5-2016

Peptide Fragmentation Studies on Doubly Charged Proline and Pipecolic Acid Containing Pentapeptides and Methods Development for an HPLC/MS Instrument for Proteomics Experiments

Hannah Emily Smith
College of William and Mary

Peptide Fragmentation Studies on Doubly Charged Proline and Pipecolic Acid Containing Pentapeptides and Methods Development for an HPLC/MS Instrument for Proteomics Experiments

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Science in Chemistry from
The College of William and Mary

by

Hannah Emily Smith

Accepted for Honors

Professor John C. Poutsma, Director

Professor Kristin Wustholz

Professor Lisa M. Landino

Professor Mainak Patel

Williamsburg, VA
May 2, 2016

Table of Contents

Dedication

The work in this thesis is dedicated to my grandfather, Ronald H. Smith, who is my inspiration and gives me the motivation for all that I do.

Acknowledgements

I would first like to thank Professor John C. Poutsma for the opportunity to be a member of his lab and for his continuing support and mentorship for the last year and a half. Without his unending patience, this thesis would not have happened. I would also like to thank members of the Ionlab for their support and help throughout this process and keeping time spent in lab always enjoyable. I would specifically like to thank Katie Henke for her help teaching me the basics of peptide synthesis as well as Zach Hasan and Anton Lachowicz for commiserating with me during the process of writing this thesis and preparing for the defense. I would like to thank my family, especially my parents, for always supporting me no matter what I do. I would also like to give special thanks to my grandmother, JoNelle Smith, for supporting this research project so heavily and for always being interested in hearing about my research. Finally I would like to thank my friends for always being there for me, through thick and thin. Funding for this work was provided by the Charles Center Honors Fellowship program.

List of Figures

List of Tables

Abstract

This thesis presents fragmentation studies as well as proteomics studies performed using mass spectrometry. Peptide fragmentation was performed on both singly and doubly charged peptides containing either a proline or pipecolic acid residue, as well as either an arginine or lysine residue using collision-induced dissociation. Results showed that altering the position of pipecolic acid or proline in relation to arginine or lysine can affect the formation of doubly charged product ions, which may be a result in change of the proton affinity of the basic residue. Proteomics experiments were performed on a new HPLC/LTQ instrument using a standardized sample of six digested proteins in order to optimize a method by which future proteomics studies can be performed. The protein sample underwent separation via the HPLC with a 90 or 63.3-minute gradient, then data dependent scanning in the LTQ mass spectrometer. These data dependent scans were processed using the SEQUEST searching algorithm in order to identify all possible proteins. Results showed that a 63.5-minute linear gradient of acetonitrile/H_2O was sufficient to successfully identify all proteins in the standard sample using SEQUEST, at a flow rate of 0.2 mL/min in the HPLC with initial injection size of 10 µl into a ACE superC18 reverse phase column, using CID only in the mass spectrometer to induce fragmentation.

Chapter 1: Introduction

1.1 Proteins: Structure and Function

Proteins are macromolecules found in biological systems that have a variety of crucial roles in the existence and maintenance of life on earth. Proteins can act as catalysts, transporters for other molecules, and act in signaling mechanisms, among many other roles [1]. Proteins are composed of long chains of amino acids linked together with peptide bonds, and are usually several hundred amino acids in length. Peptide bonding is the result of a dehydration reaction between two peptides. A group of several amino acids bonded together is known as a polypeptide [2].

Amino acids consist of an amine and carboxylic acid functional groups and vary based on a single side chain, often referred to as the "R group." The basic structure for an amino acid is shown below.

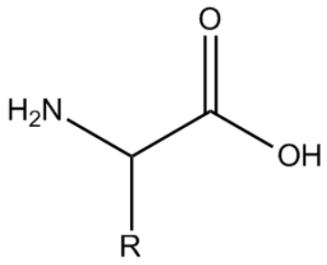

Figure 1.1 Structure of an Amino Acid

There are 20 amino acids that comprise proteins and are coded for by DNA and RNA, also known as proteinogenic amino acids (PAAs), as well as hundreds of other nonproteinogenic amino acids (NPAAs), which are not coded for by DNA and RNA and typically do not make up the structure of proteins. Each amino acid has a side chain that can be basic, acidic, nonpolar, polar, charged, or uncharged and these characteristics are crucial for function of the overall protein. Many NPAAs of interest are very similar in structure to PAAs, such as pipecolic acid

and proline, which differ based on the number of carbons in the ring side chain – where proline has a five membered ring and pipecolic acid has a six membered ring (see Figure 1.2)[3].

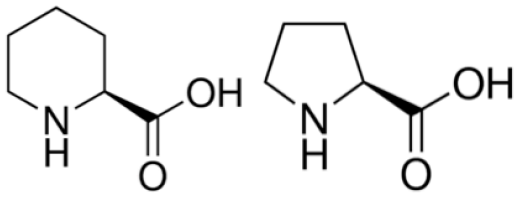

Figure 1.2 L-pipecolic acid (left) and L-proline (right)

Proteins have four different levels of organization, which give rise to overall structure and ultimately functionality. The primary structure refers to the base order of the peptide chain that composes the protein, where the main structural component is the peptide bonds that compose the backbone of the polypeptide. Next, secondary structure refers to the formation of alpha helices or beta pleated sheets that arise from hydrogen bonding interactions between the side chains on each amino acid. Tertiary structure gives rise to the overall three dimensional structure of the protein and is a result of interactions between secondary structures by way of hydrogen bonding, disulfide bonds, salt bridges, and van der Waals interactions. Finally, quaternary structure is a result of interactions between different protein subunits. For example, human alcohol dehydrogenase is a protein enzyme that functions in the body to facilitate the conversion of alcohol functional groups to ketones or aldehydes. Alcohol dehydrogenase is a dimer consisting of either α, β, or γ subunits which individually consist of mainly alpha helices as shown in figure 1.3 [4].

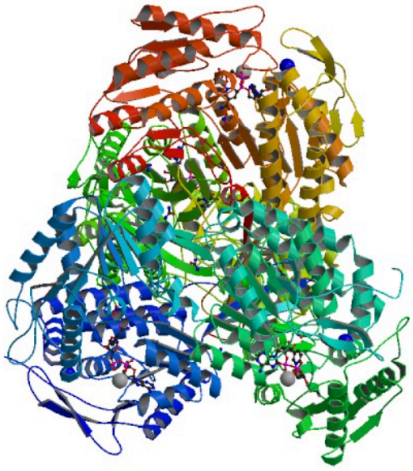

Figure 1.3 Structure of human alcohol dehydrogenase. [4]

Amino acid mutations are known to account for approximately half the known gene lesions that result in human disease [5]. The structure of the protein is directly related to function of the protein, and as previously noted, the overall structure of the protein is a result of the ordering of amino acids in the peptide chain [2,6]. Thus, change in this primary structure has the potential to have disastrous effects on the functionality of a protein. The most common example of this is sickle cell anemia, which is caused by a single point mutation that changes glutamic acid to valine. Glutamic acid at physiological pH carries a negative charge and is hydrophilic, whereas valine is nonpolar and hydrophobic. The overall structure of a hemoglobin subunit with this mutation is drastically changed to a "sickled" shape instead of a globular, rounded shape, because this change in side chain properties drastically disrupts the chemical interactions in the native protein [7].

1.2 Proteomics

Proteomics is the study of the structure and function of proteins. Since the sequencing technique of the genome of different species has been developed, the study of proteomics has advanced greatly and includes a variety of methodologies [8,9]. The genome directly codes for

all proteins of a species, however modification of DNA or RNA in a cell affects the expression of proteins on a cellular level, as does post-translational modifications. Therefore, though genomic studies can provide the fundamental amino acid primary structure of proteins, proteomics is necessary to determine more specifically what roles proteins play in different biological settings. For example, a cell infected by a virus may begin to express different proteins than originally coded for by DNA, which can cause disease. Proteomics allows researchers to probe what proteins might be found in that cell infected by a virus.

As each of the 20 PAAs have distinct masses (with the exception of isoleucine and leucine which are isomers), mass spectrometry can be a powerful tool for the field of proteomics as it is is able to distinguish between ions with different mass to charge ratios.

1.2.1 Bottom-Up Methodology

The bottom-up method of mass-spectrometry-based proteomics uses high-performance liquid chromatography (HPLC) in combination with automated tandem mass spectrometry to identify and quantify proteins [8-10]. The general procedure is as shown in figure 1.4 below.

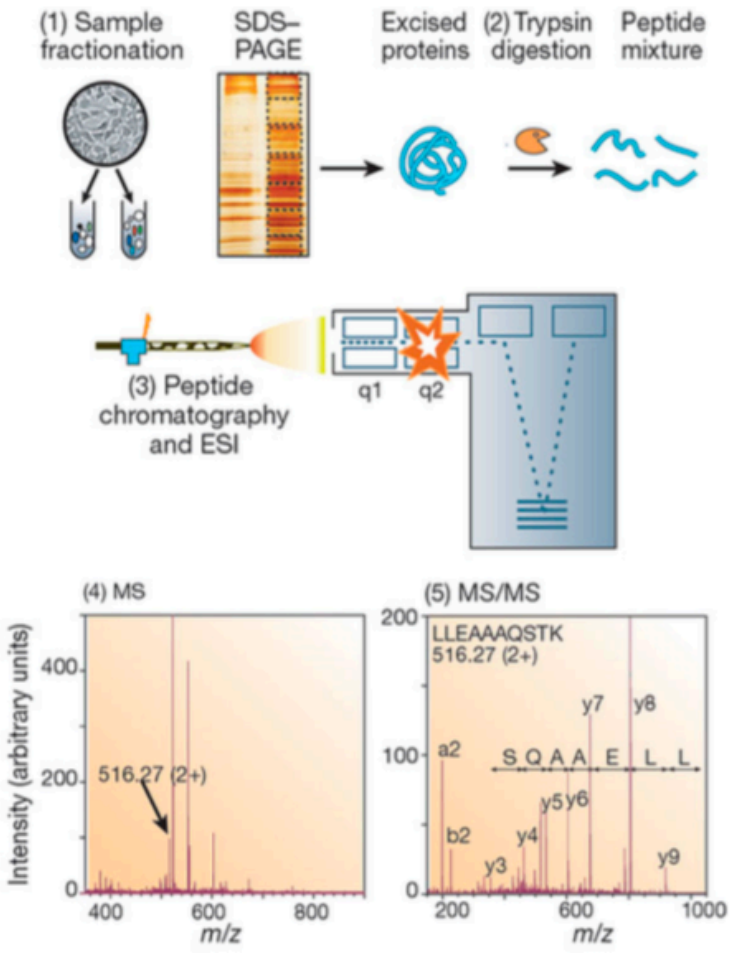

Figure 1.4 Procedure for bottom-up proteomics studies [8].

The process begins by first identifying a protein of interest or a sample. For example, this could mean comparing a normal skin cell versus a cancerous skin cell in order to determine which proteins have been differentially expressed. The sample must first be harvested from the organism of interest. Then, this sample is mechanically disrupted and purified in order to separate the protein from other components of the cell. A cell must first be lysed, because other steps in the process will not work with a cell membrane or a cell wall intact. Usually, the cell is physically disrupted by a mechanical process (such as bead beating), then a chemical detergent is used to lyse the cell and extract the proteins. Some agents commonly used for this purpose

include Triton X-100, NP-40, and sodium dodecycl sulfate. The detergent of choice will depend on the sample in question [8].

Following lysis of the cell, the chemical used in the lysis step must be removed from the system. Detergents such as sodium dodecyl sulfate can hinder the digestion process and will ruin columns used in HPLC. In order to purify the protein mixture left from the cell, a precipitation technique is used, often with acetone [8,9].

After the removal of any chemical agents from the lysis process, the protein mixture must then be digested, in order to produce fragments of a size that can be easily analyzed with a mass spectrometer. There are two options for digestion commonly used: in-solution and in-gel digestion. In in-gel digestion, the protein mixture is first separated using a gel-based technique prior to digestion. In in-solution digestion, the protein mixture is not first separated, and the entire mixture is digested directly. This technique is also known as shotgun proteomics. One major benefit of shotgun proteomics is that the sample loss by the process of gel separation is greatly minimized. Additionally, because the protein mixture is being digested directly in solution, it is easier to control and account for the overall concentration of the sample, which is important when running effective mass spectrometry experiments [8].

In in-gel digestion, protein separation is accomplished prior to digestion using gel-separation. There are a multitude of options for gel-separation, and the choice will depend on the sample and experiment in question. Researchers can use one-dimensional gel separation or two-dimensional separations. Both mass and isoelectric point of the proteins can be used to separate the protein mixtures. Both one-dimensional and two-dimensional separations have their own advantages and disadvantages. One-dimensional separation is more simple, and results in less overall loss of the sample, by virtue of the proteins only traveling in one dimension. However,

the proteins are separated less effectively. In contrast, two-dimensional separation is very efficient in terms of separation of the proteins, but has greater propensity for loss of sample [11]. With both versions of gel-separation, the concentration of the final protein digest is not easily controlled because the concentration of the final gel separated bands is not known [8].

After gel separation, the gels are stained in order to make the protein bands visible. The most common stains used in conjunction with mass spectrometry are known as Coomassie Brilliant Blue and silver. Application of the stain leads to differential intensities of the band color, which can be used as a rough estimate of protein concentration in each band. In essence, the darker the band, the more protein. By analyzing band color intensity from the two different cells, one can identify which proteins are up- or down-regulated based on which bands are more intense in color. Protein bands of interest can then be excised from the gel, and digested using a proteolytic enzyme of choice [8, 11].

The most commonly used proteolytic enzyme in mass spectrometry based proteomics is trypsin [8]. Trypsin is a naturally occurring serine protease that is found in humans in the small intestine, and plays a role in digestion and absorptions of proteins found in food. Trypsin cleaves peptides preferentially at the carboxyl side of arginine and lysine, except when either are followed by a proline residue. Trypsin functions by way of a charge relay in the “catalytic triad” of the enzyme, which consists of a serine, aspartate, and histidine residue. This charge relay increases the nucleophilicity of serine and induces the catalytic mechanism [12].

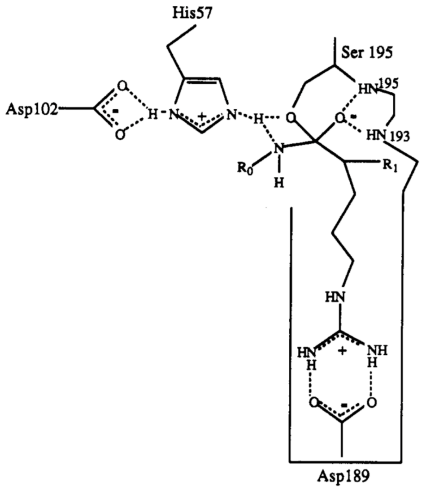

Figure 1.5 Transition state of the catalytic triad of trypsin [12]

After the tryptic digest is completed, the resulting polypeptides are extracted from the gel and put in to a solution of typically methanol/water or acetonitrile/water and acidified at a low concentration as preparation for insertion into the HPLC/MS system, which will produce protonated $[M+H]^+$ ions of the digested peptides from which can they be identified based on their mass to charge ratio [13, 14].

The method by which the proteomics experiment utilizes HPLC varies depending on whether or not an in solution digestion or in gel digestion was performed. If an in gel digestion was used, then typically the sample is inserted directly into the HPLC where it is allowed to go through the column. The most commonly used column in HPLC proteomics is the C18 reverse phase column. In a reverse phase column, the polypeptide mixture is separated based on hydrophobicity of the overall polypeptide, using a gradient of low to high acidified organic solvents. Commonly used solvents are acetonitrile/water or methanol/water with formic acid at a low percentage of total volume – typically 0.1% [8]. A schematic of separation of an analyte by reverse phase chromatography is shown below.

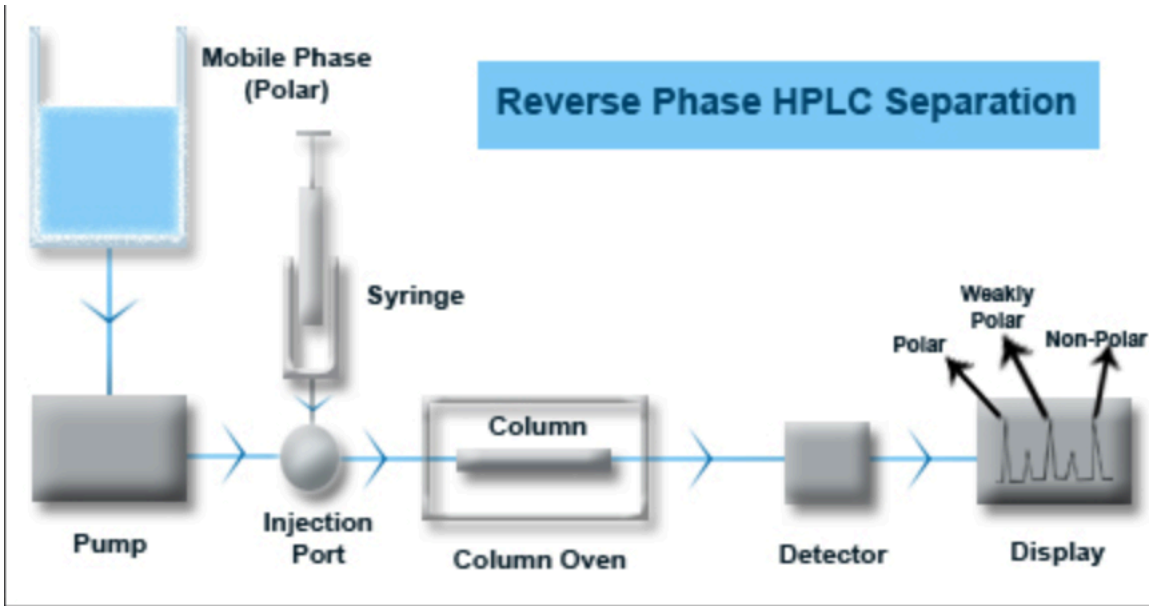

Figure 1.6 Separation of an analyte using reverse phase HPLC. (Obtained from http://lab-training.com/landing/free-hplc-training-programme-6/)

These resulting separated peptides are sent directly into the mass spectrometer, where they are isolated and allowed to undergo activation to form fragment ions. This activation is often either collision-induced dissociation or electron transfer/capture dissociation. The induced fragmentation occurs at the backbone of the peptide. Since each amino acid has a side chain with a unique molecular mass (except for leucine and isoleucine, which are isomers) the mass of the fragment ions can then be used to reconstruct the primary sequence of the peptides, which can be matched to known protein sequences obtained from genomic databases. This procedure is time-consuming, so automated computer software is used to identify peptides, and ultimately, the proteins from which they originated [8].

