E. R. Translation of Ludwig Boltzmann's Paper ``On the Relationship between the Second Fundamental Theorem of the Mechanical Theory of Heat and Probability Calculations Regarding the Conditions for Thermal Equilibrium'' 6pt Sitzungberichte Der Kaiserlichen Akademie Der Wissenschaften. Mathematisch-Naturwissen Classe. Abt. II, LXXVI 1877, Pp 373-435 (Wien. Ber. 1877, 76:373-435). Reprinted in Wiss. Abhandlungen, Vol. II, Reprint 42, p. 164-223, Barth, Leipzig, 1909. *Entropy* **2015**, *17*, 164–223.

(12) Pan, Y.; Zhou, H.; Mahsut, A.; Rohm, R. J.; Berejnaia, O.; Price, O.; Chen, Y.; Castro-Perez, J.; Lassman, M. E.; McLaren, D.; et al. Static and Turnover Kinetic Measurement of Protein Biomarkers Involved in Triglyceride Metabolism Including ApoB48 and ApoA5 by LC/MS/MS. *J. Lipid Res.* **2014**, *55* (6), 1179–1187.

(13) Lin, K.-C. Understanding Product Optimization: Kinetic versus Thermodynamic Control. *J. Chem. Educ.* **1988**, *65* (10), 857.

(14) Borisova, K. K.; Kvyatkovskaya, E. A.; Nikitina, E. V.; Aysin, R. R.; Novikov, R. A.; Zubkov, F. I. Classical Example of Total Kinetic and Thermodynamic Control: The Diels–Alder Reaction between DMAD and Bis-Furyl Dienes. *J. Org. Chem.* **2018**, *83* (8), 4840–4850.

(15) Sykes, P. *A Guidebook to Mechanism in Organic Chemistry*; Pearson Education India, 1986.

(16) Lasaga, A. C. Transition State Theory. *Rev. Miner. (United States)*.

(17) Leffler, J. E. Parameters for the Description of Transition States. *Science* **1953**, *117* (3039), 340–341.

(18) Hammond, G. S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955**, *77* (2), 334–338.

(19) Manz, T. A.; Sholl, D. S. A Dimensionless Reaction Coordinate for Quantifying the Lateness of Transition States. *J. Comput. Chem.* **2009**, *31* (7), NA-NA.

(20) Baraban, J. H.; Changala, P. B.; Mellau, G. C.; Stanton, J. F.; Merer, A. J.; Field, R. W. Spectroscopic Characterization of Isomerization Transition States. *Science* **2015**, *350* (6266), 1338–1342.

(21) Polanyi, J. C.; Zewail, A. H. Direct Observation of the Transition State. *Acc. Chem. Res.* **1995**, *28* (3), 119–132.

(22) Viswanathan, B. *Catalysis: Principles and Applications*; Alpha Science Int'l Ltd., 2002.

(23) Behr, A.; Neubert, P. *Applied Homogeneous Catalysis*; John Wiley & Sons, 2012.

(24) Nguyen, S. T.; Johnson, L. K.; Grubbs, R. H.; Ziller, J. W. Ring-Opening Metathesis Polymerization (ROMP) of Norbornene by a Group VIII Carbene Complex in Protic Media. *J. Am. Chem. Soc.* **1992**, *114* (10), 3974–3975.

(25) Matthias Scholl; Sheng Ding; Choon Woo Lee, and; Grubbs*, R. H. Synthesis and Activity of a New Generation of Ruthenium-Based Olefin Metathesis Catalysts Coordinated with 1,3-Dimesityl-4,5-Dihydroimidazol-2-Ylidene Ligands§. **1999**.

(26) Jason S. Kingsbury; Harrity, J. P. A.; Peter J. Bonitatebus, Jr., A.; Hoveyda, A. H. A Recyclable Ru-Based Metathesis Catalyst. **1999**.

(27) Miyaura, N.; Yamada, K.; Suzuki, A. A New Stereospecific Cross-Coupling by the Palladium-Catalyzed Reaction of 1-Alkenylboranes with 1-Alkenyl or 1-Alkynyl Halides. *Tetrahedron Lett.* **1979**, *20* (36), 3437–3440.

(28) Miyaura, N.; Suzuki, A. Stereoselective Synthesis of Arylated (E)-Alkenes by the Reaction of Alk-1-Enylboranes with Aryl Halides in the Presence of Palladium Catalyst. *J. Chem. Soc. Chem. Commun.* **1979**, *0* (19), 866.

(29) King, A. O.; Yasuda, N. Palladium-Catalyzed Cross-Coupling Reactions in the Synthesis of Pharmaceuticals; Springer, Berlin, Heidelberg; pp 205–245.

(30) Van Santen, R. Catalysis in Perspective: Historic Review. In *Catalysis*; 2012.

(31) Winter, M. J. *D-Block Chemistry*.

(32) Hieber, W.; Marin, R. Stickoxydverbindungen von Kobalt(I)-Halogeniden. *Zeitschrift für Anorg. und Allg. Chemie* **1939**, *240* (3), 241–260.

(33) Cornils, B.; Herrmann, W. A.; Rasch, M. Otto Roelen, Pioneer in Industrial Homogeneous Catalysis. *Angew. Chemie Int. Ed. English* **1994**, *33* (21), 2144–2163.

(34) Heck, R. F.; Breslow, D. S. *The Reaction of Cobalt Hydrotetracarbonyl with Olefins*; 1961.

(35) Salinas-Olvera, J. P.; Gómez, R. M.; Cortés-Guzmán, F. Structural Evolution: Mechanism of Olefin Insertion in Hydroformylation Reaction. **2008**.

(36) Skoog, D. A.; Holler, F. J.; Nieman, T. A. Principles of Instrumental Analysis. *Clin. Chem. Ed.* **1994**, *40* (8), 1612.

(37) Thomas, A. A.; Denmark, S. E. Pre-Transmetalation Intermediates in the Suzuki-Miyaura Reaction Revealed: The Missing Link. *Science* **2016**, *352* (6283), 329–332.

(38) Seeman, J. I. Ernest L. Eliel: A Life of Purpose, Determination, and Integrity. *Chirality* **2002**, *14* (2–3), 98–109.

(39) Frye, S. V.; Eliel, E. L.; Cloux, R. Rapid-Injection Nuclear Magnetic Resonance Investigation of the Reactivity of .Alpha.- and .Beta.-Alkoxy Ketones with Dimethylmagnesium: Kinetic Evidence for Chelation. *J. Am. Chem. Soc.* **1987**, *109* (6), 1862–1863.

(40) Thomas, A. A.; Zahrt, A. F.; Delaney, C. P.; Denmark, S. E. Elucidating the Role of the Boronic Esters in the Suzuki–Miyaura Reaction: Structural, Kinetic, and Computational Investigations. *J. Am. Chem. Soc.* **2018**, *140* (12), 4401–4416.

(41) Thomas, A. A.; Denmark, S. E. Ernest L. Eliel, a Physical Organic Chemist with the Right Tool for the Job: Rapid Injection Nuclear Magnetic Resonance; 2017; pp 105–134.

(42) Thomas, A. A.; Wang, H.; Zahrt, A. F.; Denmark, S. E. Structural, Kinetic, and Computational Characterization of the Elusive Arylpalladium(II)Boronate Complexes in the Suzuki–Miyaura Reaction. *J. Am. Chem. Soc.* **2017**, *139* (10), 3805–3821.

(43) Eliel, E. L.; Gianni, M. H. Conformational Equilibria by Nuclear Magnetic Resonance Spectroscopy. *Tetrahedron Lett.* **1962**, *3* (3), 97–101.