Bottom-up proteomics has several important advantages. First, bottom-up experiments can easily be accomplished on a much wider variety of instruments. Peptides can be fragmented using a variety of common techniques that are already found in conjunction with mass spectrometers in many laboratories today. However, the process can be very time consuming as it involves an extensive number of steps – cell lysing, gel separation, protein digestion. In

addition, there are multiple points that introduce both potential loss of sample and introduction of contaminant, which can hinder the successful identification of as many proteins as possible. Overall, bottom-up proteomics is known to be a robust methodology for successfully identifying a large number of proteins with excellent repeatability [8].

1.2.2 Top-Down Methodology

In top-down proteomics studies, rather than initially subjecting whole proteins to a digestion procedure prior to analysis, a whole protein is ionized directly into the mass spectrometer where it undergoes some form of fragmentation by way of electron capture dissociation or electron transfer dissociation [8,15]. A general diagram of this process is shown below in figure 1.7.

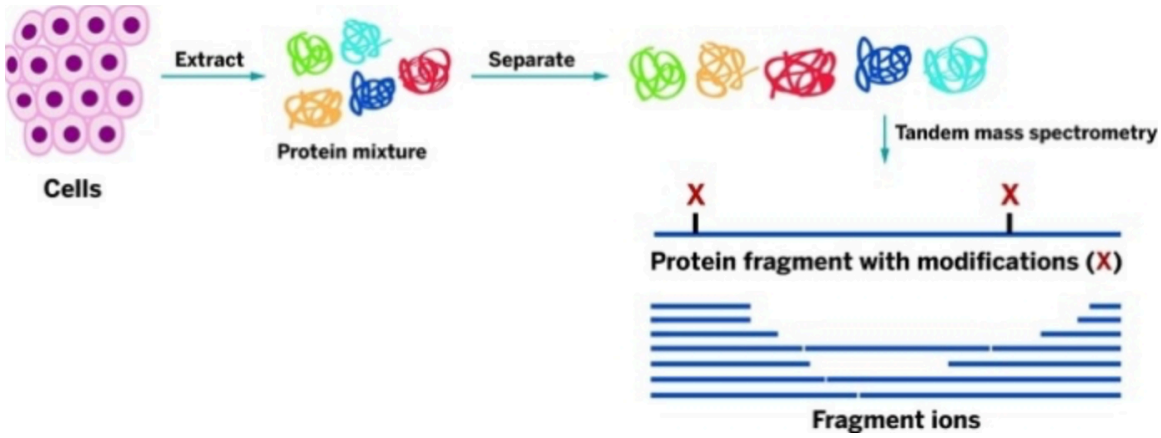

Figure 1.7 Procedure for top-down proteomics studies. (Obtained from http://clp.northwestern.edu/news/top-down-proteomics-becomes-reality)

This technique is desirable, as it allows researchers to study an entire protein by itself, as opposed to a protein that has been modified by digestion, which means that more information regarding the original structure and any post-translational modification is conserved. Additionally, top down proteomics allows for identifications of smaller proteins. Small proteins may be digested into pieces that are too small and would not be recognized in the bottom up

method. Again, prior to introduction into the mass spectrometer, the cells of interest must first be lysed in order to extract the proteins, however without the digestion step the process becomes significantly less time consuming [8].

However, there are also several disadvantages to top down proteomics. Top down proteomics experiments typically have very low signal intensity for a fragmentation event, thus have lower sensitivity than bottom up proteomics. Also, the process of directly ionizing a large protein and fragmenting it can be difficult, and many mass spectrometry instruments currently in use on a large scale are not sophisticated enough to complete this task. There is a significant cost associated with mass spectrometers that have the capability of performing top down experiments [16].

1.2.3 Sequence Database Searching

One technique to identify proteins based on an HPLC-MS/MS study involves using a computer to automatically search a database of computer generated fragmentation spectra. A program identifies the parent mass of each fragmentation spectrum obtained, generates a series of peptides that would match that parent mass, then creates theoretical fragmentation spectra using characterized rules of peptide fragmentation based on the mechanisms believed to be involved. These theoretical fragmentation spectra are then compared with the experimental spectra in question and the program attempts to create a match [17]. Researchers have developed a variety of different libraries of proteins that correspond to different organisms that have had their genome sequenced. These libraries contain sequence information for all proteins in an organism that are known. Therefore, for each study, the program creates a collection of found peptide fragments based on their spectra, and then compares them to known peptide sequences in proteins in order to identify a protein [17-20].

A commonly used program for this purpose is SEQUEST, which uses established rules to create theoretical fragmentation spectra for a variety of candidate peptides that have the same molecular mass as the peptide that was isolated and fragmented in the mass spectrometer. The spectrum from the peptide being analyzed is compared to each theoretical spectrum and SEQUEST provides a list of matching peptides with different confidence values (including cross correlation, Xcorr) based on how similar SEQUEST believes the match between the two spectra to be. Once the peptide has been identified, SEQUEST can then search a database of protein primary sequences to find which protein originally contained that peptide [19,20].

Unfortunately, protein identification through SEQUEST is not a perfect system. The rules of peptide fragmentation in the mass spectrometer are not completely understood. Currently, SEQUEST assumes that fragmentation of the peptide occurs in a random fashion, which is not always true [3]. The presence of some amino acids in a peptide chain can lead to "selective fragmentation," where the spectrum displays an increased abundance of a certain fragment, which can lead to skewed results as interpreted by SEQUEST. For example, the residues Pro, Pip, Asp, Glu, and Orn have all been shown to cause selective cleavages than can confound SEQUEST [3, 21-23]. If SEQUEST is assuming uniform fragmentation when in reality, polypeptides can be preferentially fragmenting, current studies in proteomics that use the SEQUEST database could be missing out on important proteins that could lead to improvements in health and disease treatment.

Importantly, these databases can only be used in order to identify known proteins from a genome that has already been sequenced. Furthermore, the time and computational resources required can be extensive and varies based of the complexity of the database of known protein sequences [19]. Samples from bacterial cells that only produce a small number of proteins can be

analyzed more quickly, because the library being searched will be relatively small. However, identifying a protein from a more complicated organism such as a human, can require much more time and more substantial computational resources. To alleviate this disadvantage, researchers have developed specialized, smaller databases to search against. Many labs have developed their own databases specific to their own needs. For example, Weinman et al created a specialized database for data obtained from a MALDI-TOF mass spectrometer for samples from a *Sinorhizobium meliloti* bacterium [20].

One significant challenge with utilizing SEQUEST and other similar protein searching algorithms is evaluating incorrect versus correct identifications. Multiple different probabilistic approaches have been developed in order to try and introduce a method to assess results from searching algorithms such as SEQUEST without the need for human interpretation, though this research is still in progress [24, 25].

1.3 Fragmentation

1.3.1 Fragmentation Mechanism

The mechanisms regarding how peptides fragment are not yet fully understood. However, there exists some general assumptions and nomenclature that define spectra that the mass spectrometer produces. For example, researchers know that peptides will fragment into actual pieces of the original peptide chain as opposed to side chain losses under low energy conditions, and will additionally commonly lose H_2O, NH_3, and CO molecules. For low energy collision-induced dissociation (CID) events of the peptide bond itself, researchers refer to fragments that contain the C-terminus of the original peptide chain as y_n^+ ions and fragments that contain the N-terminus are referred to as b_n^+ ions, where the n refers to the number of amino acid residues from

the end of peptide (the N-terminal end or C-terminal end, depending on the fragment). The full naming classification of peptide fragmentation is shown in the figure below.

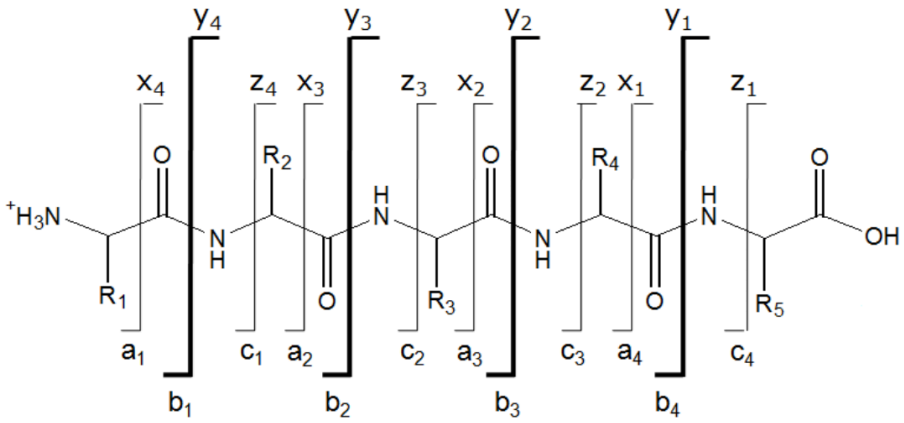

Figure 1.8 Nomenclature for peptide fragment ions [26]

The mechanism by which peptides fragment are now thought to vary based on whether or not the protonated peptide contains a mobile proton, or no mobile proton. Studies on selective cleavages in the mass spectrometer focus on the ratio of b-type ions to y-type ions, and try to determine whether the ratios can be predicted more effectively based on the composition of the peptide [3].

The mobile proton theory was developed through the work of Wysocki and Gaskell, with supporting data from various others and is now commonly accepted as the ideal model for fragmentation of a peptide in low energy collision events, such as CID [27-29]. In this model, the cleavage of a protonated peptide requires the presence of a "mobile proton" at the site of fragmentation, and is thought to be "charge-directed." This mobile proton upon ionization of the peptide is initially sequestered to the more basic sites on the peptide, such as at arginine, lysine, or histidine residues. With application of energy, this proton can be mobilized and "trigger charge site initiated mechanisms" that result in fragmentation [28]. The energy required to

initiate the mobility of a proton to induce fragmentation depends on the residues of the peptides in question, and is highest with arginine and decreases with gas-phase basicity of the residue [27].

Studies have also shown that fragmentation producing b_n^+ ions generally results in an oxazolone structure, which is related to interactions of the charge on the peptide with the N-terminal carbonyl [13, 30]. This is the result of the mechanism that follows activation of the mobile proton, which induces dissociation. This mechanism is known as the b_x-y_z pathway and involves nucleophilic attack to a carbon from a protonated peptide bond. Following this nucleophilic attack, an oxazolone b-type ion and y-type neutral fragment are produced and dissociate. Alternatively, proton transfer can occur from the protonated b_x^+ ion to the neutral y_z fragment creating a y_z^+ ion and a b_x neutral. The fragment that retains the mobile proton and creates an ion detectable by the mass spectrometer depends on factors such as the relative proton affinities of residues in the peptide [30]. A general structure for both b_n^+ and y_n^+ ions is shown below.

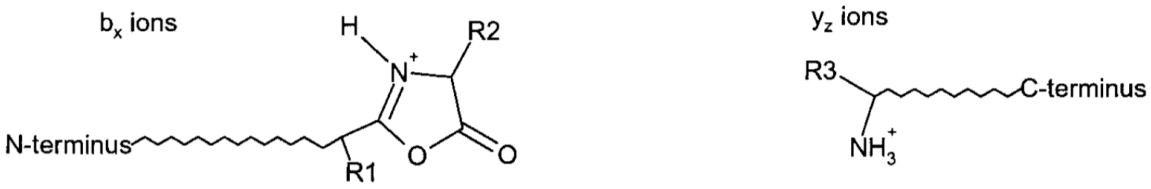

Figure 1.10 Structure of b-type and y-type ions [30]

Studies on peptide fragmentation often also involve comparison to theoretical energy calculations, in an attempt to discern more information on potential mechanisms of the fragmentation. Researchers can use quantum-modeling software to examine the energetics of different conformations of the peptides. Using these methods, researchers can predict what conformers are most likely to occur based on their energetics, predict a possible mechanism for

fragmentation, and then compare these mechanisms to actual experimental results. When the quantum theoretical models do not agree with experimental results, it can be assumed that there are other factors at play that have not yet been considered. For example, computational modeling has suggested that the pipecolic acid fragmentation should have similar energetics for both the C-terminal cleavage and the N-terminal cleavage, which suggests that both versions of fragmentation should be seen in equal quantities in the mass spectra, however the experimental results show that the C-terminal cleavage appears more frequently [3].

In the mobile proton model described above, there exists only one "mobile proton" which initiates fragmentation. In the doubly charged system, there are two protons that must be considered. The doubly charged system can result when a peptide contains an internal or C-terminal basic amino acid residue. Of the 20 amino acids coded by the human genome, arginine and lysine are both considered basic amino acids. This doubly protonated system, in a more practical application, will occur when a protein is digested using a trypsin proteolytic enzyme – as mentioned above, trypsin will digest peptides at both arginine and lysine residues. Some studies suggest that protonation that occurs during ionization in the mass spectrometer will occur on the side chains of arginine and lysine when a peptide contains these residues. In contrast, when a peptide does not contain these residues, they suggest protonation occurs on the peptide backbone or N-terminus instead [21].

1.3.2 Preferential Fragmentation

The proline effect refers to increased fragmentation of a peptide chain N-terminal to the residue proline [3, 22, 32]. This means, by the above nomenclature for fragmented ions, that a spectrum of a peptide containing a proline residue will form an abundance of y_n^+ ions as opposed

to b_n^+ ions. A variety of studies have been completed with the intent of determining a method by which this mechanism occurs. Generally, these studies have used peptides with varying numbers of alanine residues, as the alanine amino acid contains a $–CH_3$ R-group which is relatively unreactive and with relatively low proton affinity, as compared to many other amino acids, along with a proline residue placed somewhere within the peptide chain. This position can be varied along the alanine residues, to determine if the position plays a role in any potential fragmentation mechanisms [22]. Researchers have also performed analyses on databases consisting of existing known spectra including amino acid residues of particular interest. For example, one research group studied a tandem mass spectral database of doubly protonated tryptic peptides, in order to examine the effect of internal basic residues on the fragmentation patterns [21].

A corollary to this proline effect is the recently discovered "pipecolic acid effect," wherein the peptide is cleaved C-terminal to pipecolic acid, selectively producing mainly b_n^+ instead of y_n^+ ions. Pipecolic acid is an amino acid quite similar to proline in structure, differing only by one carbon in the side chain ring (as shown above in Figure 1.2), so the fact that it fragments in what can be considered an opposite fashion suggests that the mechanism by which proline fragments is specific to only proline; there must be something unique about proline's structure that allows this cleavage to occur. While proline contains a 5-membered ring as an R group, pipecolic acid contains a 6-membered ring. Again, the fact that these two very similar peptides selectively fragment in different ways means that the ring structure or the number of atoms in the ring must itself be relevant in the mechanism. There also exist two other analogs to proline: Azetidine-2-carboxylic acid (Aze) and N-methylalanine (NMeA), which are a four membered ring version and an acyclic version, respectively. Another study determined that Aze will fragment similarly to proline and NMeA will fragment similarly to pipecolic acid [3]. It

should be noted that only proline is one of the 20 standard amino acids that are coded for by the human genetic code. As a result, other amino acids being studied that are not coded for by the human genome are also important, because they can be misincorporated into proteins in the human body, sometimes resulting in disease [33]. Thus, identification of these amino acids in proteomics research is important.

It has also been shown that the D- and L-enantiomers of proline result in different fragmentation spectra for peptides that contain them. Specifically, L-proline incorporated in the pentapeptide AAXAA (where X is representing either enantiomer of proline, A is representing alanine) will produce mostly y_n^+ fragments, where D-proline will not abide by the proline effect, producing more b_n^+ fragments [32]. As the enantiomers have the same chemical formula, this means that their 3-dimensional structures must also affect fragmentation behavior.

It has been hypothesized that the differential fragmentation between proline, azetidine, N-methylalanine, and pipecolic acid occurs as a result of the mobile proton effect. Researchers have suggested that the structures of Aze and Pro are "rigid," meaning that the mobile proton is isolated – so the y_n^+ fragment is the only energetically favorable mechanism. In contrast, pipecolic acid and N-methylalanine are more flexible, and the mobile proton is able to travel to different, but still energetically-preferred sites [3].

Researchers have also studied fragmentation patterns when proline is combined in a peptide chain along with different amino acid residues, other than the "control" alanine. In one instance, researchers studied the effects of serine, leucine, valine, phenyalanine, and tryptophan on the proline effect in pentapeptides, with one proline residue. The proton affinity of each of the amino acid residues is also involved in the fragmentation pattern by way of the mobile proton mechanism, and each of these amino acids has a different proton affinity. These researchers

found that as the proton affinity increases, the ratio of y_2^+ to b_3^+ fragments will decrease. That is, as the proton affinity decreases, the proline effect is less likely to take hold on the fragmentation spectra [22].

Most existing studies on the proline and pipecolic acid effects have examined only singly charged peptides. However, in a doubly charged system, the fragmentation pattern is quite altered. As stated above, doubly charged peptides can occur with the presence of a basic amino acid, such as lysine or arginine [21].

Chapter 2: Experimental Procedures

2.1 Peptide Fragmentation

2.1.1 Synthesis

All peptides in this experiment were synthesized in lab using a standard solid state technique that was first developed by Merrifield in the 1960s and has been subsequently optimized throughout the years to increase speed of synthesis, purity, and yield. This method is known to produce high purity, crude peptides with extended synthesis times [34-36]. In this technique, peptide residues are added stepwise to an initial residue bound by the C terminus to a Wang resin, which are commercially available.

Fmoc-Pro-OH and Fmoc-Ala-OH residues and Fmoc-Arg(Pbf) and Fmoc-Lys(Boc) Wang resins were purchased from Chem Pep. The Fmoc-L-pipecolic acid residue was purchased from Neosystem. 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU) was purchased from ChemPep. DCM and DMF were purchased from Fischer Scientific. N,N-Diisopropylethylamine (DIEA) and piperidine were purchased from Sigma Aldrich.

The amino acid attached to the Wang resin is initially protected on the N-terminus by a Fmoc protecting group. Other amino acid residues are also commercially available without Wang resin or any protecting group on the C-terminus, however an Fmoc protecting group also protects them on the N-terminus. With stepwise removal of the Fmoc protecting group and addition of the new amino acid residue, peptides can be synthesized starting from the C terminus and moving to the desired N terminus. After the desired number of amino acids has been linked together, the newly synthesized polypeptide can be cleaved from the Wang resin and precipitated out of solution to be collected in solid form. In addition to the Fmoc protecting group found on the N-terminus of all amino acids, the side chain of lysine is bound to a tert-butyloxycarbonyl

(Boc) protecting group as well as arginine side chain is bound to a 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl (Pbf) protecting group. An example of one step of the synthesis process for a Pro-Lys dipeptide is as follows.

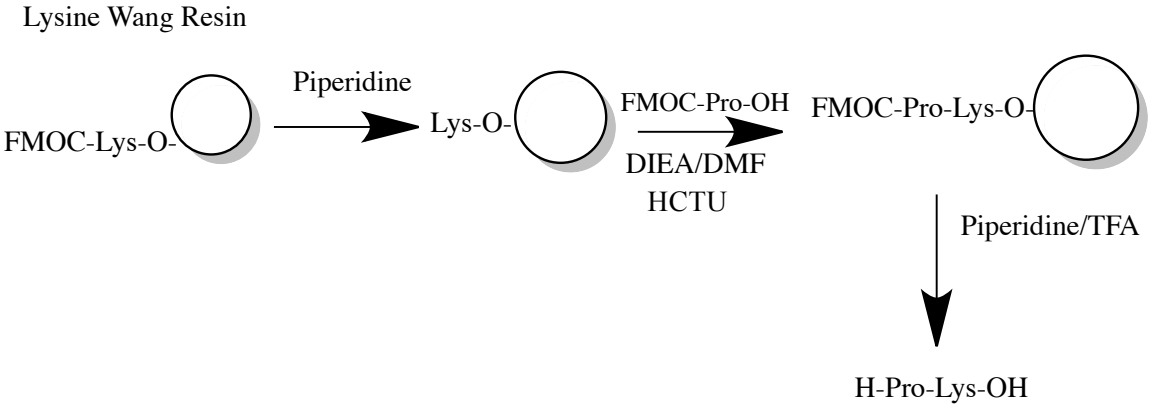

Figure 2.1 Example synthesis of Pro-Lys dipeptide

As shown in the figure above, the first main step in the coupling process involves addition of piperidine in order to remove the Fmoc protecting group from the N-terminus of the residue attached to the Wang resin. Next DIEA, DMF, and HCTU are used to couple the next residue to the now free N-terminus of the lysine residue. Finally following another deprotection step, trifluoroacetic acid (TFA) is used to cleave the C-terminus of the lysine residue from the Wang resin.

The synthesis begins by measuring out 0.2 g of Wang resin of choice and placing it in a Henke Sass Wolf 10 mL syringe with built in filter. This filter allows a series of washes to be performed and permits DCM/DMF/piperidine to flow through the filter while preventing the Wang resins from washing out. Roughly 5 mL of a solution mixture of 50/50 DMF/DCM is added to the syringe. The syringe is then placed on a Vortex Genie 2 Digital and is shaken at speed of about 1000 RPM for 30 minutes, to swell the resin. Next, the plunger is depressed to remove all solution and the resin is shaken with 5 mL of DMF for 1 minute two times. Then, the

Fmoc protecting group at the N-terminus of the amino acid is removed using a 20/80 mixture of piperidine/DMF. About 5 ml of the piperidine/DMF solution is added, the syringe is placed on the vortexer for 5 minutes, then the solution is removed, 5 mL more piperidine/DMF is added and the syringe is shaken on the vortexer for 30 minutes. Following this, the resin undergoes a series of 1-minute washes - twice with DMF then four times with DCM, in order to ensure all piperidine has been removed. The resin is washed twice more with DMF. A solution is made of the desired amino acid residue to be added to the peptide with DIEA and HCTU. The mass of the amino acid residue and HCTU as well as the volume of DIEA depends on the peptide being synthesized. This solution is combined with roughly 5 mL of DMF and is added to the syringe and placed on the vortexer for 1 hour. Following the coupling of the new amino acid, the peptide is washed with a series of 1-minute additions of DMF and DCM. After these final washes, the addition of the amino acid residue is complete. The peptide attached to the Wang resin can then undergo this same process multiple times to reach the desired peptide length, can be stored in the refrigerator temporarily, or can undergo the cleavage process to remove it from the Wang Resin. If stored in the refrigerator, prior to beginning any other additions or cleavage steps, the Wang resin must again be swollen using the 50/50 DMF/DCM solution.

The cleavage process starts with two 1-minute rinses using DMF, followed by the same deprotecting process using 20/80 piperidine/DMF as described above. Following this, the resin is washed using DMF/DCM, again as described above. Following the wash steps, all liquid is removed from the syringe as much as is possible with depression of the plunger. Then, 10 mL cleavage solution is added to the syringe, which is then placed on the vortexer for 2 hours. The cleavage solution consists of 95% TFA, 2.5% triisopropyl silane, and 2.5% deionized water. After two hours, the peptides have been cleaved from the resin and will pass through the filter

and are placed into a 100 mL round bottom flask, leaving behind the Wang resin. Then, 30 mL cold anhydrous ethyl ether is added to the solution in the round bottom, which results in formation of a white cloudy precipitate as the peptide crashes out of solution. This solution is allowed to remain in the freezer overnight to allow the peptide to crash out of solution completely.

After at least 12 hours in the freezer, the solution is added 1 mL at a time to an Eppendorf tube and centrifuged. The peptide collects at the bottom of the tube and the supernatant is discarded. After all of the solution has been centrifuged, a small amount of the peptide is added to a 50/50 mixture of methanol/water to obtain a concentration of roughly 10^{-5} M. With the addition of 1% formic acid to protonate the peptides, the solution is ready for analysis in the mass spectrometer.

2.1.2 Mass Spectrometry Analysis

Peptide fragmentation studies were performed in a Finnigan Thermo LCQ ion trap mass spectrometer with an electrospray source. The parent mass of the peptide of interest was isolated, then allowed to undergo collision-induced dissociation (CID), a process in which kinetic energy of the peptide is translated into internal energy after collision from with inert gas. CID energy used for fragmentation of the peptide varied depending on the peptide and was adjusted by steps in order to maintain intensity of about 15% of the parent mass. Experimental fragmentation spectra were compared with theoretical fragmentation peaks predicted from the Protein Prospector program v 5.16.0 developed and published by the University of California San Francisco.

2.2 Proteomics

2.2.1 Sample Preparation and Description

The sample used for all proteomics HPLC/MS optimization experiments was a MS Qual/Quant QC Mix purchased from Sigma-Aldrich. As purchased this sample contains 6 human proteins purified and predigested with trypsin as well as 14 stable isotope labeled peptides. The 6 human proteins are present in known varied concentration ranges (range varies 25 fold), pre purified with a C18 LC column. The stable isotope peptides range in concentration over 3 orders of magnitude. Also provided with the sample is a FASTA library file that allows the proteins to be searched for using SEQUEST. The proteins included in the MS Qual/Quant QC Mix and their corresponding molecular weights, concentrations, and stable isotope labeled peptides are listed in the table below.

Protein[1] (UniProt Accession Number)	Calc'd MW (Da)	Approximate protein per vial[2]		Corresponding SIL peptide sequences[3]	SIL Peptide Content per vial (pmol)[4]	Theoretical Ratio[5] Light : Heavy (Protein : SIL Peptide)
		pmol	µg			
Carbonic Anhydrase I (P00915)	28739	100	2.9	GGPFSDSY[R]	100	1
				VLDALQAI[K]	50	2
Carbonic Anhydrase II (P00918)	29115	100	2.9	AVQQPDGLAVLGIFL[K]	10	10
				SADFTNFDP[R]	2	50
NAD(P)H dehydrogenase (P15559)	30736	20	0.62	EGHLSPDIVAEQ[K]	20	1
				ALIVLAHSE[R]	10	2
C-reactive Protein (P02741)	23047	20	0.46	ESDTSYVSL[K]	2	10
				GYSIFSYAT[K]	0.4	50
Peptidyl-prolyl cis-trans isomerase A (P62937)	20176	4	0.08	FEDENFIL[K]	8	0.5
				VSFELFAD[K]	4	1
				TAENF[R]	2	2
Catalase (P04040)	59625	4	0.24	GAGAFGYFEVTHDIT[K]	20	0.2
				FSTVAGESGSADTV[R]	0.4	10
				NLSVEDAA[R]	0.08	50

Table 2.1 Composition of MS Qual/Quant QC Mix sample (obtained from http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/9/msqc1dat.pdf)

The MS Qual/Quant QC Mix was reconstituted as instructed by the Sigma-Aldrich website by adding 20 microliters of water/acetonitrile mixture with 0.1% formic acid. This reconstituted sample was then diluted further with 1 mL of 95/5 water/acetonitrile solution. The protein sample when not in use was stored in freezer at -20 °C.

2.2.2 HPLC

The sample of predigested proteins was inserted directed into the HPLC instrument using an automated sampling device. The sampling device inserted between 2-10 μL at a time into the system. From there, the sample was run through the HPLC column using a time dependent gradient with a solvent mixture of H_2O, acetonitrile, and 0.1% formic acid. Percentage volume of the H_2O and acetonitrile varied based on the time in the run, but ranged from 98% H_2O and 2% acetonitrile to 98% acetonitrile and 2% H_2O. Differential pumping from two solvent sources achieved specified solvent gradients. Data was initially obtained with an ACE 3 C18 30 x 2.1 mm reverse phase HPLC column. This column was later replaced by an ACE Excel 3 superC18 30 x 2.1 mm reverse phase column, which was used to obtain the remainder of the data.

2.2.3 Data Dependent Mass Spectrometry

For the proteomics experiments described in this thesis, a Thermo Finnigan LTQ-XL linear ion trap mass spectrometer was utilized with an electrospray ionization source. The LTQ was run with electrospray voltage of 3-5 kV, sheath gas flow rate of 35, and auxiliary gas flow rate of 30. A blank of 98% water and 2% acetonitrile with 0.1% formic acid was run between all runs of the Qual/Quant QC sample.

2.2.4 Scanning Criteria

The LTQ was run using scanning criteria created in XCalibur. The LTQ first took a full mass scan of the entire sample, then isolated and fragmented in sequence the five most intense peaks from the full scan. This process was then repeated for the entire time period of the HPLC gradient. Depending on the fragmentation mechanism of choice (either CID or both CID and ETD) there were either 4 or 8 product mass scan events. CID was performed using default charge state of 2, isolation width of 2 m/z, normalized collision energy of 35%, activation Q of

0.250, and activation time of 30 ms. ETD was performed using default charge state of 2, isolation width of 2 m/z, and activation time of 100 ms. Scan events were performed in centroid mode. The mass range of the full mass scans was 350 m/z to 1900 m/z.

Data was obtained in conjunction with use of a dynamic exclusion list, which allowed for exclusion from fragmentation events any parent mass that repeated twice within 30 seconds of each other. Once excluded, the mass was added to an exclusion list where it would remain for 180 seconds. There were a maximum of 50 masses that could be added to the exclusion list.

2.2.5 SEQUEST

The resulting data from the entire 63.3 or 90-minute run on the mass spectrometer was then run through the SEQUEST searching database. The data was compared to a FASTA database from Sigma Aldrich, which corresponded to their LC Quant/Qual QC calibration product and contained full sequence information for the six proteins contained within the calibration sample. SEQUEST matched the fragmentation data to this library then provided total number of matched proteins, as well as a confidence value for each protein that corresponds with the certainty by which SEQUEST made the match. In addition, SEQUEST provides a total list of all peptides identified, regardless of whether or not they were eventually matched to a protein in the library. Each peptide is listed with its molecular weight, retention time from the HPLC column as well as an Xcorr value. From this, data was collected corresponding to successful number of proteins identified, time at which the peptides with highest confidence were eluted, and time at which any peptide was eluted. This data was then used to attempt to optimize the procedure in order to maximize total number of proteins identified.

The data was run through SEQUEST with mass range 350 through 5000 Da, activation type CID, through the FASTA database provided by Sigma Aldrich, with Trypsin enzyme digest,

precursor mass tolerance of 1 Da, fragment mass tolerance of 0.6 Da, b-type and y-type ions calculated, and static side chain modification of carbamidomethyl (+57 Da) on all cysteine residues. The one data set obtained using both ETD and CID used activation type ETD and calculated b, y, z, and c ions. All other parameters remained the same.

Chapter 3: Results and Discussion

3.1 Peptide Fragmentation Studies

A total of 8 peptides of form AAXAK, AAXAR, AXAAK, and AXAAR (X= pipecolic acid or proline) were synthesized and analyzed under collision-induced dissociation in the mass spectrometer. The proton affinity for proline is about 937 kJ/mol, pipecolic acid is 944 kJ/mol, lysine is 984 kJ/mol, and arginine is 1025 kJ/mol [3, 37, 38]. The structures for the proline and pipecolic acid containing peptides are shown below in figures 3.1 and 3.2, respectively

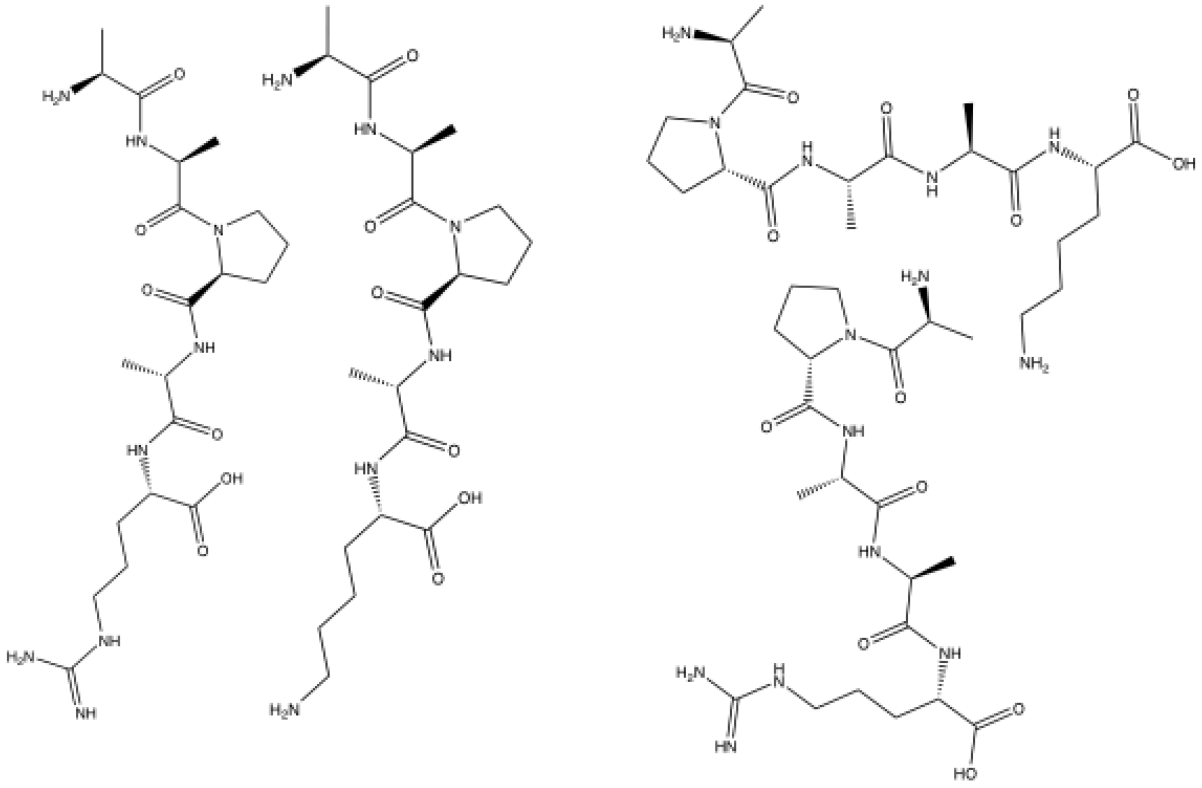

Figure 3.1 Proline-containing pentapeptides synthesized: from left AAPAR, AAPAK; top right APAAK bottom right AAPAR