(44) Eliel, E. L.; Ro, R. S. Conformational Analysis. III. Epimerization Equilibria of Alkylcyclohexanols [1]. *J. Am. Chem. Soc.* **1957**, *79* (22), 5992–5994.

(45) Naito, S.; Aida, S.; Kasahara, T.; Miyao, T. Infrared Spectroscopic Study on the Reaction Mechanism of CO Hydrogenation over Pd/CeO2. *Res. Chem. Intermed.* **2006**, *32* (3–4), 279–290.

(46) Preston T. Snee; Haw Yang; Kenneth T. Kotz; Christine K. Payne, and; Harris*, C. B. Ultrafast Infrared Studies of the Reaction Mechanism of Silicon−Hydrogen Bond Activation by H5-CpV(CO)4. **1999**.

(47) Zscherp, C.; Barth, A. Reaction-Induced Infrared Difference Spectroscopy for the Study of Protein Reaction Mechanisms†. **2001**.

(48) Yunker, L. P. E.; Ahmadi, Z.; Logan, J. R.; Wu, W.; Li, T.; Martindale, A.; Oliver, A. G.; McIndoe, J. S. Real-Time Mass Spectrometric Investigations into the Mechanism of the Suzuki–Miyaura Reaction. *Organometallics* **2018**, *37* (22), 4297–4308.

(49) Ewan, H. S.; Iyer, K.; Hyun, S.-H.; Wleklinski, M.; Cooks, R. G.; Thompson, D. H. Multistep Flow Synthesis of Diazepam Guided by Droplet-Accelerated Reaction Screening with Mechanistic Insights from Rapid Mass Spectrometry Analysis. *Org. Process Res. Dev.* **2017**, *21* (10), 1566–1570.

(50) Yan, X.; Sokol, E.; Li, X.; Li, G.; Xu, S.; Cooks, R. G. On-Line Reaction Monitoring and Mechanistic Studies by Mass Spectrometry: Negishi Cross-Coupling, Hydrogenolysis, and Reductive Amination. *Angew. Chemie* **2014**, *126* (23), 6041–6045.

(51) Santos, L. S.; Rosso, G. B.; Pilli, R. A.; Eberlin, M. N. The Mechanism of the Stille Reaction Investigated by Electrospray Ionization Mass Spectrometry. **2007**.

(52) Ahmadi, Z.; Yunker, L. P. E.; Oliver, A. G.; McIndoe, J. S. Mechanistic Features

of the Copper-Free Sonogashira Reaction from ESI-MS. *Dalt. Trans.* **2015**, *44* (47), 20367–20375.

(53) Stoddard, R. L.; Luo, J.; van der Wal, N.; O'Rourke, N. F.; Wulff, J. E.; McIndoe, J. S. A Multi-Pronged Mechanistic Study of the Phosphine-Mediated Conjugate Addition of an Alcohol to an Acetylenic Ester. *New J. Chem.* **2014**, *38* (11), 5382–5390.

(54) Luo, J.; Oliver, A. G.; Scott McIndoe, J. A Detailed Kinetic Analysis of Rhodium-Catalyzed Alkyne Hydrogenation. *Dalt. Trans.* **2013**, *42* (31), 11312.

(55) Ahmadi, Z.; Oliver, A. G.; McIndoe, J. S. An Unexpected Pathway for Ligand Substitution in an Aryl Halide Complex of Palladium. *Chempluschem* **2013**, *78* (7), 632–635.

(56) Sharma, K. S. Mass Spectrometry—The Early Years. *Int. J. Mass Spectrom.* **2013**, *349–350*, 3–8.

(57) Thomson, J. J. *Carriers of Negative Electricity*.

(58) Squires, G. Francis Aston and the Mass Spectrograph. *J. Chem. Soc. Dalt. Trans.* **1998**, *0* (23), 3893–3900.

(59) Aston, F. W.; Baxter, G. P.; Brauner, B.; Debierne, A.; Leduc, A.; Richards, T. W.; Soddy, F.; Urbain, G. REPORT OF THE INTERNATIONAL COMMITTEE ON CHEMICAL ELEMENTS: 1923. *J. Am. Chem. Soc.* **1923**, *45* (4), 867–874.

(60) Gross, J. H. *Mass Spectrometry : A Textbook*; Springer, 2010.

(61) Kebarle, P.; Tang, L. From Ions in Solution to Ions in the Gas Phase - the Mechanism of Electrospray Mass Spectrometry. *Anal. Chem.* **1993**, *65* (22), 972A–986A.

(62) Maher, S.; Syed, S. U.; Hughes, D. M.; Gibson, J. R.; Taylor, S. Mapping the Stability Diagram of a Quadrupole Mass Spectrometer with a Static Transverse Magnetic Field Applied. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (8), 1307–1314.

(63) Keller, B. O.; Sui, J.; Young, A. B.; Whittal, R. M. Interferences and Contaminants Encountered in Modern Mass Spectrometry. *Anal. Chim. Acta* **2008**, *627* (1), 71–81.

(64) Weber, R. J. M.; Li, E.; Bruty, J.; He, S.; Viant, M. R. MaConDa: A Publicly Accessible Mass Spectrometry Contaminants Database. *Bioinformatics* **2012**, *28* (21), 2856–2857.

(65) Kim, S.; Rodgers, R. P.; Marshall, A. G. Truly "Exact" Mass: Elemental Composition Can Be Determined Uniquely from Molecular Mass Measurement at ~0.1 MDa Accuracy for Molecules up to ~500 Da. *Int. J. Mass Spectrom.* **2006**, *251* (2–3), 260–265.

(66) Brenton, A. G.; Godfrey, A. R. Accurate Mass Measurement: Terminology and Treatment of Data. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (11), 1821–1835.

(67) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrom. Rev.* **1998**, *17* (1), 1–35.

(68) Marshall, A. G.; Hendrickson, C. L. Fourier Transform Ion Cyclotron Resonance Detection: Principles and Experimental Configurations. *Int. J. Mass Spectrom.* **2002**, *215* (1–3), 59–75.

(69) Comisarow, M. B.; Marshall, A. G. Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **1974**, *25* (2), 282–283.

(70) Fourier Transform Ion Cyclotron Resonance (FT-ICR) - MagLab https://nationalmaglab.org/education/magnet-academy/watch-play/interactive/fourier-transform-ion-cyclotron-resonance-ft-icr (accessed Dec 11, 2018).

(71) Griffiths, D. J. *Introduction to Electrodynamics*; Prentice Hall New Jersey, 1962.

(72) Lemaire, J.; Thomas, S.; Lopes, A.; Louarn, E.; Mestdagh, H.; Latappy, H.; Leprovost, J.; Heninger, M. Compact FTICR Mass Spectrometry for Real Time Monitoring of Volatile Organic Compounds. *Sensors (Switzerland)* **2018**, *18* (5), 1–15.

(73) Mass Spectrometer, March 29, 1996.

(74) Perry, R. H.; Cooks, R. G.; Noll, R. J. Orbitrap Mass Spectrometry: Instrumentation, Ion Motion and Applications. *Mass Spectrom. Rev.* **2008**, *27* (6), 661–699.

(75) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. The Orbitrap: A New Mass Spectrometer. *J. Mass Spectrom.* **2005**, *40* (4), 430–443.

(76) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. **2000**.

(77) Scigelova, M.; Hornshaw, M.; Giannakopulos, A.; Makarov, A. Fourier Transform Mass Spectrometry. *Mol. Cell. Proteomics* **2011**, *10* (7), M111.009431.

(78) Brewer, R. G.; Shoemaker, R. L. Optical Free Induction Decay. *Phys. Rev. A* **1972**, *6* (6), 2001–2007.

(79) Verma, T. S.; Levine, S. N.; Meng, T. H. Y. *Transient Modeling Synthesis: A Exible Analysis/Synthesis Tool for Transient Signals*.