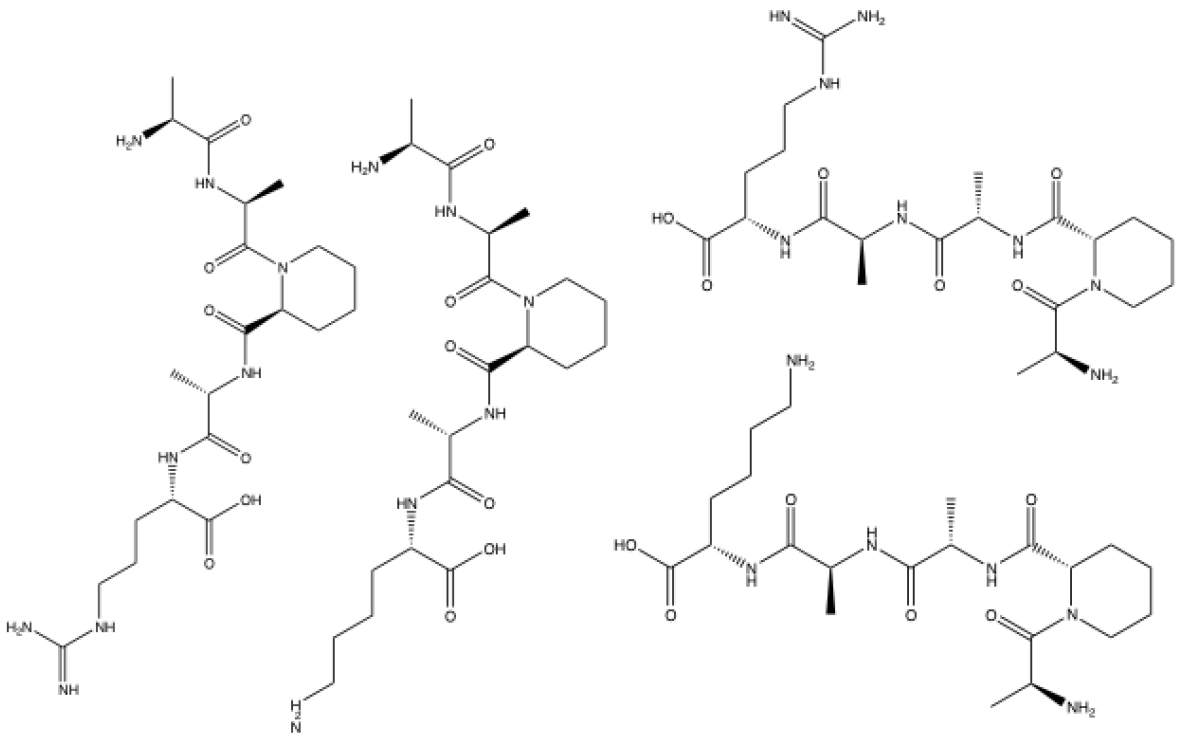

Figure 3.2 Pipecolic acid-containing peptides synthesized: from left AAPipAR, AAPipAK, top right APipAAR bottom right APipAAK.

First, peptides of the form AAXAK and AAXAR were studied.

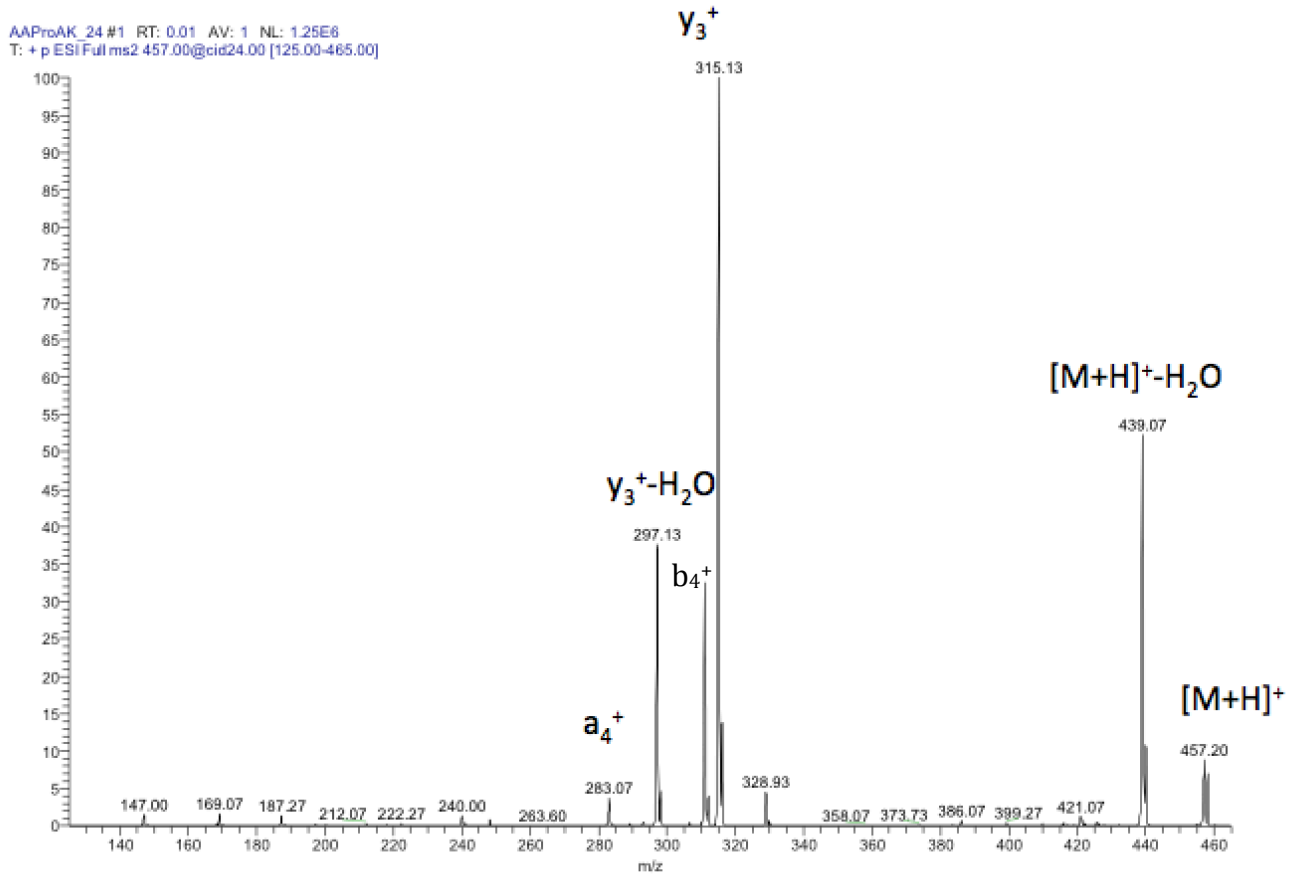

Figure 3.3 AAPAK singly charged 24% CID spectrum

The fragmentation of AAPAK shows three main product peaks, loss of water from the parent mass, a y_3^+ ion and y_3^+ ion with neutral loss of water, and a b_4^+ at 311 m/z. This is very consistent with the results expected when considering the proline effect.

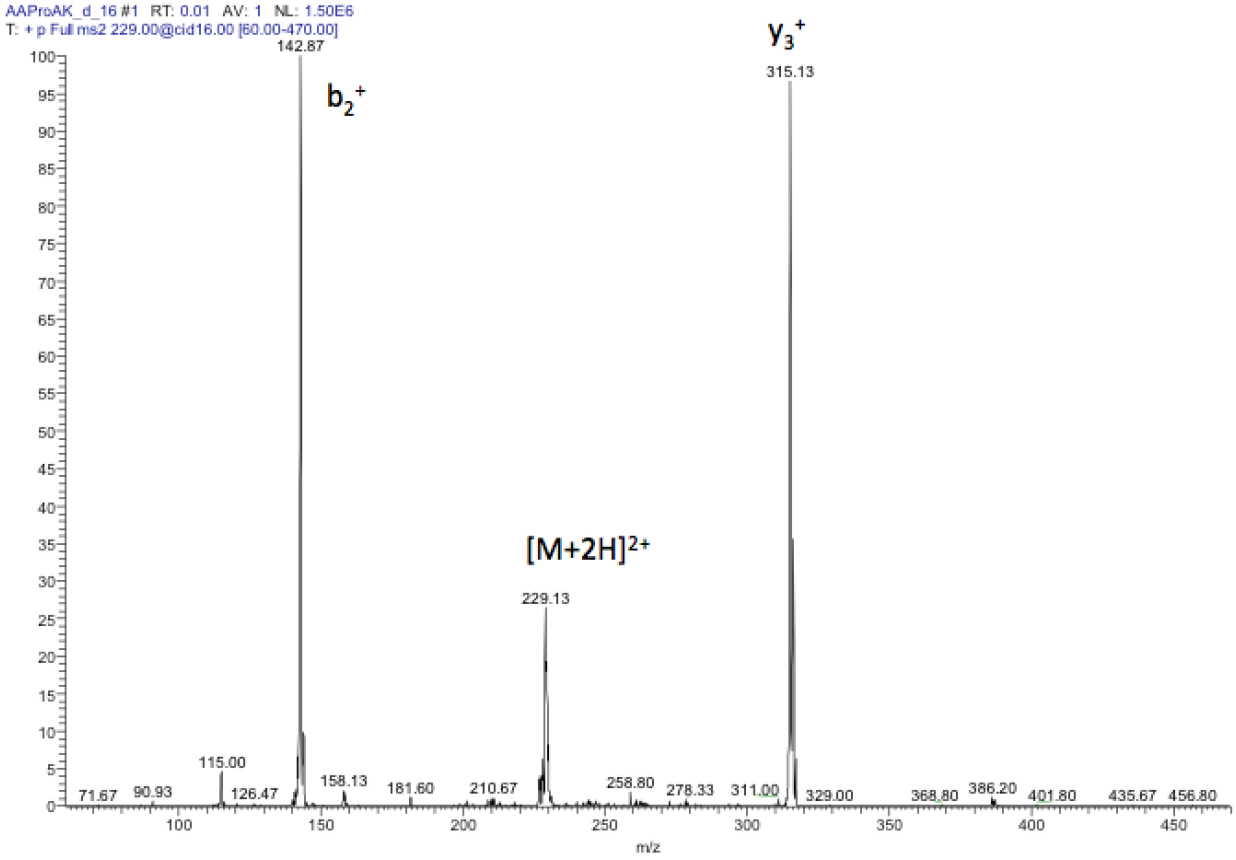

Figure 3.4 AAPAK doubly charged 16% CID spectrum

However, in the doubly charged version of AAPAK, there are now two main product peaks, the y_3^+ and b_2^+ ion. This means the addition of a second proton is forcing the formation of b-type ions and this second proton does not necessarily remain on the lysine side chain, where the proline effect is still directing cleavage to produce the y_3^+ ion as well. If the second proton were remaining on the side chain of lysine, the spectrum would show a doubly charged y-type ion.

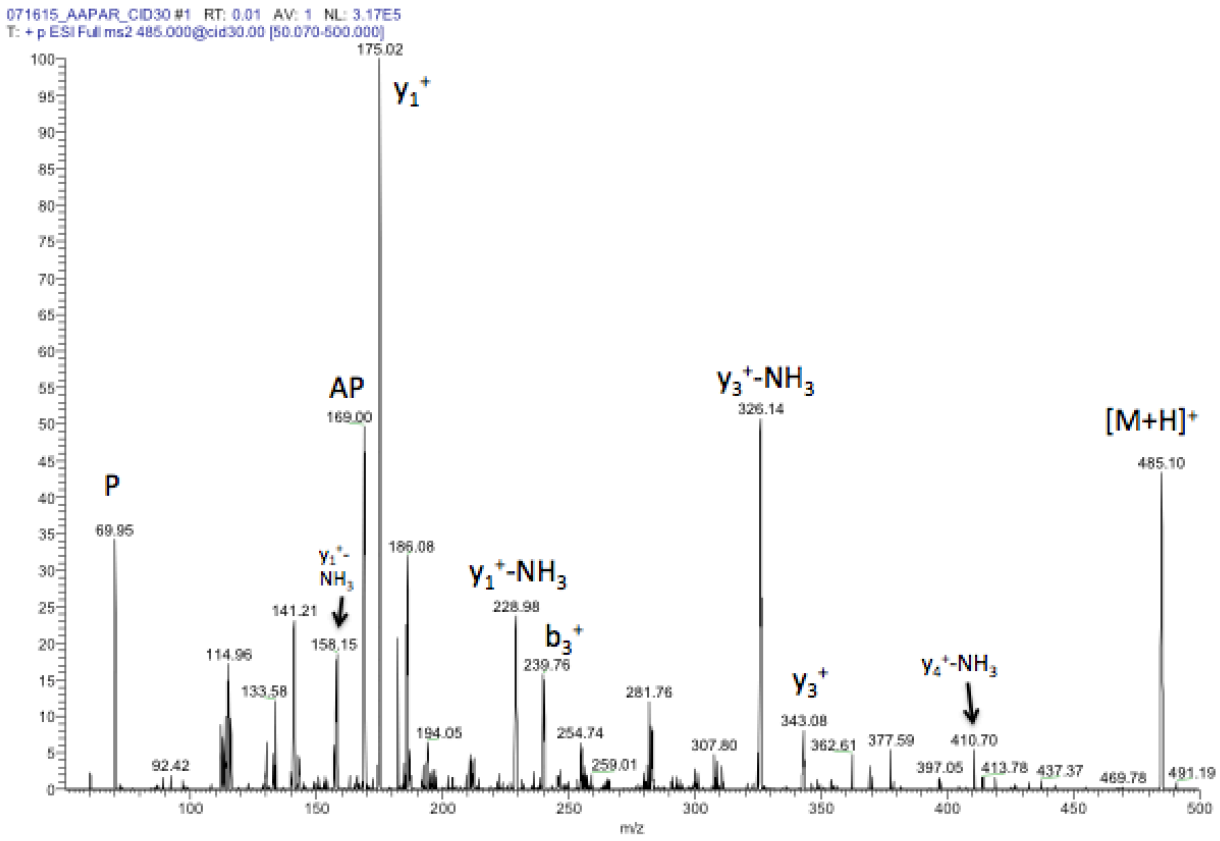

Figure 3.5 AAPAR singly charged 30% CID spectrum

With replacement of the lysine residue with an arginine residue for the AAPAR peptide, there is a marked increase of product ions formed. The majority of the product peaks are y-type ions, several with neutral loss of ammonia. However, there is noted the presence of one b_3^+ ion peak, as well as internal fragmentation of an alanine-proline segment (denoted as AP) as well as a fragment resulting from just proline (denoted as P). A spectrum in which the proline-effect was active would be dominated by the y_3^+ ion, thus the basicity of the arginine residue is affecting the conformational stability of the peptide and causing a more random fragmentation pattern.

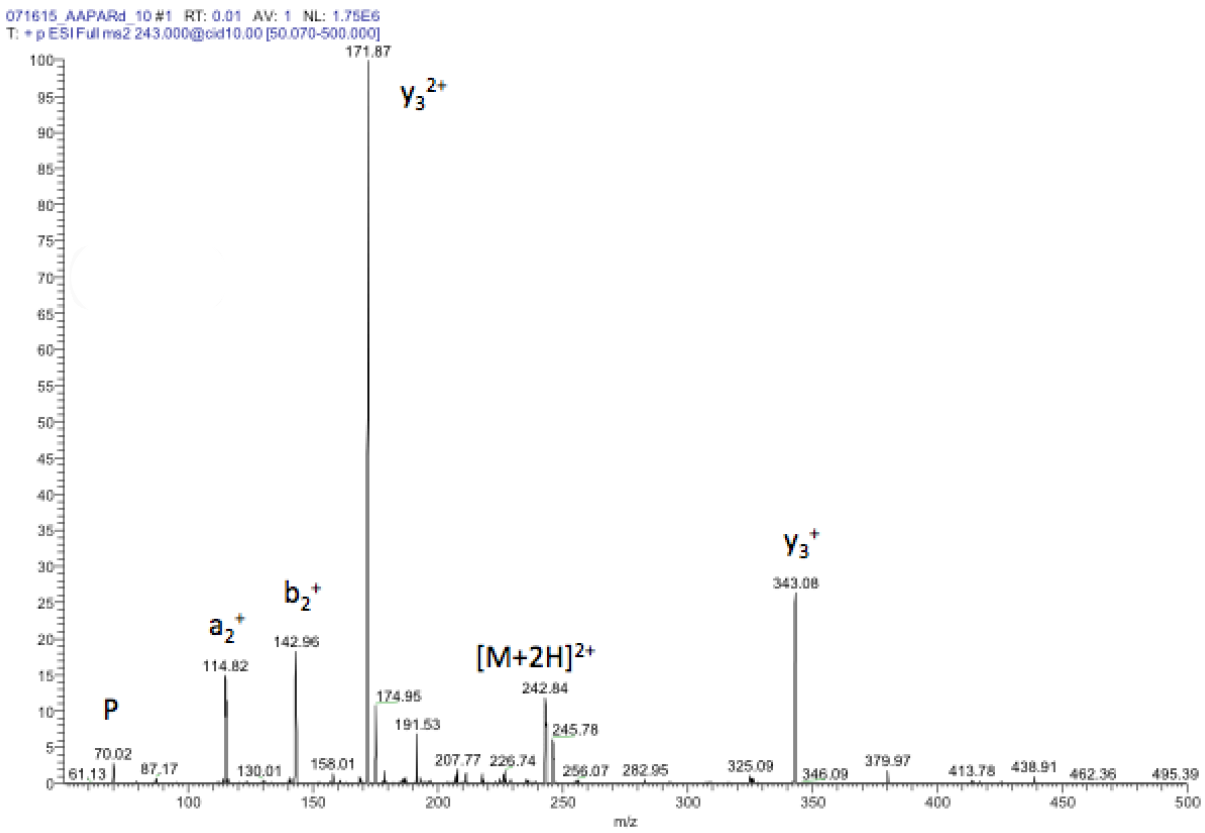

Figure 3.6 AAPAR doubly charged 10% CID spectrum

In the doubly charged spectrum of AAPAR, there is again y_3^+ ion formation however now the primary product is the y_3^{2+} ion fragment. In contrast to the lysine version of this peptide, this means that the second proton on the peptide is no longer free to move about the peptide but likely remains sequestered on the arginine residue. This is consistent with the fact that the proton affinity for arginine is greater than that of lysine, therefore arginine has greater tendency to keep the additional proton upon the induction of fragmentation. There is also presence of a small amount of b_2^+ and a_2^+ fragments, as well as a fragment from proline. This spectrum is consistent with the proline effect, because the most abundant fragments are y_3^+-derived ions.

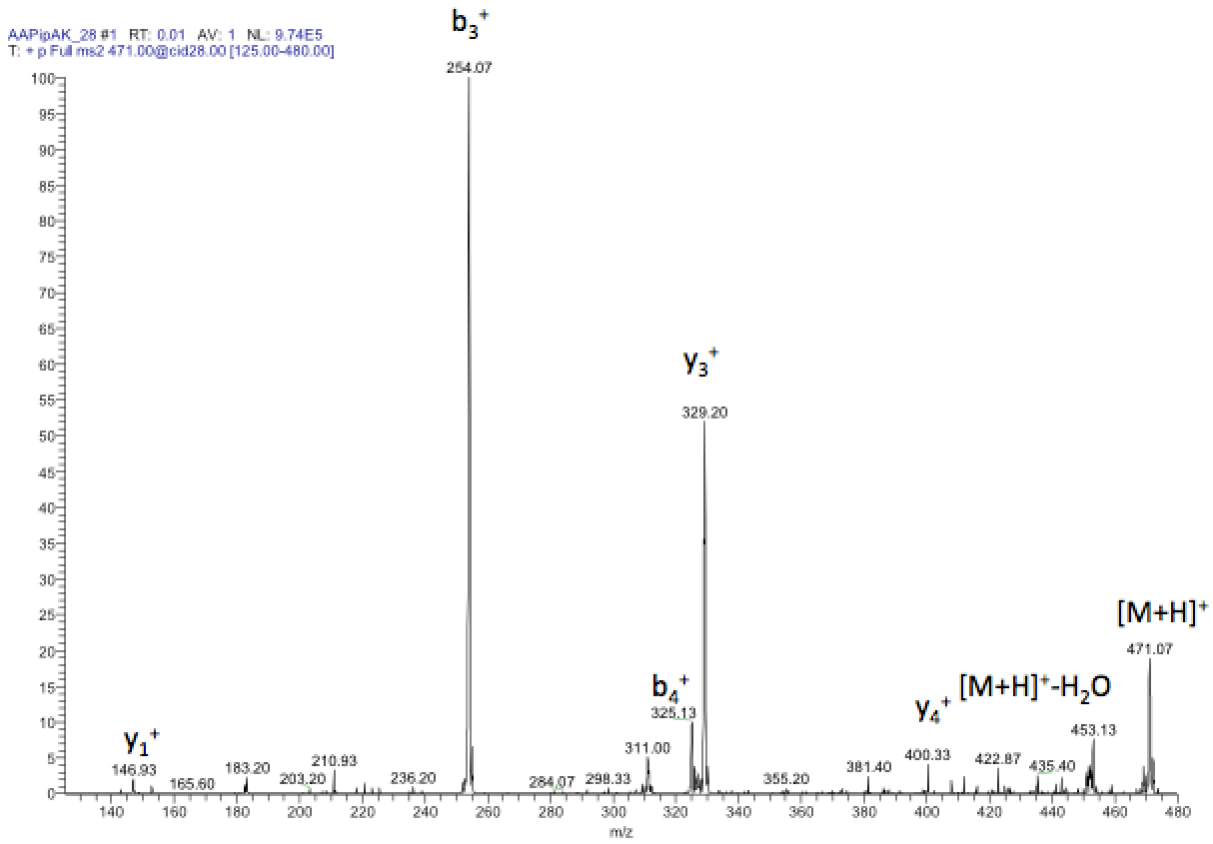

Figure 3.7 AAPipAK singly charged 28% CID spectrum

In the singly charged peptide AAPipAK, there is primary production of both b_3^+ and y_3^+ fragments, as well as some minor contribution of b_4^+, y_4^+, and y_1^+ fragments. As there is presence of both b-type and y-type ions, the pipecolic acid effect must be somewhat dampened by the presence of the lysine residue, but is still present because the relative abundance of the b_3^+ ion is higher than that of the y_3^+ ion.

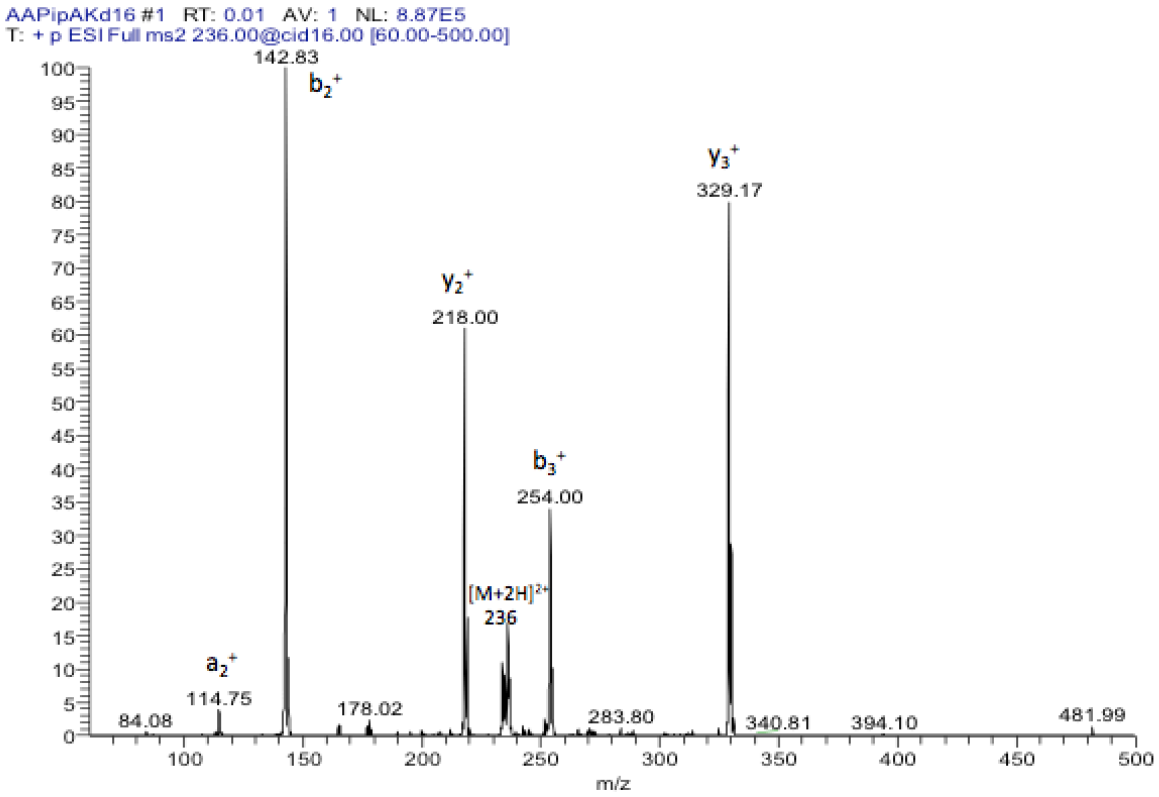

Figure 3.8 AAPipAK doubly charged 16% CID spectrum

In the AAPipAK doubly charged spectrum, there are fairly equal amounts of both the b_2^+ and y_3^+ ions, as well as some lesser amounts of the b_3^+ and y_2^+ ions. In addition, there is a small amount of a_2^+ ion. Here we see that in comparison to the singly charged spectrum, the pipecolic acid effect is even more significantly dampened. The b_2^+ ion is still the most abundant peak and the abundance of y_3^+ is much greater than the abundance of the b_3^+ ion, which is a significant change from the singly-charged spectrum as shown above. Notably, there is a lack of doubly-charged fragments in this spectrum. Upon cleavage of the backbone to form the b_3^+ ion, the second proton likely remains sequestered on the lysine side chain because it has a greater proton affinity than the remainder of the residues, which produces the complementary y_2^+ ion.

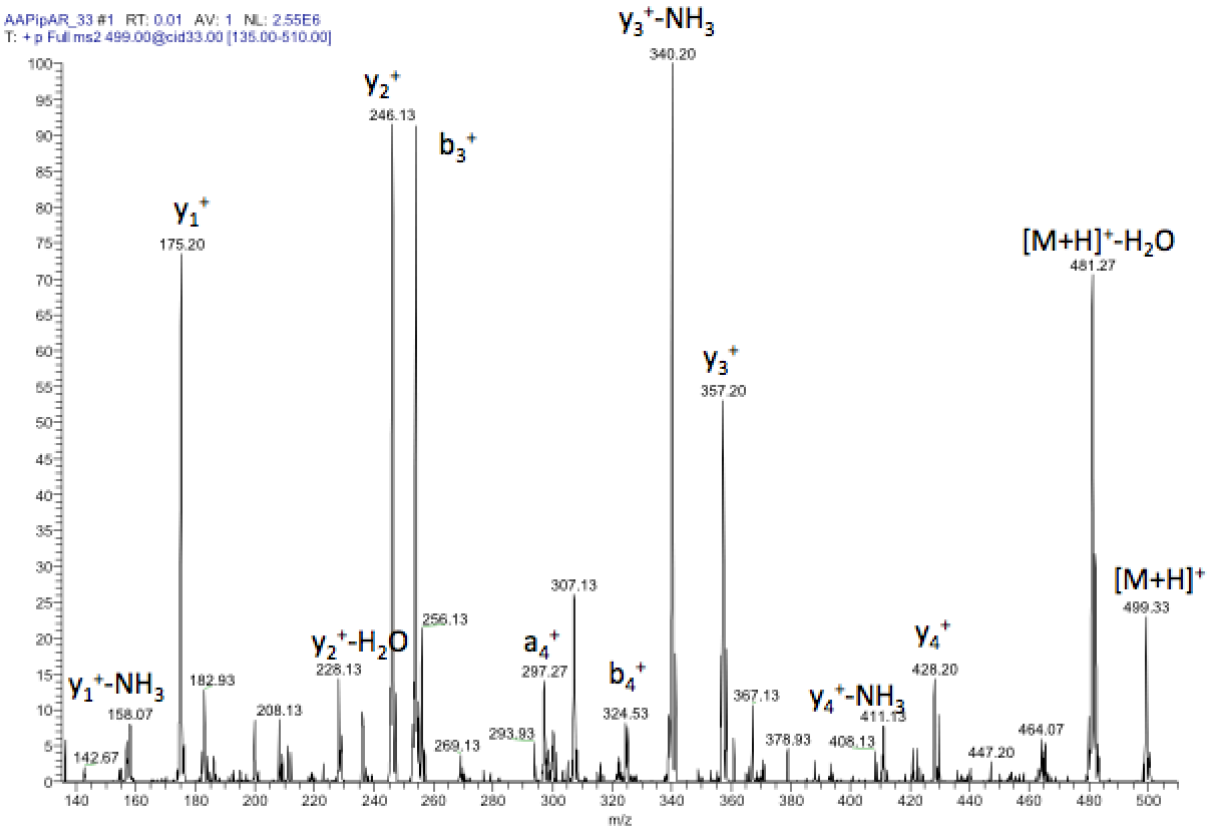

Figure 3.9 AAPipAR singly charged 33% CID spectrum

Similar to the proline version of this peptide, with singly charged AAPipAR there is again a large number of fragment ions produced upon CID. Primary product fragments include y_4^+, y_3^+, y_3^+ minus ammonia, y_2^+, and y_1^+. There is also a significant amount of b_3^+ product ion as well as minor production of b_4^+ and a_4^+ ions.

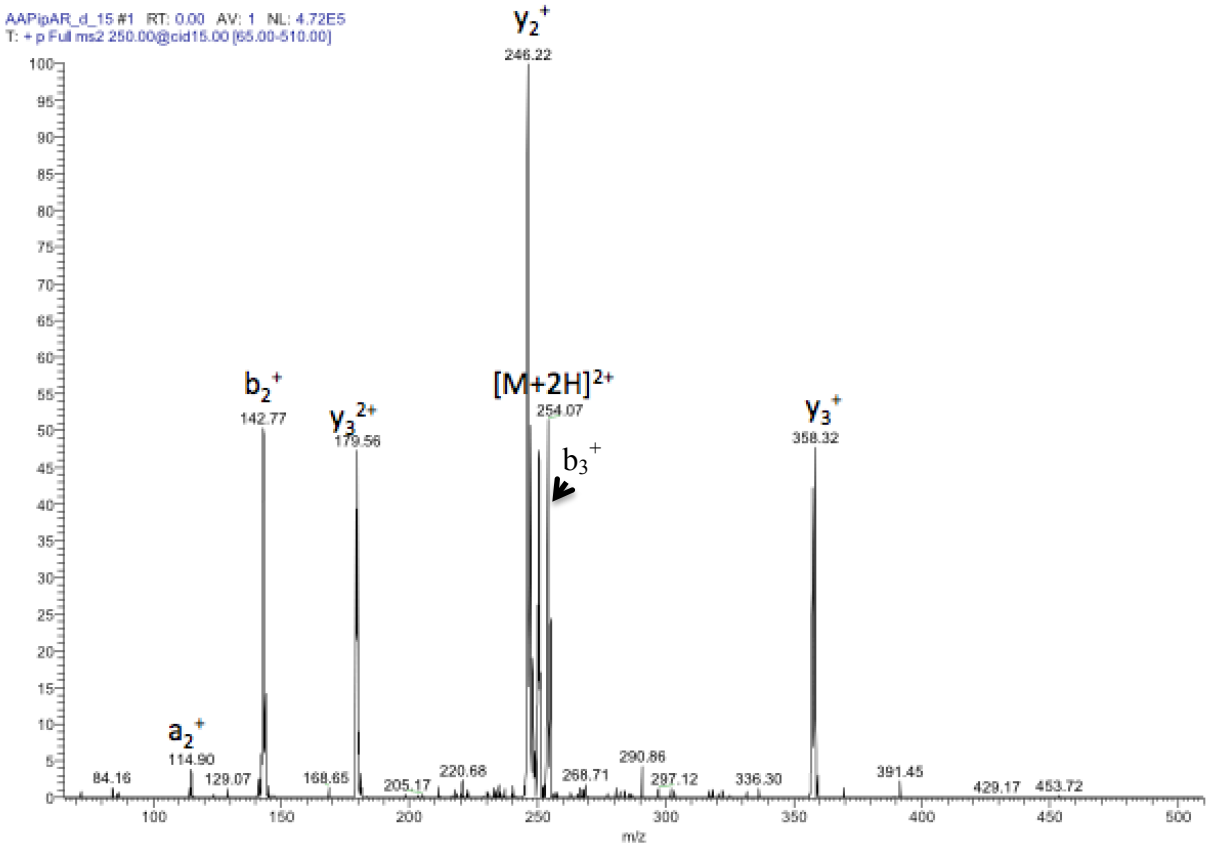

Figure 3.10 AAPipAR doubly charged 15% CID spectrum

In the doubly charged fragmentation of AAPipAR, the major product peaks include singly charged y_2^+, y_3^+, b_3^+ and b_2^+ fragments as well as y_3^{2+} fragment. There is a very minor contribution from an a_2^+ ion. The most abundant ion is the y_2^+, showing that the pipecolic acid effect is active, though not exclusive. The presence of the doubly charged y-type ion means that the proton affinity of arginine when separated by one alanine residue from pipecolic acid is high enough to sequester the second proton in the peptide on the arginine side chain. Therefore, this second proton must be free to participate in fragmentation along the backbone of the peptide.

Next, peptides of the form AXAAR and AXAAK (X= pipecolic acid or proline) were studied.