(80) Mellon, F. A. MASS SPECTROMETRY | Principles and Instrumentation. *Encycl. Food Sci. Nutr.* **2003**, 3739–3749.

(81) Somogyi, Á. Mass Spectrometry Instrumentation and Techniques. *Med. Appl. Mass Spectrom.* **2008**, 93–140.

(82) Miller, P. E.; Denton, M. B. The Quadrupole Mass Filter: Basic Operating Concepts. *J. Chem. Educ.* **1986**, *63* (7), 617.

(83) Konenkov, N. V.; Sudakov, M.; Douglas, D. J. Matrix Methods for the Calculation of Stability Diagrams in Quadrupole Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (6), 597–613.

(84) Konenkov, N. V.; Kratenko, V. I. Characteristics of a Quadrupole Mass Filter in the Separation Mode of a Few Stability Regions. *Int. J. Mass Spectrom. Ion Process.* **1991**, *108* (2–3), 115–136.

(85) McLafferty, F. Tandem Mass Spectrometry. *Science (80-.).* **1981**, *214* (4518), 280–287.

(86) Johnson, J. V.; Yost, R. A.; Kelley, P. E.; Bradford, D. C. Tandem-in-Space and Tandem-in-Time Mass Spectrometry: Triple Quadrupoles and Quadrupole Ion Traps. *Anal. Chem.* **1990**, *62* (20), 2162–2172.

(87) Price, P. Standard Definitions of Terms Relating to Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1991**, *2* (4), 336–348.

(88) Fenn, J. B. Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angew. Chemie Int. Ed.* **2003**, *42* (33), 3871–3894.

(89) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459.

(90) Taylor, G. Disintegration of Water Drops in an Electric Field. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1964**, *280* (1382), 383–397.

(91) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9.

(92) Wilm, M. Principles of Electrospray Ionization. *Mol. Cell. Proteomics* **2011**, *10* (7).

(93) Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; Ng, K. F.; Suen, M. W. M.; Tai, H. L. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin. Biochem. Rev.* **2003**, *24* (1), 3–12.

(94) Hesketh, A. V.; Nowicki, S.; Baxter, K.; Stoddard, R. L.; McIndoe, J. S. Simplified Real-Time Mass Spectrometric Analysis of Reactions. *Organometallics* **2015**, *34* (15), 3816–3819.

(95) Janusson, E.; Hesketh, A. V.; Bamford, K. L.; Hatlelid, K.; Higgins, R.; McIndoe, J. S. Spatial Effects on Electrospray Ionization Response. *Int. J. Mass Spectrom.* **2015**, *388*, 1–8.

(96) Yunker, L. P. E.; Stoddard, R. L.; McIndoe, J. S. Practical Approaches to the ESI-MS Analysis of Catalytic Reactions. *J. Mass Spectrom.* **2014**, *49* (1), 1–8.

(97) Vikse, K. L.; Ahmadi, Z.; Luo, J.; van der Wal, N.; Daze, K.; Taylor, N.; McIndoe, J. S. Pressurized Sample Infusion: An Easily Calibrated, Low Volume Pumping System for ESI-MS Analysis of Reactions. *Int. J. Mass Spectrom.* **2012**, *323–324*, 8–13.

(98) Zhu, H.; Janusson, E.; Luo, J.; Piers, J.; Islam, F.; McGarvey, G. B.; Oliver, A. G.;

Granot, O.; McIndoe, J. S. Phenol-Selective Mass Spectrometric Analysis of Jet Fuel. *Analyst* **2017**, *142* (17), 3278–3284.

(99) Janusson, E.; Zijlstra, H. S.; Nguyen, P. P. T.; MacGillivray, L.; Martelino, J.; McIndoe, J. S. Real-Time Analysis of Pd 2 (Dba) 3 Activation by Phosphine Ligands. *Chem. Commun.* **2017**, *53* (5), 854–856.

(100) Dean, N. L.; McIndoe, J. S. Fluoride-Mediated Rearrangement of Phenylfluorosilanes. *Can. J. Chem.* **2018**, *96* (6), 587–590.

(101) Vikse, K. L.; Woods, M. P.; McIndoe, J. S. Pressurized Sample Infusion for the Continuous Analysis of Air- And Moisture-Sensitive Reactions Using Electrospray Ionization Mass Spectrometry. *Organometallics* **2010**, *29* (23), 6615–6618.

(102) Burgmayer, S. J. N. Use of a Titanium Metallocene as a Colorimetric Indicator for Learning Inert Atmosphere Techniques. *J. Chem. Educ.* **1998**, *75* (4), 460.

(103) Yeung, D.; Penafiel, J.; Zijlstra, H. S.; McIndoe, J. S. Oxidation of Titanocene(III): The Deceptive Simplicity of a Color Change. *Inorg. Chem.* **2018**, *57* (1), 457–461.

(104) Birmingham, J. M.; Fischer, A. K.; Wilkinson, G. The Reduction of Bis-Cyclopentadienyl Compounds. *Naturwissenschaften* **1955**, *42* (4), 96–96.

(105) Wilkinson, G.; Birmingham, J. M. Bis-Cyclopentadienyl Compounds of Ti, Zr, V, Nb and Ta. *J. Am. Chem. Soc.* **1954**, *76* (17), 4281–4284.

(106) Manzer, L. E.; Mintz, E. A.; Marks, T. J. 18. Cyclopentadienyl Complexes of Titanium(III) and Vanadium(III); Wiley-Blackwell, 2007; pp 84–86.

(107) Nugent, W. A.; RajanBabu, T. V. Transition-Metal-Centered Radicals in Organic Synthesis. Titanium(III)-Induced Cyclization of Epoxy Olefins. *J. Am. Chem. Soc.* **1988**, *110* (25), 8561–8562.

(108) Cangönül, A.; Behlendorf, M.; Gansäuer, A.; van Gastel, M. Radical-Based Epoxide Opening by Titanocenes. *Inorg. Chem.* **2013**, *52* (20), 11859–11866.

(109) Wu, X.; Hao, W.; Ye, K.-Y.; Jiang, B.; Pombar, G.; Song, Z.; Lin, S. Ti-Catalyzed Radical Alkylation of Secondary and Tertiary Alkyl Chlorides Using Michael Acceptors. *J. Am. Chem. Soc.* **2018**, *140* (44), 14836–14843.

(110) Speckmeier, E.; Fischer, T. G.; Zeitler, K. A Toolbox Approach To Construct Broadly Applicable Metal-Free Catalysts for Photoredox Chemistry: Deliberate Tuning of Redox Potentials and Importance of Halogens in Donor–Acceptor Cyanoarenes. *J. Am. Chem. Soc.* **2018**, *140* (45), 15353–15365.

(111) Streuff, J.; Feurer, M.; Frey, G.; Steffani, A.; Kacprzak, S.; Weweler, J.; Leijendekker, L. H.; Kratzert, D.; Plattner, D. A. Mechanism of the Ti^{III} -Catalyzed Acyloin-Type Umpolung: A Catalyst-Controlled Radical Reaction. *J. Am. Chem. Soc.* **2015**, *137* (45), 14396–14405.

(112) Gansäuer, A.; Barchuk, A.; Keller, F.; Schmitt, M.; Grimme, S.; Gerenkamp, M.; Mück-Lichtenfeld, C.; Daasbjerg, K.; Svith, H. Mechanism of Titanocene-Mediated Epoxide Opening through Homolytic Substitution. **2007**.

(113) Zhao, Y.; Weix, D. J. Enantioselective Cross-Coupling of *Meso* -Epoxides with Aryl Halides. *J. Am. Chem. Soc.* **2015**, *137* (9), 3237–3240.