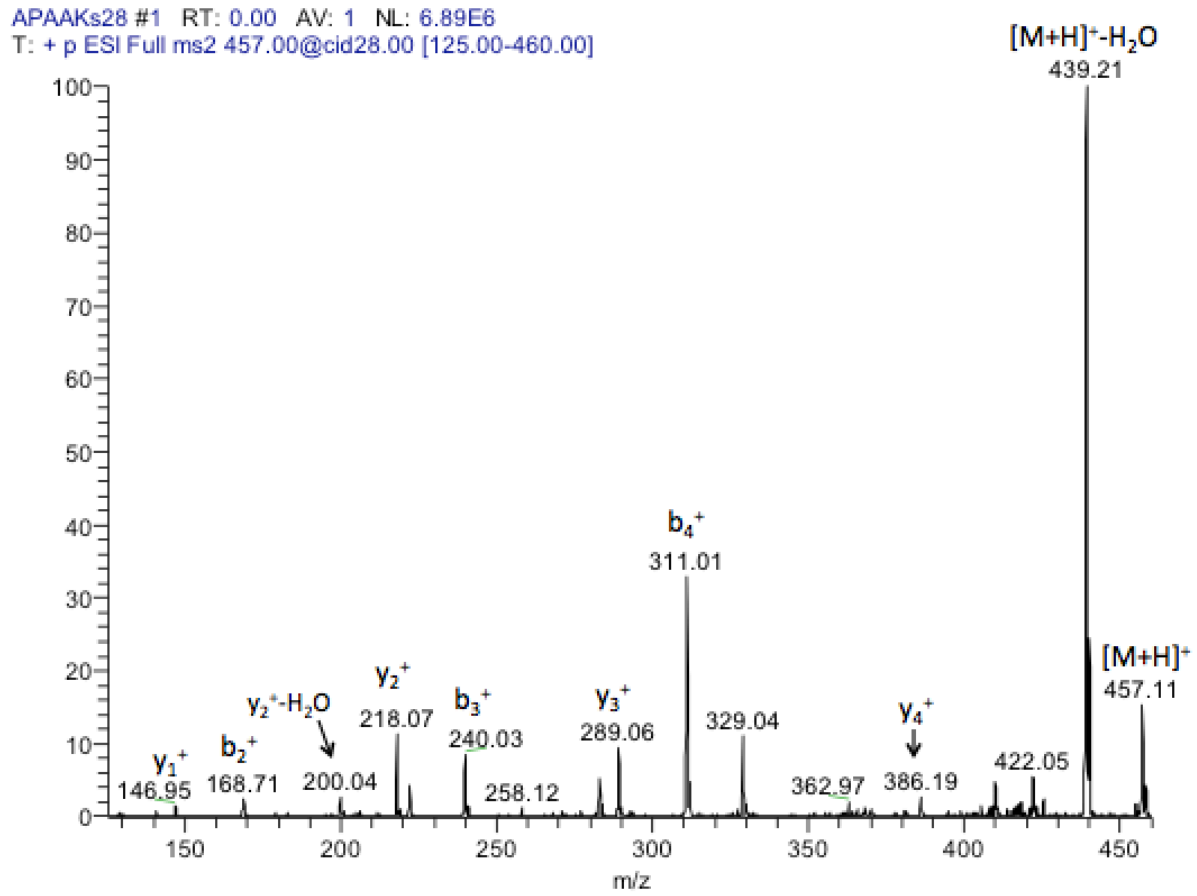

Figure 3.11 APAAK singly charged 28% CID spectrum

Analysis of the spectrum above for singly charged APAAK is somewhat limited, as the primary product ion upon CID is the parent mass with neutral loss of water. The second most abundant peak is that of the b_4^+ ion, which is not consistent with the proline effect. The spectrum does show small amounts of y_4^+, y_3^+, y_2^+, and y_1^+ ions, however they are in equal abundance with the b_3^+ and b_2^+ ions, and the b_4^+ ion dominates overall.

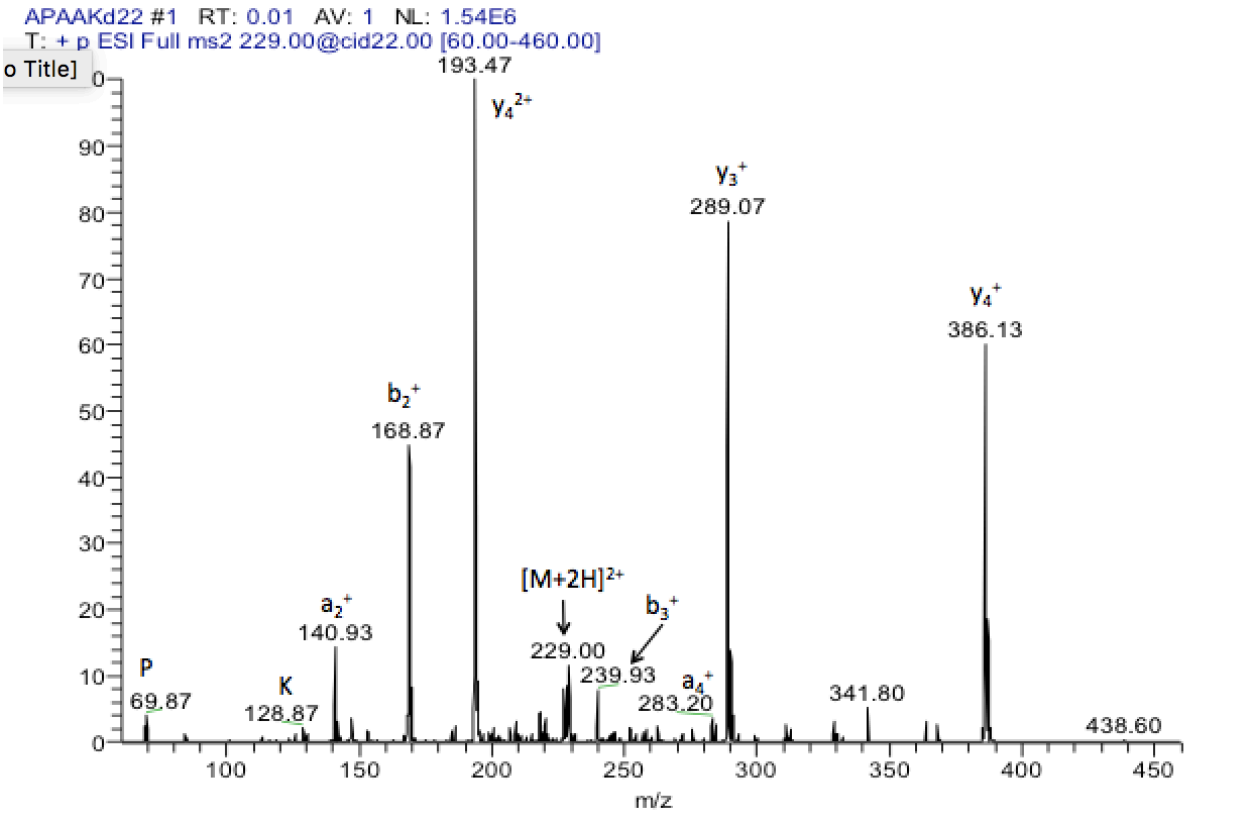

Figure 3.12 APAAK doubly charged 22% CID spectrum

In the spectrum for doubly charged APAAK, the most abundant peaks are that of the y_4^+, y_3^+, and y_4^{2+} ions. There is also a lesser amount of b_2^+ ion and minimal amounts of the b_3^+ and a_4^+ ions and some fragmentation resulting from the proline and lysine residues (denoted by P and K on the spectrum). The presence of the y_4^{2+} ion likely means that the second proton here is remaining sequestered on the lysine residue during fragmentation.

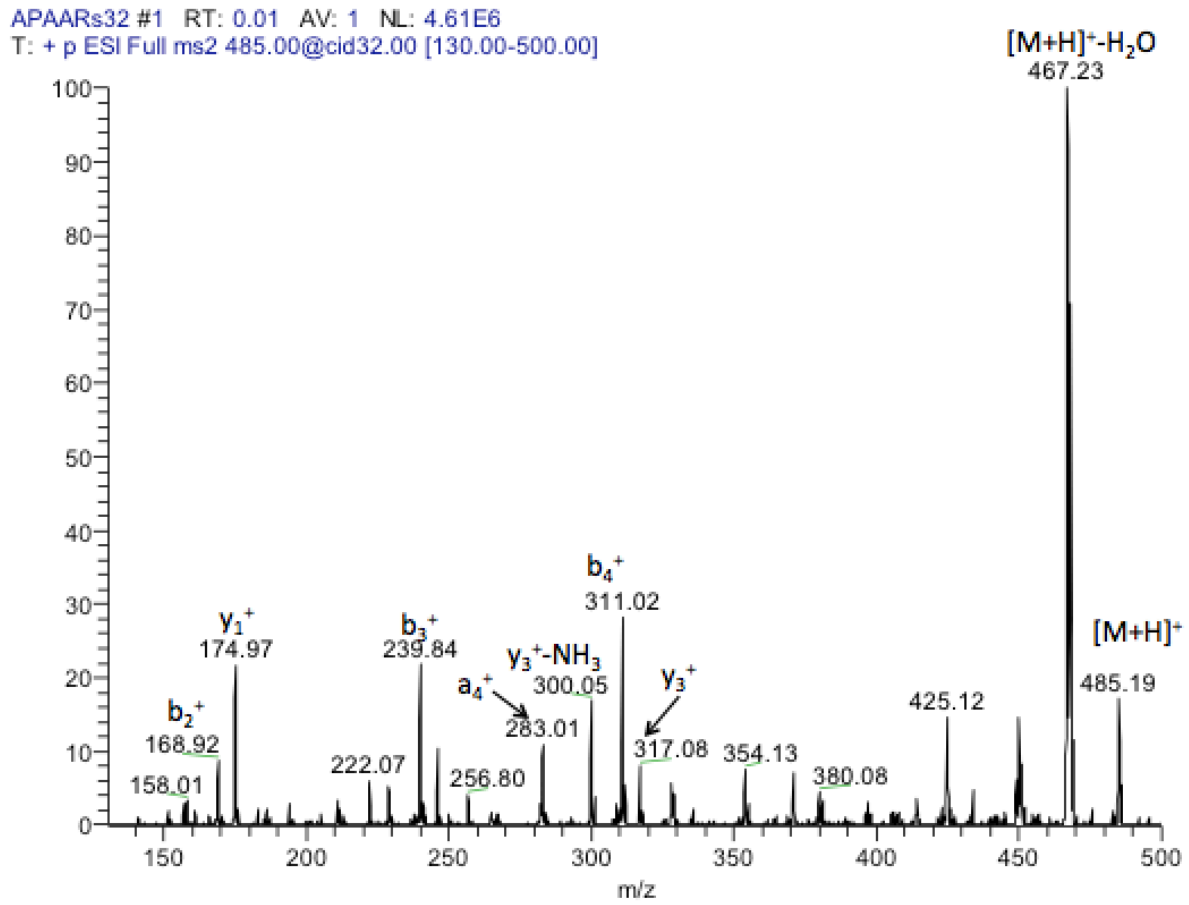

Figure 3.13 APAAR singly charged 32% CID spectrum

In the spectrum for singly charged APAAR, similarly to the above spectrum for singly charged APAAK, the primary product peak is that of loss of neutral water from the parent mass. There is also production of numerous fragments both from b-type and y-type ions in similar abundance. The analysis is again limited as the abundance of all product ion peaks is small in comparison to that of the $[M+H]^+$-H_2O ion.

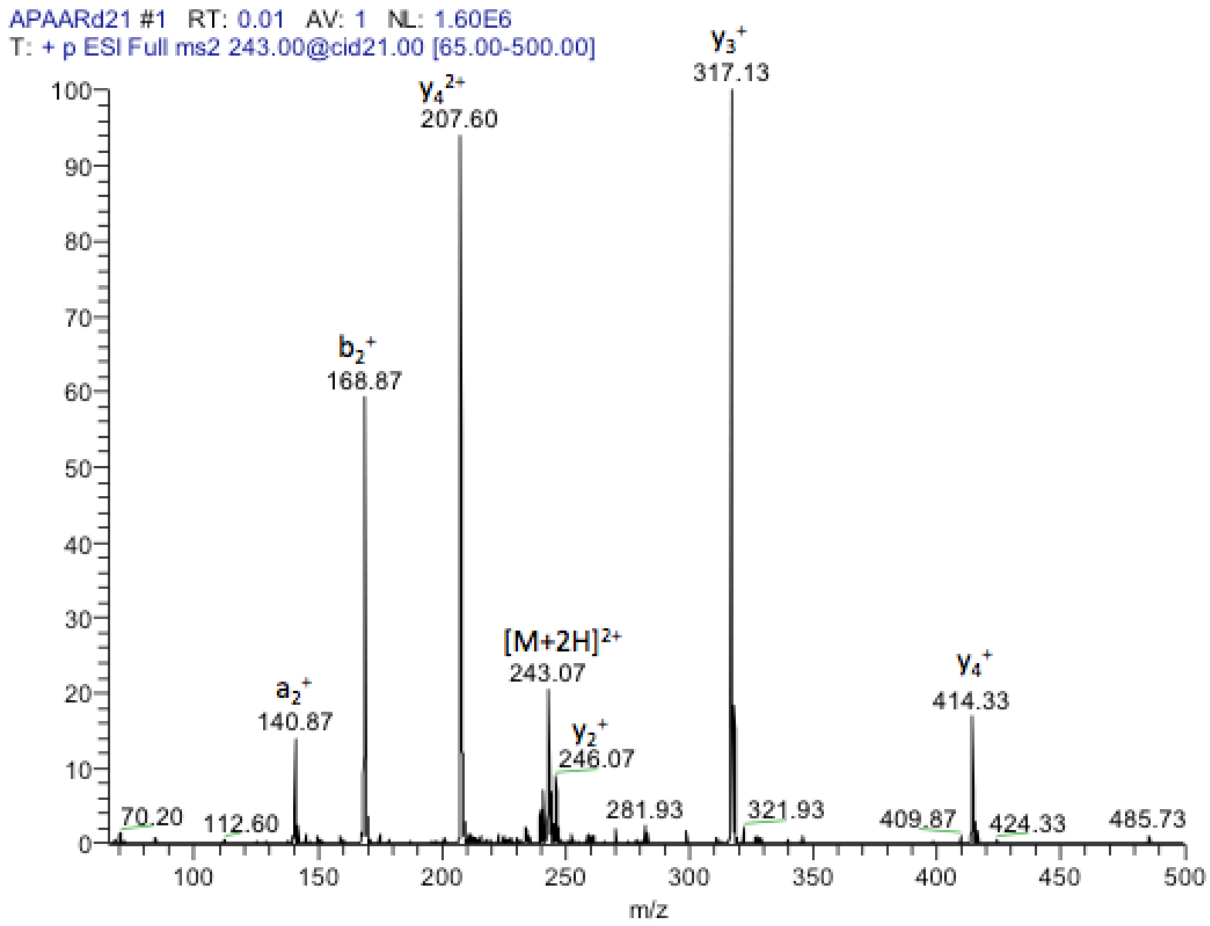

Figure 3.14 APAAR doubly charged 21% CID spectrum

In the spectrum of doubly charged APAAR, there is a pattern very similar to that of the doubly charged APAAK with primary production of y_4^{2+}, y_3^{+}, and y_4^{+} ions and a lesser amount of b_2^{+} and a_2^{+} ions. As compared to APAAK, there is a lesser amount of y_4^{+} ion, because the arginine residue has a greater proton affinity than lysine and is therefore more likely to retain the second proton on the y_4^{2+} ion.

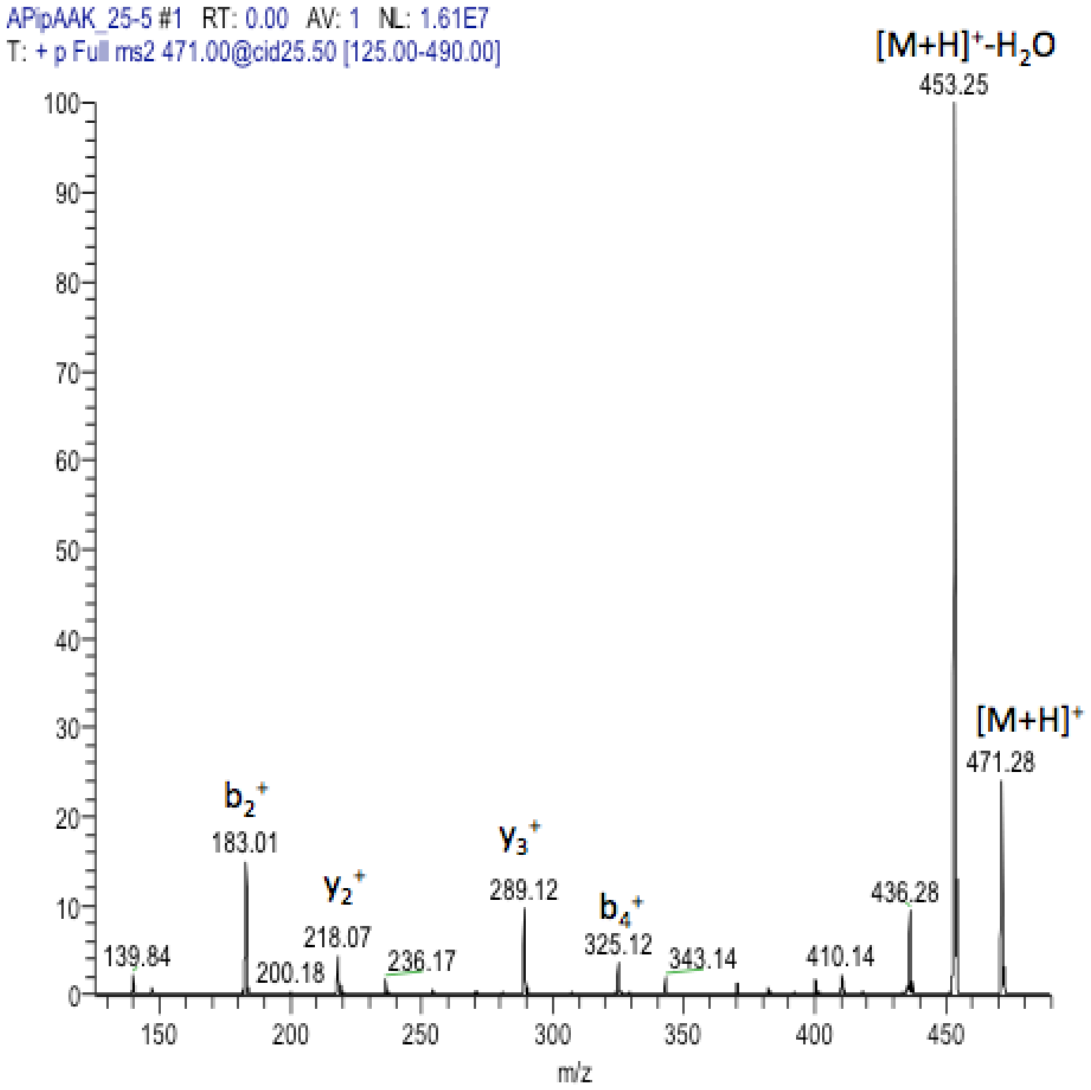

Figure 3.15 APipAAK singly charged 25% CID spectrum

In singly charged APipAAK, analysis of product ions is difficult, as the dominating peak in this spectrum is that of the parent mass with loss of neutral water. There is minor contribution of b_2^+, b_4^+, y_2^+, and y_3^+ ions. The production of b_2^+ and y_3^+ ions is consistent with the pipecolic acid effect.

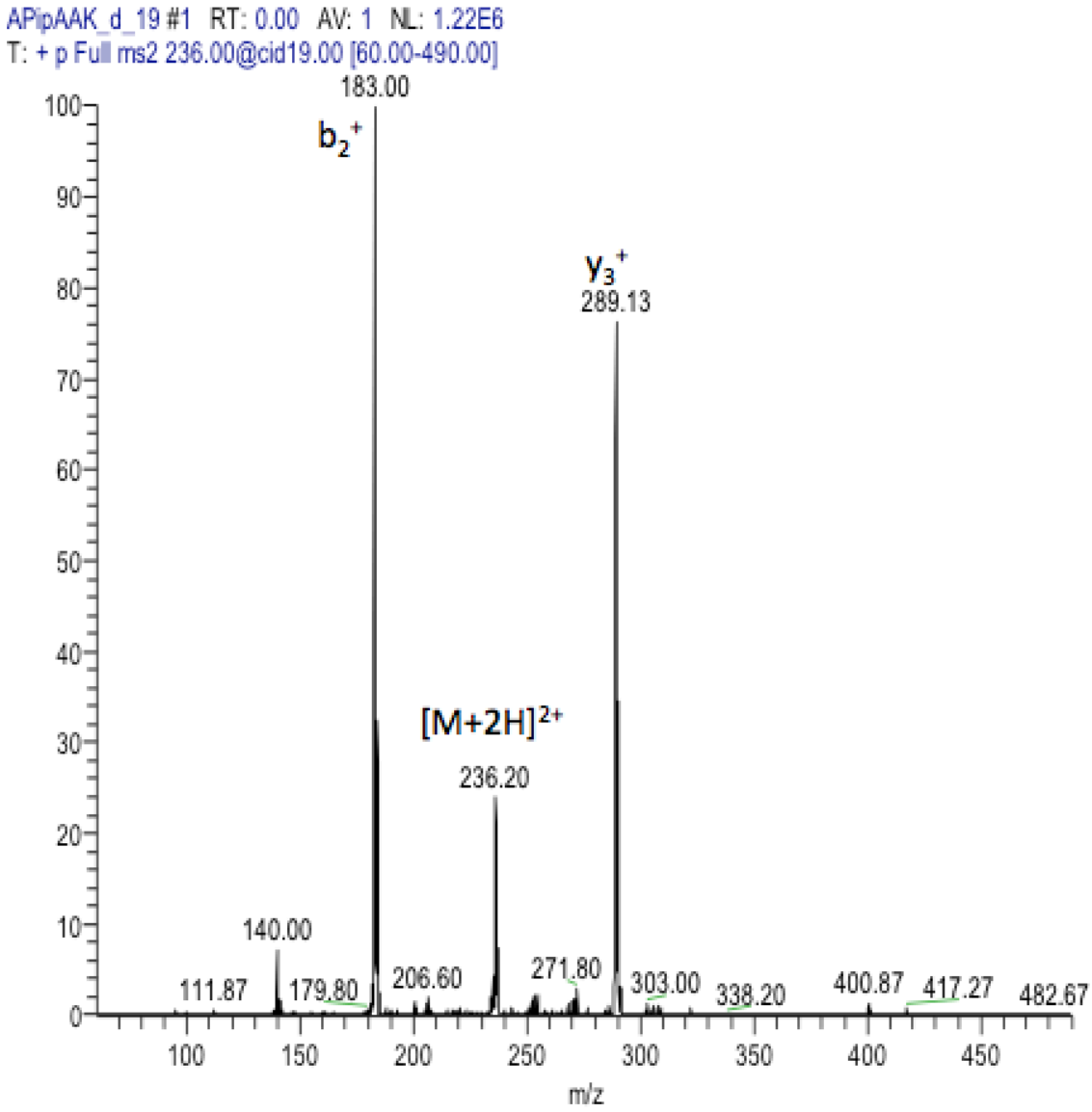

Figure 3.16 APipAAK doubly charged 19% CID spectrum

Similar to the previous doubly charged pipecolic acid peptide, in the spectrum of doubly charged APipAAK, there are primarily both b-type and y-type fragments, all arising from the pipecolic acid effect. Unlike the arginine version of this peptide however, there are only b_2^+ and y_3^+ fragments and no doubly charged fragments, so the second proton is likely not being sequestered on the lysine residue. This makes sense, because the proton affinity of arginine is greater than that of lysine and therefore is more likely to remain on the side chain of arginine during fragmentation.

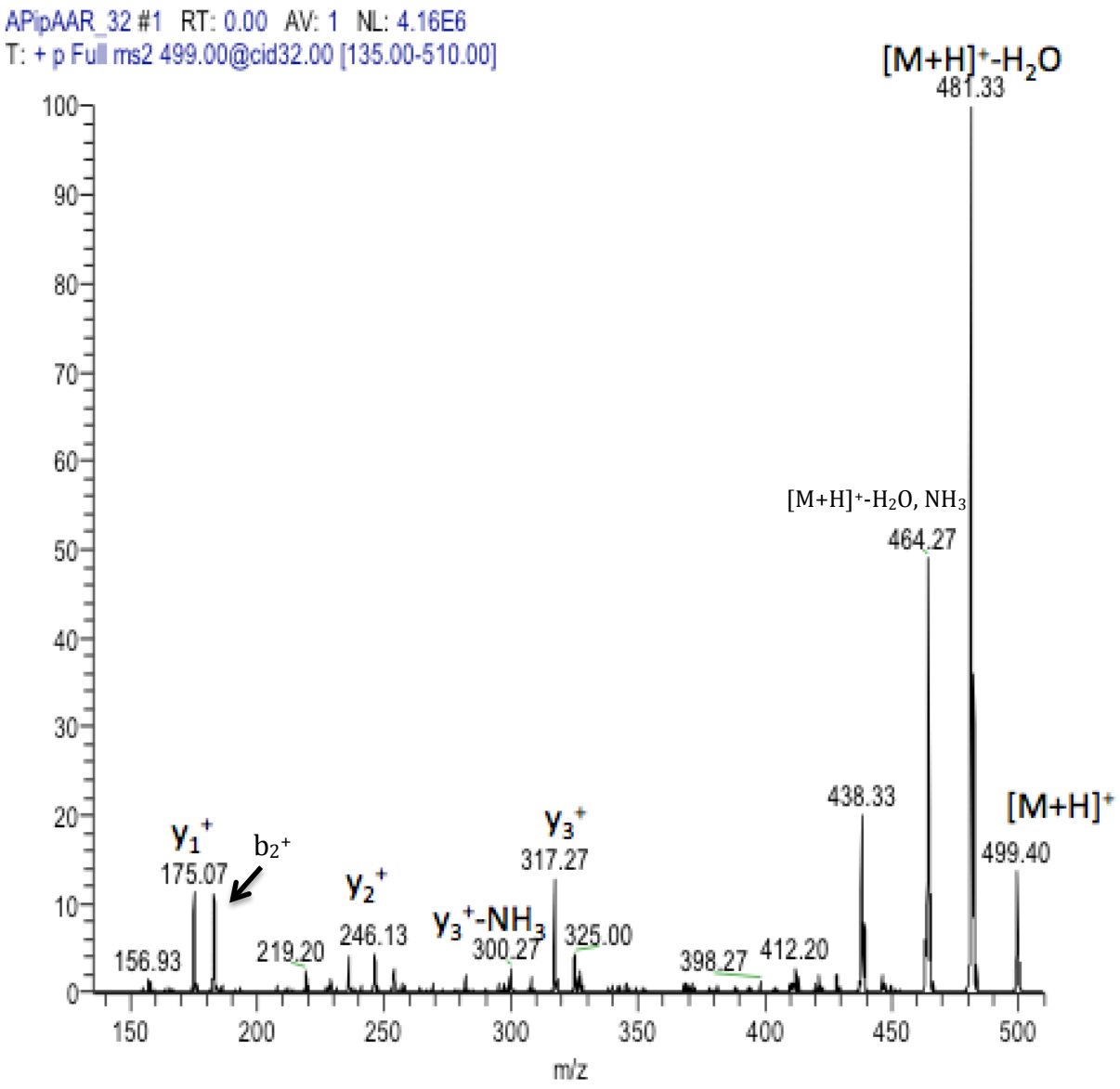

Figure 3.17 APipAAR singly charged 32% CID spectrum

In the spectrum of APipAAR singly charged, main product ions are the b_2^+, y_1^+, y_2^+, and y_3^+ fragments. The peak at 438 was unable to be identified based on predicted fragmentation patterns under CID.

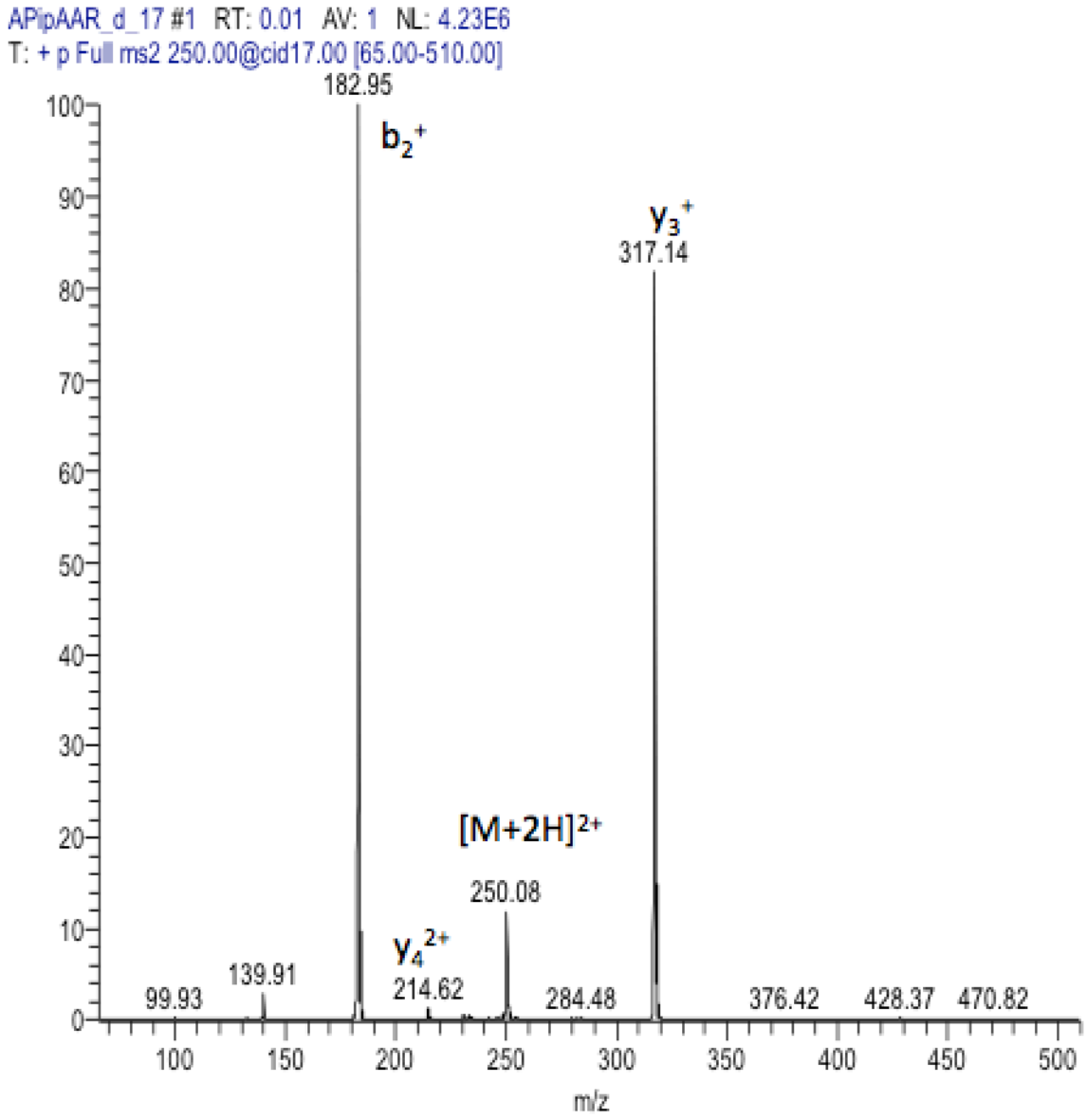

Figure 3.18 APipAAR doubly charged 17% CID spectrum

Just like the spectrum for AAPipAK doubly charged, in the fragmentation spectrum for APipAAR doubly charged there is formation of primarily b_2^+ and y_3^+ fragments, without any doubly charged fragments, consistent with the pipecolic acid effect. This is in contrast to the AAPipAR doubly charged spectrum where there was formation of a doubly charged y-type ion. This means that changing the position of the pipecolic acid residue alters the movement of the

second proton and somehow allows it to participate in fragmentation without being isolated to the arginine side chain.

3.2 Proteomics

An initial solvent gradient used was developed and optimized by a previous lab for a different instrument. This initial solvent gradient lasted for a total of 90 minutes and was as follows, where Pump B refers to the solvent consisting of 98% ACN, 2% H_2O, and 0.1% formic acid, and the remainder of the percentage comes from pump A which consists of 98% H_2O, 2% ACN, and 0.1% formic acid

5 min – 5% Pump B
5 min – 5%-25% Pump B
10 min – 25%-35% Pump B
50 min – 35%-65% Pump B
10 min – 65%-85% Pump B
5 min – 95% Pump B
5 min – 5% Pump B

A graphical representation of this gradient curve is shown below.

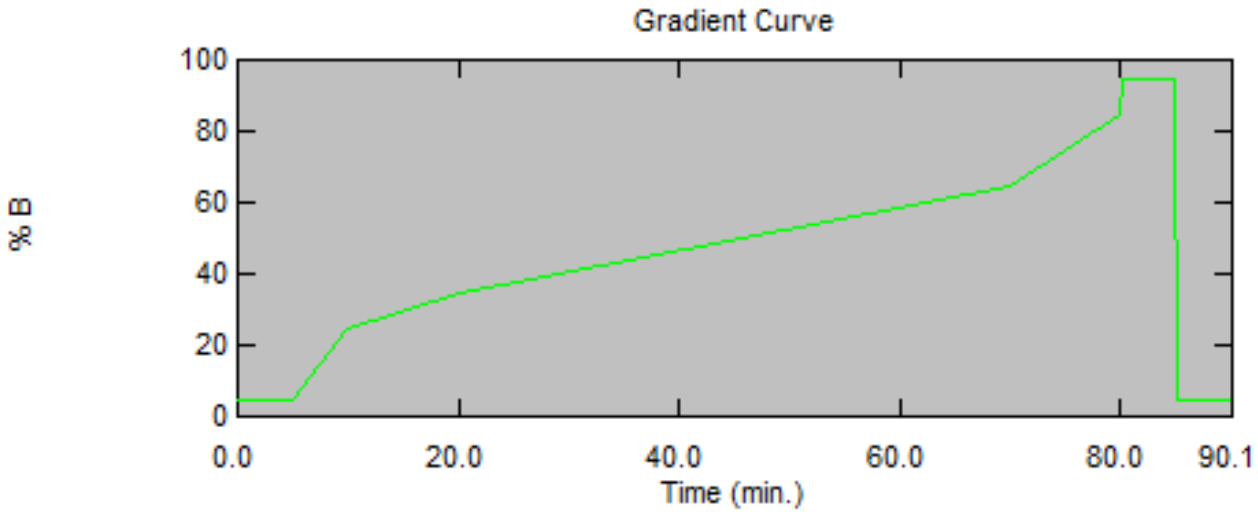

Figure 3.19 HPLC 90 minute gradient curve

The first sets of data collected were run through a ACE 3 C18 column that had been obtained previously in the lab, however the integrity of this column was unknown, as the end caps had been left off of the column during storage. The first run on this column was a 10 microliter injection, set at a flow rate of 0.2 mL/min, with sheath gas flow of 60, and auxiliary gas flow of 20.