(114) Gansäuer, A.; Justicia, J.; Fan, C.-A.; Worgull, D.; Piestert, F. Reductive C–C Bond Formation after Epoxide Opening via Electron Transfer. In *Metal Catalyzed Reductive C–C Bond Formation*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp 25–52.

(115) Gansäuer, A.; Hildebrandt, S.; Michelmann, A.; Dahmen, T.; von Laufenberg, D.;

Kube, C.; Fianu, G. D.; Flowers, R. A. Cationic Titanocene(III) Complexes for Catalysis in Single-Electron Steps. *Angew. Chemie Int. Ed.* **2015**, *54* (24), 7003–7006.

(116) Hildebrandt, S.; Gansäuer, A. Synthesis of Dihydropyrrolizine and Tetrahydroindolizine Scaffolds from Pyrroles by Titanocene(III) Catalysis. *Angew. Chemie* **2016**, *128* (33), 9871–9874.

(117) Zhang, Y.-Q.; Vogelsang, E.; Qu, Z.-W.; Grimme, S.; Gansäuer, A. Titanocene-Catalyzed Radical Opening of N-Acylated Aziridines. *Angew. Chemie Int. Ed.* **2017**, *56* (41), 12654–12657.

(118) Richrath, R. B.; Olyschläger, T.; Hildebrandt, S.; Enny, D. G.; Fianu, G. D.; Flowers, R. A.; Gansäuer, A. Cp_2TiX Complexes for Sustainable Catalysis in Single-Electron Steps. *Chem. - A Eur. J.* **2018**, *24* (24), 6371–6379.

(119) Jungst, R.; Sekutowski, D.; Davis, J.; Luly, M.; Stucky, G. Structural and Magnetic Properties of Di-.Mu.-Chloro-Bis[Bis(.Eta.5-Cyclopentadienyl)Titanium(III)] and Di-.Mu.-Bromo-Bis[Bis(.Eta.5-Methylcyclopentadienyl)Titanium(III)]. *Inorg. Chem.* **1977**, *16* (7), 1645–1655.

(120) RajanBabu, T. V.; Nugent, W. A. Selective Generation of Free Radicals from Epoxides Using a Transition-Metal Radical. A Powerful New Tool for Organic Synthesis. *J. Am. Chem. Soc.* **1994**, *116* (3), 986–997.

(121) RajanBabu, T. V.; Nugent, W. A.; Beattie, M. S. Free Radical-Mediated Reduction and Deoxygenation of Epoxides. *J. Am. Chem. Soc.* **1990**, *112* (17), 6408–6409.

(122) Wu, Y.; Wang, X.; Luo, Y.; Wang, J.; Jian, Y.; Sun, H.; Zhang, G.; Zhang, W.; Gao, Z. Solvent Strategy for Unleashing the Lewis Acidity of Titanocene

Dichloride for Rapid Mannich Reactions. *RSC Adv.* **2016**, *6* (19), 15298–15303.

(123) Henriques, D. S. G.; Zimmer, K.; Klare, S.; Meyer, A.; Rojo-Wiechel, E.; Bauer, M.; Sure, R.; Grimme, S.; Schiemann, O.; Flowers, R. A.; et al. Highly Active Titanocene Catalysts for Epoxide Hydrosilylation: Synthesis, Theory, Kinetics, EPR Spectroscopy. *Angew. Chemie Int. Ed.* **2016**, *55* (27), 7671–7675.

(124) Kuntzleman, T. S.; Jacobson, E. C. Teaching Beer's Law and Absorption Spectrophotometry with a Smart Phone: A Substantially Simplified Protocol. *J. Chem. Educ.* **2016**, *93* (7), 1249–1252.

(125) Knutson, T. R.; Knutson, C. M.; Mozzetti, A. R.; Campos, A. R.; Haynes, C. L.; Penn, R. L. A Fresh Look at the Crystal Violet Lab with Handheld Camera Colorimetry. *J. Chem. Educ.* **2015**, *92* (10), 1692–1695.

(126) Kehoe, E.; Penn, R. L. Introducing Colorimetric Analysis with Camera Phones and Digital Cameras: An Activity for High School or General Chemistry. *J. Chem. Educ.* **2013**, *90* (9), 1191–1195.

(127) Gee, C. T.; Kehoe, E.; Pomerantz, W. C. K.; Penn, R. L. Quantifying Protein Concentrations Using Smartphone Colorimetry: A New Method for an Established Test. *J. Chem. Educ.* **2017**, *94* (7), 941–945.

(128) Dangkulwanich, M.; Kongnithigarn, K.; Aurnoppakhun, N. Colorimetric Measurements of Amylase Activity: Improved Accuracy and Efficiency with a Smartphone. *J. Chem. Educ.* **2018**, *95* (1), 141–145.

(129) Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85* (11), 5288–5296.

(130) Roesky, H. W.; Haiduc, I.; Hosmane, N. S. Organometallic Oxides of Main Group

and Transition Elements Downsizing Inorganic Solids to Small Molecular Fragments. *Chem. Rev.* **2003**, *103* (7), 2579–2596.

(131) Heshmatpour, F.; Wocadlo, S.; Massa, W.; Dehnicke, K.; Bottomley, F.; Day, R. W. Organometallische Oxide: Die Kristallstruktur von [(η-C5H5) Ti]8 (μ-O)12/Organometallic Oxides: The Crystal Structure of [(η-C5H5) Ti]8 (μ-O)12. *Zeitschrift für Naturforsch. B* **1994**, *49* (6), 827–830.

(132) Huffman, J. C.; Stone, J. G.; Krusell, W. C.; Caulton, K. G. Homogeneous Hydrogenation of Carbon Monoxide. *J. Am. Chem. Soc.* **1977**, *99* (17), 5829–5831.

(133) Flores, J. C.; Mena, M.; Royo, P.; Serrano, R. The Formation of Acetone Complexes from the Reaction of CO with [Ti (C 5 Me 5) MeY]2 (μ-O)(Y= Me, Cl) and Their Decomposition Reactions. *J. Chem. Soc. Chem. Commun.* **1989**, No. 10, 617–618.

(134) Skapski, A. C.; Troughton, P. G. H. The Crystal and Molecular Structure of Cyclotetra [μ-Oxo-Chloro-π-Cyclopentadienyltitanium (IV)]. *Acta Crystallogr. Sect. B* **1970**, *26* (6), 716–722.

(135) Klein, H. P.; Thewalt, U.; Doppert, K.; SANCHEZDELGADO, R. DETERMINATION, STRUCTURE AND PROPERTIES OF THE TRINUCLEAR TITANIUM COMPLEX (PI-C5H5) TICL-O-TI (PI-C5H5) CL-O-TICL-(PI-C5H5) 2. CHCL3. *J. Organomet. Chem.* **1982**, *236* (2), 189–195.

(136) Honold, B.; Thewalt, U.; Herberhold, M.; Alt, H. G.; Kool, L. B.; Rausch, M. D. BINUCLEAR DICYCLOPENTADIENYLTITANIUM COMPLEXES WITH OXYGEN BRIDGES-THE STRUCTURE OF [CP2TI] 2 (MU-O). *J. Organomet.*

Chem. **1986**, *314* (1–2), 105–111.

(137) Aguado-Ullate, S.; Carbó, J. J.; González-del Moral, O.; Martín, A.; Mena, M.; Poblet, J.-M.; Santamaría, C. Ammonia Activation by M3-Alkylidyne Fragments Supported on a Titanium Molecular Oxide Model. *Inorg. Chem.* **2011**, *50* (13), 6269–6279.

(138) Eid, S.; Zalewski, A.; Smieško, M.; Ernst, B.; Vedani, A.; Eid, S.; Zalewski, A.; Smieško, M.; Ernst, B.; Vedani, A. A Molecular-Modeling Toolbox Aimed at Bridging the Gap between Medicinal Chemistry and Computational Sciences. *Int. J. Mol. Sci.* **2013**, *14* (1), 684–700.

(139) Metzger, R. M. (Robert M. .; Wiley InterScience (Online service). *The Physical Chemist's Toolbox*; Wiley, 2012.

(140) Boström, J.; Brown, D. G.; Young, R. J.; Keserü, G. M. Expanding the Medicinal Chemistry Synthetic Toolbox. *Nat. Rev. Drug Discov.* **2018**, *17* (10), 709–727.

(141) Roughley, S. D.; Jordan, A. M. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *J. Med. Chem.* **2011**, *54* (10), 3451–3479.

(142) Goh, G. B.; Hodas, N. O.; Vishnu, A. Deep Learning for Computational Chemistry. *J. Comput. Chem.* **2017**, *38* (16), 1291–1307.

(143) Besnard, J.; Ruda, G. F.; Setola, V.; Abecassis, K.; Rodriguiz, R. M.; Huang, X.-P.; Norval, S.; Sassano, M. F.; Shin, A. I.; Webster, L. A.; et al. Automated Design of Ligands to Polypharmacological Profiles. *Nature* **2012**, *492* (7428), 215–220.

(144) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity.

Nature **2018**, *559* (7714), 377–381.

(145) Modeling Programmable Logic Controllers for Logic Verification. *IEEE Control Syst.* **1994**, *14* (2), 53–59.

(146) David, R. Grafcet: A Powerful Tool for Specification of Logic Controllers. *IEEE Trans. Control Syst. Technol.* **1995**, *3* (3), 253–268.

(147) Han, R.; Guo, L.; Ghanem, M. M.; Guo, Y. Lightweight Resource Scaling for Cloud Applications. In *2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012)*; IEEE, 2012; pp 644–651.

(148) El-Rewini, H.; Abd-El-Barr, M. *Advanced Computer Architecture and Parallel Processing*; John Wiley, 2005.

(149) French, W. R.; Zimmerman, L. J.; Schilling, B.; Gibson, B. W.; Miller, C. A.; Townsend, R. R.; Sherrod, S. D.; Goodwin, C. R.; McLean, J. A.; Tabb, D. L. Wavelet-Based Peak Detection and a New Charge Inference Procedure for MS/MS Implemented in ProteoWizard's MsConvert. *J. Proteome Res.* **2015**, *14* (2), 1299–1307.

(150) Holman, J. D.; Tabb, D. L.; Mallick, P. Employing ProteoWizard to Convert Raw Mass Spectrometry Data. *Curr. Protoc. Bioinforma.* **2014**, *46*, 13.24.1-9.

(151) Kochen, M. A.; Chambers, M. C.; Holman, J. D.; Nesvizhskii, A. I.; Weintraub, S. T.; Belisle, J. T.; Islam, M. N.; Griss, J.; Tabb, D. L. Greazy: Open-Source Software for Automated Phospholipid Tandem Mass Spectrometry Identification. *Anal. Chem.* **2016**, *88* (11), 5733–5741.

(152) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A Cross-Platform

Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–920.

(153) Yunker, L.; Yeung, D.; McIndoe, J. S. PythoMS: A Python Framework to Simplify and Assist in the Processing and Interpretation of Mass Spectrometric Data. **2018**.

(154) Usage Statistics of JavaScript for Websites, December 2018 https://w3techs.com/technologies/details/cp-javascript/all/all (accessed Dec 11, 2018).

(155) Stack Overflow Developer Survey 2018 https://insights.stackoverflow.com/survey/2018#most-popular-technologies (accessed Dec 11, 2018).

(156) Could Python's Popularity Outperform JavaScript in the Next Five Years? https://hackernoon.com/could-pythons-popularity-outperform-javascript-in-the-next-five-years-abed4e307224 (accessed Dec 11, 2018).

(157) Electron | Build cross platform desktop apps with JavaScript, HTML, and CSS. https://electronjs.org/ (accessed Dec 11, 2018).

(158) Build Apps with JavaScript | Meteor https://www.meteor.com/ (accessed Dec 11, 2018).

(159) Angular https://angular.io/ (accessed Dec 11, 2018).

(160) Zehethofer, N.; Pinto, D. M.; Volmer, D. A. Plasma Free Fatty Acid Profiling in a Fish Oil Human Intervention Study Using Ultra-Performance Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22* (13), 2125–2133.

(161) Duncan, K. D.; Volmer, D. A.; Gill, C. G.; Krogh, E. T. Rapid Screening of Carboxylic Acids from Waste and Surface Waters by ESI-MS/MS Using Barium Ion Chemistry and On-Line Membrane Sampling. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (3), 443–450.

(162) Zhang, W.; Zhao, P. X. Quality Evaluation of Extracted Ion Chromatograms and Chromatographic Peaks in Liquid Chromatography/Mass Spectrometry-Based Metabolomics Data. *BMC Bioinformatics* **2014**, *15 Suppl 11* (Suppl 11), S5.

(163) Berendsen, B. J. A.; Wegh, R. S.; Meijer, T.; Nielen, M. W. F. The Assessment of Selectivity in Different Quadrupole-Orbitrap Mass Spectrometry Acquisition Modes. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (2), 337–346.

(164) Tao, W. A.; Wollscheid, B.; O'Brien, R.; Eng, J. K.; Li, X.; Bodenmiller, B.; Watts, J. D.; Hood, L.; Aebersold, R. Quantitative Phosphoproteome Analysis Using a Dendrimer Conjugation Chemistry and Tandem Mass Spectrometry. *Nat. Methods* **2005**, *2* (8), 591–598.

(165) Yao, M.; Ma, L.; Humphreys, W. G.; Zhu, M. Rapid Screening and Characterization of Drug Metabolites Using a Multiple Ion Monitoring-Dependent MS/MS Acquisition Method on a Hybrid Triple Quadrupole-Linear Ion Trap Mass Spectrometer. *J. Mass Spectrom.* **2008**, *43* (10), 1364–1375.

(166) Scholz, K.; Dekant, W.; Völkel, W.; Pähler, A. Rapid Detection and Identification of N-Acetyl-L-Cysteine Thioethers Using Constant Neutral Loss and Theoretical Multiple Reaction Monitoring Combined with Enhanced Product-Ion Scans on a Linear Ion Trap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (12), 1976–1984.

(167) Lu, W.; Kimball, E.; Rabinowitz, J. D. A High-Performance Liquid Chromatography-Tandem Mass Spectrometry Method for Quantitation of Nitrogen-Containing Intracellular Metabolites. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (1), 37–50.

(168) Cody, R. B.; Freiser, B. S. Collision-Induced Dissociation in a Fourier-Transform Mass Spectrometer. *Int. J. Mass Spectrom. Ion Phys.* **1982**, *41* (3), 199–204.

(169) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol.* **2005**, *402*, 148–185.

(170) Johnson, A. R.; Carlson, E. E. Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation. *Anal. Chem.* **2015**, *87* (21), 10668–10678.

(171) Dyson, P. J.; Hearley, A. K.; Johnson, B. F. G.; Khimyak, T.; McIndoe, J. S.; Langridge-Smith, P. R. R. Mass Spectrometric Method for the Rapid Characterization of Transition Metal Carbonyl Cluster Reaction Mixtures. **2001**.

(172) Dyson, P. J.; Johnson, B. F. G.; McIndoe, J. S.; Langridge-Smith, P. R. R. Energy-Dependent Electrospray Ionisation Mass Spectrometry: Applications in Transition Metal Carbonyl Chemistry. *Rapid Commun. Mass Spectrom.* **2000**, *14* (5), 311–313.

(173) Dyson, P. J.; Hearley, A. K.; Johnson, B. F. G.; McIndoe, J. S.; Langridge-Smith, P. R. R.; Whyte, C. Combining Energy-Dependent Electrospray Ionisation with Tandem Mass Spectrometry for the Analysis of Inorganic Compounds. *Rapid Commun. Mass Spectrom.* **2001**, *15* (12), 895–897.

(174) Butcher, C. P. G.; Johnson, B. F. G.; McIndoe, J. S.; Yang, X.; Wang, X.-B.;

Wang, L.-S. Collision-Induced Dissociation and Photodetachment of Singly and Doubly Charged Anionic Polynuclear Transition Metal Carbonyl Clusters: Ru3Co(CO)13−, Ru6C(CO)162−, and Ru6(CO)182−. *J. Chem. Phys.* **2002**, *116* (15), 6560–6566.

(175) Butcher, C. P. G.; Dyson, P. J.; Johnson, B. F. G.; Langridge-Smith, P. R. R.; McIndoe, J. S.; Whyte, C. On the Use of Breakdown Graphs Combined with Energy-Dependent Mass Spectrometry to Provide a Complete Picture of Fragmentation Processes. *Rapid Commun. Mass Spectrom.* **2002**, *16* (16), 1595–1598.

(176) Husheer, S. L. G.; Forest, O.; Henderson, M.; McIndoe, J. S. EDit: A Computer Program to Assist in the Presentation of Energy-Dependent Mass Spectra. *Rapid Commun. Mass Spectrom.* **2005**, *19* (10), 1352–1354.

(177) Sawyer, S. A Market-Based Perspective on Information Systems Development. *Commun. ACM* **2001**, *44* (11), 97–102.

(178) Chia, P. H.; Heiner, A. P.; Asokan, N. Use of Ratings from Personalized Communities for Trustworthy Application Installation; Springer, Berlin, Heidelberg, 2012; pp 71–88.

(179) Seithe, M.; Morina, J.; Glöckner, A. Bonn EXperimental System (BoXS): An Open-Source Platform for Interactive Experiments in Psychology and Economics. *Behav. Res. Methods* **2016**, *48* (4), 1454–1475.

(180) Tinkercad | Create 3D digital designs with online CAD https://www.tinkercad.com/ (accessed Dec 11, 2018).

(181) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open

Source Software for Rapid Proteomics Tools Development. *Bioinformatics* **2008**, *24* (21), 2534–2536.

(182) TkInter - Python Wiki https://wiki.python.org/moin/TkInter (accessed Dec 11, 2018).

(183) Bruno, B. J.; Miller, G. D.; Lim, C. S. Basics and Recent Advances in Peptide and Protein Drug Delivery. *Ther. Deliv.* **2013**, *4* (11), 1443–1467.

(184) Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. The Future of Peptide-Based Drugs. *Chem. Biol. Drug Des.* **2013**, *81* (1), 136–147.

(185) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(186) Yates, J. R.; Eng, J. K.; McCormack, A. L.; Schieltz, D. Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. *Anal. Chem.* **1995**, *67* (8), 1426–1436.

(187) Skinner, O. S.; Kelleher, N. L. Illuminating the Dark Matter of Shotgun Proteomics. *Nat. Biotechnol.* **2015**, *33* (7), 717–718.

(188) Matallana-Surget, S.; Leroy, B.; Wattiez, R. Shotgun Proteomics: Concept, Key Points and Data Mining. *Expert Rev. Proteomics* **2010**, *7* (1), 5–7.

(189) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R.; III. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343.

(190) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data-Independent Acquisition-Based SWATH-MS for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2018**, *14* (8), e8126.

(191) Krasny, L.; Bland, P.; Kogata, N.; Wai, P.; Howard, B. A.; Natrajan, R. C.; Huang, P. H. SWATH Mass Spectrometry as a Tool for Quantitative Profiling of the Matrisome. *J. Proteomics* **2018**, *189*, 11–22.

(192) Simhadri, C.; Gignac, M. C.; Anderson, C. J.; Milosevich, N.; Dheri, A.; Prashar, N.; Flemmer, R. T.; Dev, A.; Henderson, T. G.; Douglas, S. F.; et al. Structure–Activity Relationships of Cbx7 Inhibitors, Including Selectivity Studies against Other Cbx Proteins. *ACS Omega* **2016**, *1* (4), 541–551.

(193) Tang, H.; Mechref, Y.; Novotny, M. V. Automated Interpretation of MS/MS Spectra of Oligosaccharides. *Bioinformatics* **2005**, *21 Suppl 1* (Suppl 1), i431-9.

(194) Yamagaki, T.; Makino, Y. Fragmentation of Oligosaccharides from Sodium Adduct Molecules Depends on the Position of N-Acetyl Hexosamine Residue in Their Sequences in Mass Spectrometry. *Mass Spectrom. (Tokyo, Japan)* **2017**, *6* (Spec Iss 2), S0073.

(195) Harvey, D. J. Analysis of Carbohydrates and Glycoconjugates by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry: An Update Covering the Period 1999–2000. *Mass Spectrom. Rev.* **2006**, *25* (4), 595–662.

(196) Everest-Dass, A. V.; Kolarich, D.; Campbell, M. P.; Packer, N. H. Tandem Mass Spectra of Glycan Substructures Enable the Multistage Mass Spectrometric Identification of Determinants on Oligosaccharides. *Rapid Commun. Mass Spectrom.* **2013**, *27* (9), 931–939.

(197) Reinhold, V. N.; Reinhold, B. B.; Costello, C. E. Carbohydrate Molecular Weight Profiling, Sequence, Linkage, and Branching Data: ES-MS and CID. *Anal. Chem.* **1995**, *67* (11), 1772–1784.

(198) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci. Rep.* **2011**, *1* (1), 90.

(199) Crawford, E.; Dyson, P. J.; Forest, O.; Kwok, S.; Scott McIndoe, J. Energy-Dependent Electrospray Ionisation Mass Spectrometry of Carbonyl Clusters. *J. Clust. Sci.* **2006**, *17* (1), 47–63.

(200) Nemykin, V. N.; Basu, P. Energy-Dependent Electrospray Ionization Mass Spectrometric Studies of Mononuclear Metal Carbonyls. *Inorganica Chim. Acta* **2005**, *358* (10), 2876–2882.

(201) Fusco, R.; Luca Longo; Masi, F.; Garbassi, F. Olefin Polymerization with Homogeneous Ziegler−Natta Catalysts: A DFT Quantum-Mechanical Study of the Reactions of Cp2MtCH3Cl Complexes (Mt = Ti, Zr) with Al(CH3)3 and MAO. **1997**.

(202) Chen, E. Y.; Marks, T. J. Cocatalysts for Metal-Catalyzed Olefin Polymerization: Activators, Activation Processes, and Structure-Activity Relationships. *Chem. Rev.* **2000**, *100* (4), 1391–1434.

(203) Cavallo, L.; Guerra, G.; Corradini, P. Mechanisms of Propagation and Termination Reactions in Classical Heterogeneous Ziegler−Natta Catalytic Systems: A Nonlocal Density Functional Study. **1998**.

(204) Anderson, L.; Hunter, C. L. Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. *Mol. Cell. Proteomics* **2006**, *5* (4), 573–588.

(205) Stampfer, M.; Bartley, J.; Lauffenburger, D. A.; White, F. M. Induction of

Transformation and Continuous Cell Lines from Normal Human Mammary Epithelial Cells after Exposure to Benzo[a]Pyrene. *PNAS* **2007**, *82* (8), 2394–2398.

(206) Cohen Freue, G. V.; Borchers, C. H. Multiple Reaction Monitoring (MRM): Principles and Application to Coronary Artery Disease. *Circ. Cardiovasc. Genet.* **2012**, *5* (3), 378–378.

(207) AutoHotkey https://www.autohotkey.com/ (accessed Dec 11, 2018).

(208) Zijlstra, H. S.; Linnolahti, M.; Collins, S.; McIndoe, J. S. Additive and Aging Effects on Methylalumoxane Oligomers. *Organometallics* **2017**, *36* (9), 1803–1809.

(209) Theron, R.; Wu, Y.; Yunker, L. P. E.; Hesketh, A. V.; Pernik, I.; Weller, A. S.; McIndoe, J. S. Simultaneous Orthogonal Methods for the Real-Time Analysis of Catalytic Reactions. *ACS Catal.* **2016**, *6* (10), 6911–6917.

(210) West, A. M. A.; Elber, R.; Shalloway, D. Extending Molecular Dynamics Time Scales with Milestoning: Example of Complex Kinetics in a Solvated Peptide. *J. Chem. Phys.* **2007**, *126* (14), 145104.

(211) Espenson, J. H.; York, N.; San, S. L.; Auckland Bogota, F.; Lisbon, C.; Madrid, L.; City, M.; Montreal, M.; Delhi, N.; Juan, S.; et al. *CHEMICAL KINETICS AND REACTION MECHANISMS Second Edition*.

(212) Kozuch, S.; Shaik, S. A Combined Kinetic−Quantum Mechanical Model for Assessment of Catalytic Cycles: Application to Cross-Coupling and Heck Reactions. **2006**.

(213) Denmark, S. E.; Sweis, R. F.; Wehrli, D. Fluoride-Promoted Cross-Coupling Reactions of Alkenylsilanols. Elucidation of the Mechanism through

Spectroscopic and Kinetic Analysis. **2004**.

(214) Uhe, A.; Kozuch, S.; Shaik, S. Automatic Analysis of Computed Catalytic Cycles. *J. Comput. Chem.* **2011**, *32* (5), 978–985.

(215) Bruno, N. C.; Tudge, M. T.; Buchwald, S. L. Design and Preparation of New Palladium Precatalysts for C-C and C-N Cross-Coupling Reactions. *Chem. Sci.* **2013**, *4*, 916–920.

(216) Melvin, P. R.; Balcells, D.; Hazari, N.; Nova, A. Understanding Precatalyst Activation in Cross-Coupling Reactions: Alcohol Facilitated Reduction from Pd(II) to Pd(0) in Precatalysts of the Type (η^3-Allyl)Pd(L)(Cl) and (η^3-Indenyl)Pd(L)(Cl). *ACS Catal.* **2015**, *5* (9), 5596–5606.

(217) Qian, N. On the Momentum Term in Gradient Descent Learning Algorithms. *Neural Networks* **1999**, *12* (1), 145–151.

(218) Kim, D.; Fessler, J. A. Optimized First-Order Methods for Smooth Convex Minimization. *Math. Program.* **2016**, *159* (1–2), 81–107.

(219) Barzilai, J.; Borwein, J. M. Two-Point Step Size Gradient Methods. *IMA J. Numer. Anal.* **1988**, *8* (1), 141–148.

(220) Albert Cohen. *Approximation by Greedy Algorithms*; 2007; Vol. 40.

(221) Hofinger, A. Nonlinear Function Approximation: Computing Smooth Solutions with an Adaptive Greedy Algorithm. *J. Approx. Theory* **2006**, *143* (2), 159–175.

(222) Bang-Jensen, J.; Gutin, G.; Yeo, A. When the Greedy Algorithm Fails. *Discret. Optim.* **2004**, *1* (2), 121–127.

(223) Cattell, R. B. The Description of Personality: Basic Traits Resolved into Clusters. *J. Abnorm. Soc. Psychol.* **1943**, *38* (4), 476–506.

(224) Sibson, R. SLINK: An Optimally Efficient Algorithm for the Single-Link Cluster Method. *Comput. J.* **1973**, *16* (1), 30–34.

(225) Defays, D. An Efficient Algorithm for a Complete Link Method. *Comput. J.* **1977**, *20* (4), 364–366.

(226) Agrawal, R.; Gehrke, J.; Gunopulos, D.; Raghavan, P. Automatic Subspace Clustering of High Dimensional Data. *Data Min. Knowl. Discov.* **2005**, *11* (1), 5–33.

(227) Ghodsi, A. *Dimensionality Reduction A Short Tutorial*.

(228) Yan, S.; Xu, D.; Zhang, B.; Zhang, H.; Yang, Q.; Lin, S. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29* (1), 40–51.

(229) Bingham, E.; Mannila, H. *Random Projection in Dimensionality Reduction: Applications to Image and Text Data*.

(230) Cao, L. J.; Chua, K. S.; Chong, W. K.; Lee, H. P.; Gu, Q. M. A Comparison of PCA, KPCA and ICA for Dimensionality Reduction in Support Vector Machine. *Neurocomputing* **2003**, *55* (1–2), 321–336.

(231) Tipping, M. E.; Bishop, C. M. Probabilistic Principal Component Analysis. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **1999**, *61* (3), 611–622.

(232) Vasilescu, M. A. O.; Terzopoulos, D. Multilinear Image Analysis for Facial Recognition. In *Object recognition supported by user interaction for service robots*; IEEE Comput. Soc; Vol. 2, pp 511–514.

(233) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. Metabonomics: The

Use of Electrospray Mass Spectrometry Coupled to Reversed-Phase Liquid Chromatography Shows Potential for the Screening of Rat Urine in Drug Development. *Rapid Commun. Mass Spectrom.* **2002**, *16* (20), 1991–1996.

(234) Garcia, J. S.; Vaz, B. G.; Corilo, Y. E.; Ramires, C. F.; Saraiva, S. A.; Sanvido, G. B.; Schmidt, E. M.; Maia, D. R. J.; Cosso, R. G.; Zacca, J. J.; et al. Whisky Analysis by Electrospray Ionization-Fourier Transform Mass Spectrometry. *Food Res. Int.* **2013**, *51* (1), 98–106.

(235) Coomans, D.; Massart, D. L. Alternative K-Nearest Neighbour Rules in Supervised Pattern Recognition : Part 1. k-Nearest Neighbour Classification by Using Alternative Voting Rules. *Anal. Chim. Acta* **1982**, *136*, 15–27.

(236) Altman, N. S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am. Stat.* **1992**, *46* (3), 175–185.

(237) Introduction to Random Forest Algorithm with Python https://blog.goodaudience.com/introduction-to-random-forest-algorithm-with-python-9efd1d8f0157 (accessed Dec 11, 2018).

(238) The Unreasonable Effectiveness of Random Forests – Rants on Machine Learning – Medium https://medium.com/rants-on-machine-learning/the-unreasonable-effectiveness-of-random-forests-f33c3ce28883 (accessed Dec 11, 2018).

(239) Caruana, R.; Karampatziakis, N.; Yessenalina, A. An Empirical Evaluation of Supervised Learning in High Dimensions. In *Proceedings of the 25th international conference on Machine learning - ICML '08*; ACM Press: New York, New York, USA, 2008; pp 96–103.

(240) Jones, Z.; Linder, F. *Exploratory Data Analysis Using Random Forests* *.

(241) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. **2003**.

(242) Sollich, P.; Krogh, A. *Learning with Ensembles: How over-Fitting Can Be Useful*.

(243) Chandra, A.; Yao, X. DIVACE: Diverse and Accurate Ensemble Learning Algorithm; Springer, Berlin, Heidelberg, 2004; pp 619–625.

(244) Yu, L.; Wang, S.; Lai, K. K. Credit Risk Assessment with a Multistage Neural Network Ensemble Learning Approach. *Expert Syst. Appl.* **2008**, *34* (2), 1434–1444.

(245) Wang, G.; Hao, J.; Ma, J.; Jiang, H. A Comparative Assessment of Ensemble Learning for Credit Scoring. *Expert Syst. Appl.* **2011**, *38* (1), 223–230.

(246) *Ensemble Machine Learning*; Zhang, C., Ma, Y., Eds.; Springer US: Boston, MA, 2012.

(247) Mehta, P.; Wang, C.-H.; Day, A. G. R.; Richardson, C.; Bukov, M.; Fisher, C. K.; Schwab, D. J. *A High-Bias, Low-Variance Introduction to Machine Learning for Physicists*; 2018.

(248) Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese Neural Networks for One-Shot Image Recognition. *Present. Deep Learn. Work. 2015 Int. Conf. Mach. Learn. Lille, Fr.* **2015**.

(249) Santoro, A.; Bartunov, S.; Botvinick, M.; Wierstra, D.; Lillicrap, T. One-Shot Learning with Memory-Augmented Neural Networks. **2016**.

(250) Bromley, J.; Guyon, I.; Lecun, Y.; Sickinger, E.; Shah, R. *Signature Verification Using a "Siamese" Time Delay Neural Network*.

(251) Fei-Fei, L.; Fergus, R.; Perona, P. *One-Shot Learning of Object Categories*.

(252) MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges http://yann.lecun.com/exdb/mnist/ (accessed Dec 11, 2018).

(253) Banko, M.; Brill, E. Scaling to Very Very Large Corpora for Natural Language Disambiguation. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics - ACL '01*; Association for Computational Linguistics: Morristown, NJ, USA, 2001; pp 26–33.

(254) *The Unreasonable Effectiveness of Data*; 2009.

(255) Simkin, J.; Trowbridge, C. W. Optimization Problems in Electromagnetics. *IEEE Trans. Magn.* **1991**, *27* (5), 4016–4019.

(256) The Optimization Problem. In *Practical Optimization*; Springer US: Boston, MA, 2007; pp 1–26.

(257) Segal, M. R. *Machine Learning Benchmarks and Random Forest Regression Publication Date Machine Learning Benchmarks and Random Forest Regression*; 2003.

(258) Duan, Y.; Andrychowicz, M.; Stadie, B.; Ho, J.; Schneider, J.; Sutskever, I.; Abbeel, P.; Zaremba, W.; Research Lab, B. A. *One-Shot Imitation Learning Work Done While at OpenAI*.

(259) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9* (3), 90–95.

(260) Zander, S.; Nguyen, T.; Armitage, G. Automated Traffic Classification and Application Identification Using Machine Learning. In *The IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05)l*; IEEE, 2005; pp 250–257.

(261) Blum, A. L.; Langley, P. Selection of Relevant Features and Examples in Machine Learning. *Artif. Intell.* **1997**, *97* (1–2), 245–271.

(262) Soren Bouma. One Shot Learning and Siamese Networks in Keras – Neural Tinkering https://sorenbouma.github.io/blog/oneshot/ (accessed Dec 11, 2018).

(263) Wilson, D. R.; Martinez, T. R. The General Inefficiency of Batch Training for Gradient Descent Learning. *Neural Networks* **2003**, *16* (10), 1429–1451.

(264) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. **2015**.

(265) Balles, L.; Romero, J.; Hennig, P. *Coupling Adaptive Batch Sizes with Learning Rates*; 2017.

(266) Bengio, Y. Practical Recommendations for Gradient-Based Training of Deep Architectures; Springer, Berlin, Heidelberg, 2012; pp 437–478.

(267) Keskar, N. S.; Mudigere, D.; Nocedal, J.; Smelyanskiy, M.; Tang, P. T. P. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. **2016**.

(268) Regularization for Simplicity: L_2 Regularization | Machine Learning Crash Course | Google Developers https://developers.google.com/machine-learning/crash-course/regularization-for-simplicity/l2-regularization (accessed Dec 11, 2018).

(269) Caramelli, D.; Salley, D.; Henson, A.; Aragon Camarasa, G.; Sharabi, S.; Keenan, G.; Cronin, L. Networking Chemical Robots Using Twitter for #RealTimeChem. **2018**.

(270) Fitzpatrick, D. E.; Ley, S. V. Engineering Chemistry for the Future of Chemical Synthesis. *Tetrahedron* **2018**, *74* (25), 3087–3100.

(271) Poscharny, K.; Fabry, D. C.; Heddrich, S.; Sugiono, E.; Liauw, M. A.; Rueping, M. Machine Assisted Reaction Optimization: A Self-Optimizing Reactor System for Continuous-Flow Photochemical Reactions. *Tetrahedron* **2018**, *74* (25), 3171–3175.

(272) Parrott, A. J.; Bourne, R. A.; Akien, G. R.; Irvine, D. J.; Poliakoff, M. Self-Optimizing Continuous Reactions in Supercritical Carbon Dioxide. *Angew. Chemie Int. Ed.* **2011**, *50* (16), 3788–3792.

(273) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7* (4), 308–313.

(274) Armbruster, D. A.; Overcash, D. R.; Reyes, J. Clinical Chemistry Laboratory Automation in the 21st Century - Amat Victoria Curam (Victory Loves Careful Preparation). *Clin. Biochem. Rev.* **2014**, *35* (3), 143–153.

(275) Mellon, M. G. Automation in Analytical Chemistry. *Anal. Chem.* **1958**, *30* (12), 25A–34A.

(276) Warrington, B.; Vinter, J.; Mackay, M. Automated Iterative Drug Discovery and Synthesis, 2007.

(277) Martha, C. T.; Heemskerk, A.; Hoogendoorn, J.-C.; Elders, N.; Niessen, W. M. A.; Orru, R. V. A.; Irth, H. High-Throughput Reaction Optimisation and Activity Screening of Ferrocene-Based Lewis Acid-Catalyst Complexes by Using Continuous-Flow Reaction Detection Mass Spectrometry. *Chem. - A Eur. J.* **2009**, *15* (30), 7368–7375.

(278) Wleklinski, M.; Loren, B. P.; Ferreira, C. R.; Jaman, Z.; Avramova, L.; Sobreira, T. J. P.; Thompson, D. H.; Cooks, R. G. High Throughput Reaction Screening

Using Desorption Electrospray Ionization Mass Spectrometry. *Chem. Sci.* **2018**, *9* (6), 1647–1653.

(279) Shevlin, M. Practical High-Throughput Experimentation for Chemists. **2017**, *8*, 601.

(280) Grasser, J. A.; Muggli, D. S. A High-Throughput Reaction System to Measure the Gas-Phase Photocatalytic Oxidation Activity of TiO2 Nanotubes. *Rev. Sci. Instrum.* **2009**, *80* (7), 075106.

(281) Jones, E.; Michael, S.; Sittampalam, G. S. *Basics of Assay Equipment and Instrumentation for High Throughput Screening*; Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004.

(282) Axis Travel, Home & Switch Location | Smithy - Detroit Machine Tools https://smithy.com/cnc-programmers-guide/chapter-7/page/7 (accessed Dec 11, 2018).

(283) TinyG | Synthetos.com https://www.synthetos.com/project/tinyg/ (accessed Dec 11, 2018).

(284) The 50 Failure Modes of Electric Motors - UE Systems - UE Systems http://www.uesystems.com/news/the-50-failure-modes-of-electric-motors (accessed Dec 11, 2018).

(285) Ylonen, T.; Lovnick, C. *RFC 4251 - The Secure Shell (SSH) Protocol Architecture*; 2006.

(286) Yeung, D. TinyG-PythonControl https://github.com/Yeungdb/TinyG-PythonControl.