The chromatogram from this initial run is as shown below.

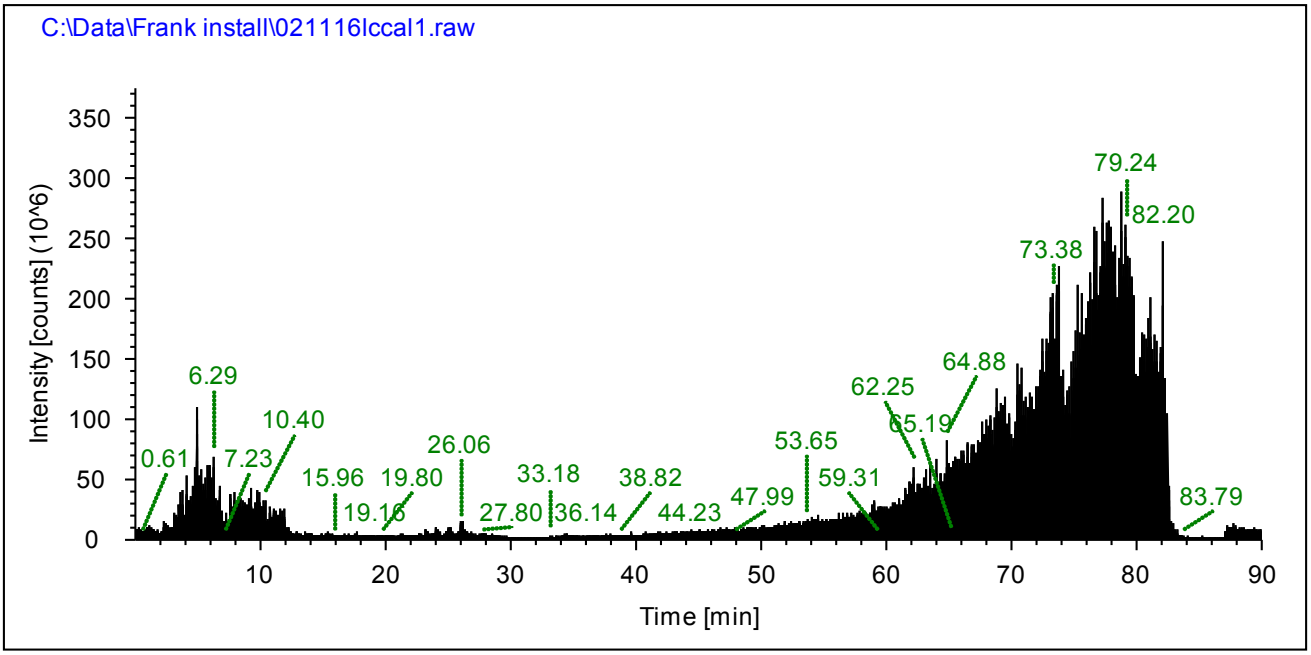

Figure 3.20 HPLC chromatogram using ACE 3 C18 column and 0.2 mL/min flow rate (90 minute gradient)

The data was then run through SEQUEST with specifications as described above. Two proteins were identified out of six, carbonic anhydrase 1 and 2 with coverage percentages of 56.15% and 70.37%, respectively. SEQUEST identified 213 peptides, with 56 peptides matched to a protein. SEQUEST matched 4 peptides with XCorr value greater than 1 to carbonic anhydrase 1: VLDALQAIK, HDTSLKPISVSYNPATAK, YSSLAEAASK, and

ADGLAVIGVLMK. There were 2 peptides matched to carbonic anhydrase 2 with greater than 1 XCorr value: VVDVLDSIK and NRQIKASFK.

A plot with retention time of peptides matched with proteins versus m/z is shown below.

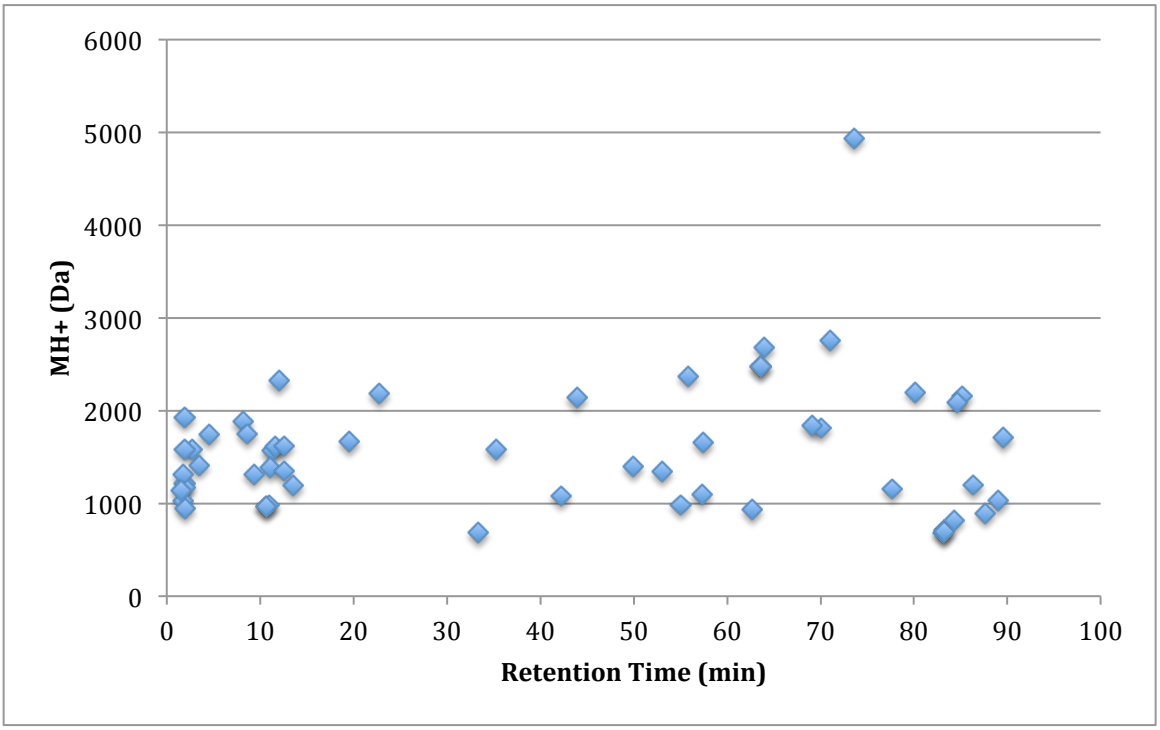

Figure 3.21 Retention Time (min) vs MH+ (Da) for ACE 3 C18 column with 0.2 mL/min flow rate (90 minute gradient)

From this plot, there are peptides being eluted from the column throughout the 90 minutes of the run, however there are is large cluster that appear at the beginning of the run from about 0-10 minutes. Otherwise, peptides eluting from about 30 minutes through the end at 90 minutes seem to be fairly evenly distributed. The overall mass of all of the peptides eluted and identified is small in comparison to later runs, with most peptides below 2000 Da. There are outliers to this average, notably the peptide with mass 4936 Da that eluted at 73.61 minutes.

With a 10 microliter injection of the same gradient at flow rate of 0.8 mL/min, this yielded the following chromatogram:

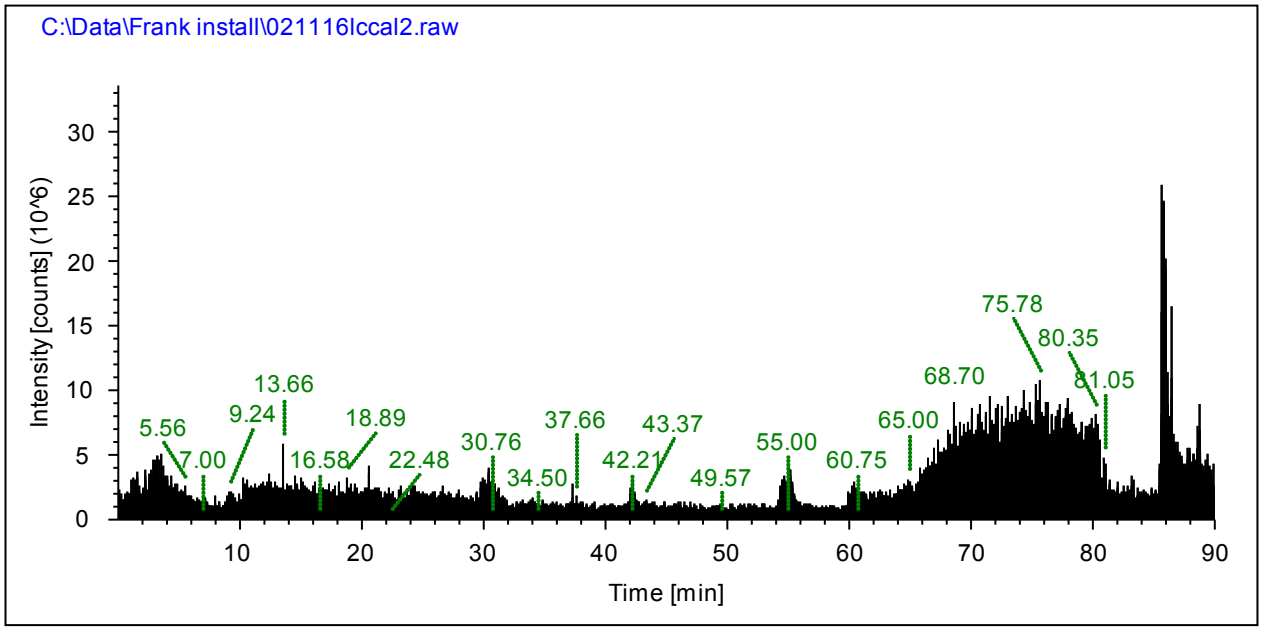

Figure 3.22 HPLC chromatogram using ACE 3 C18 column and 0.8 mL/min flow rate (90 minute gradient)

After running the data from this run in SEQUEST, 2 proteins were identified, carbonic anhydrase 1 and 2, with coverage value of 65.38% and 62.93%, respectively. In total, SEQUEST identified a total number of 197 peptides, however it only matched 49 peptides to a protein. SEQUEST matched 2 peptides to carbonic anhydrase 2 with greater than 1 XCorr value, VVDVDSIK and EPISVSSEQVLIK. SEQUEST matched 1 peptide to carbonic anhydrase 1 with XCorr greater than 1, VLDALQAIK, though it also ranked peptide YSSLAEAASK with "good" matching to that protein. A plot with retention time of peptides matched with proteins versus m/z is shown below.

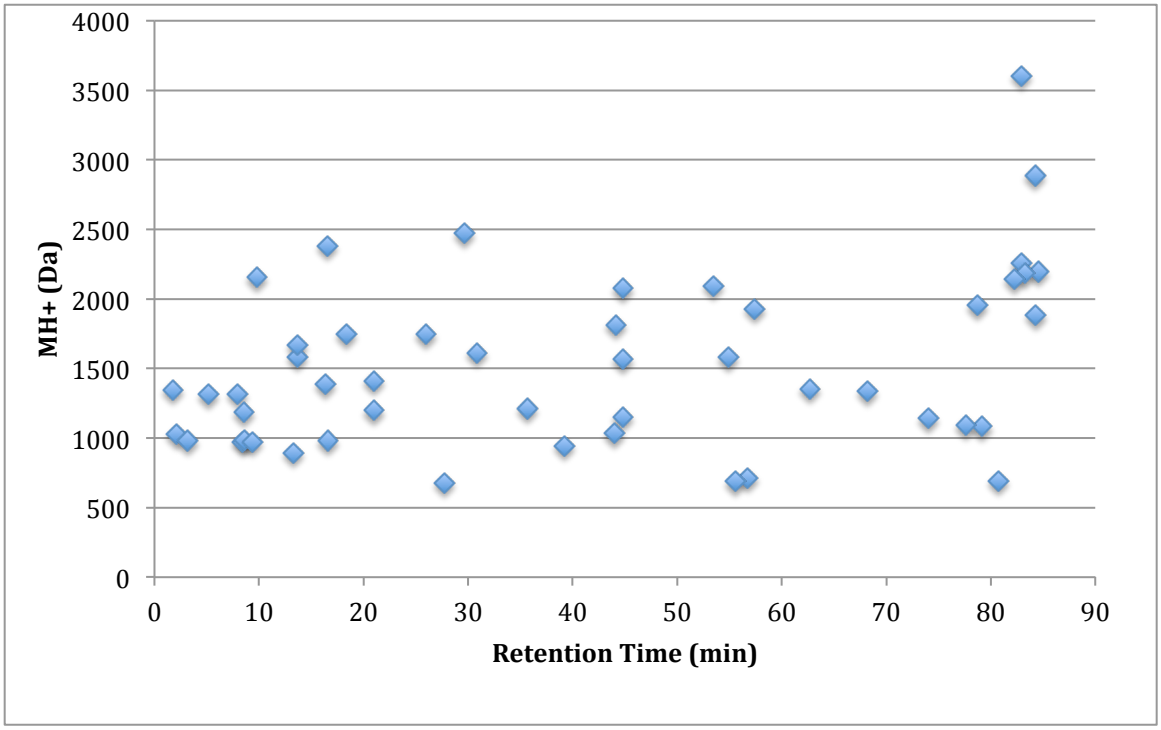

Figure 3.23 Retention Time (min) vs MH+ (Da) for ACE 3 C18 column with 0.8 mL/min flow rate (90 minute gradient)

Again, from this plot overall the peptides appear to be eluting at a fairly even distribution throughout the 90 minute run. The average mass of the peptides is below 2000, but there are some higher mass peptides that are not noted in the previous run. With this run, the peptide at 4936 Da is no longer seen.

Following these two initial runs that were somewhat unsuccessful, the older column was replaced with a new superC18 reverse phase HPLC column purchased from ACE. The next run was again a 10 microliter injection at flow rate of 0.8 mL/ml which yielded chromatogram:

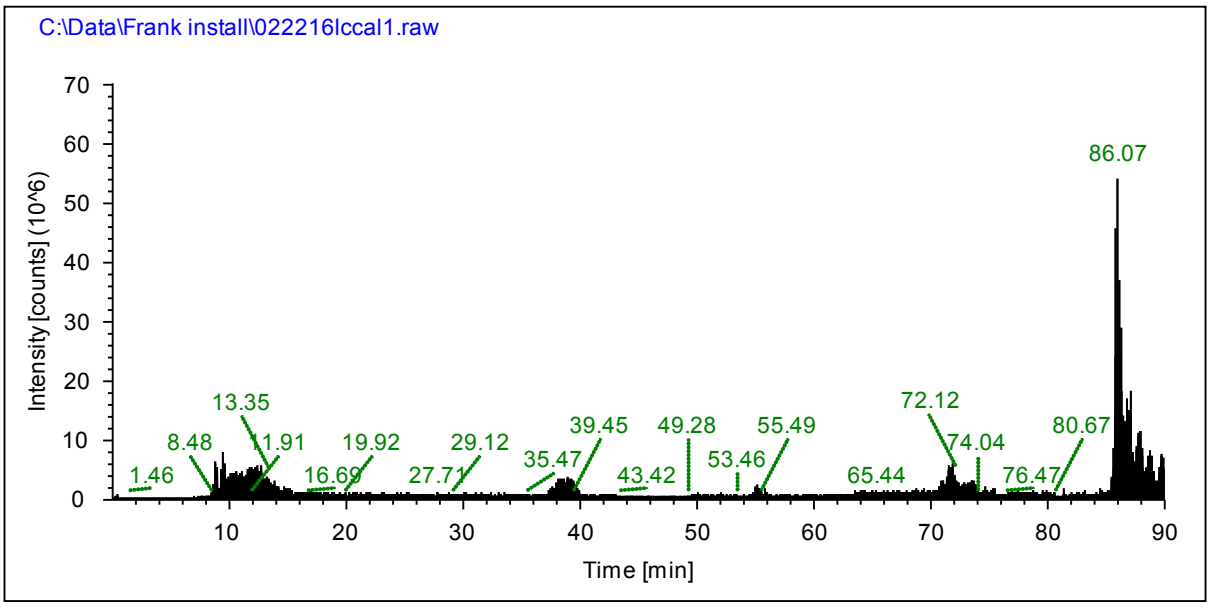

Figure 3.24 HPLC chromatogram using superC18 column and 0.8 mL/min flow rate

This time, SEQUEST was able to identify 3 proteins: carbonic anhydrase 2, NAD(P)H dehydrogenase, and C-reactive protein (Chain 19-224) with coverage percentages of 63.32%, 59.34%, and 44.23% respectively. In total SEQUEST found 215 peptide fragments and it was able to match 86 peptides to a protein. The plot of retention time versus parent mass for peptides matched to proteins for this run is shown below.

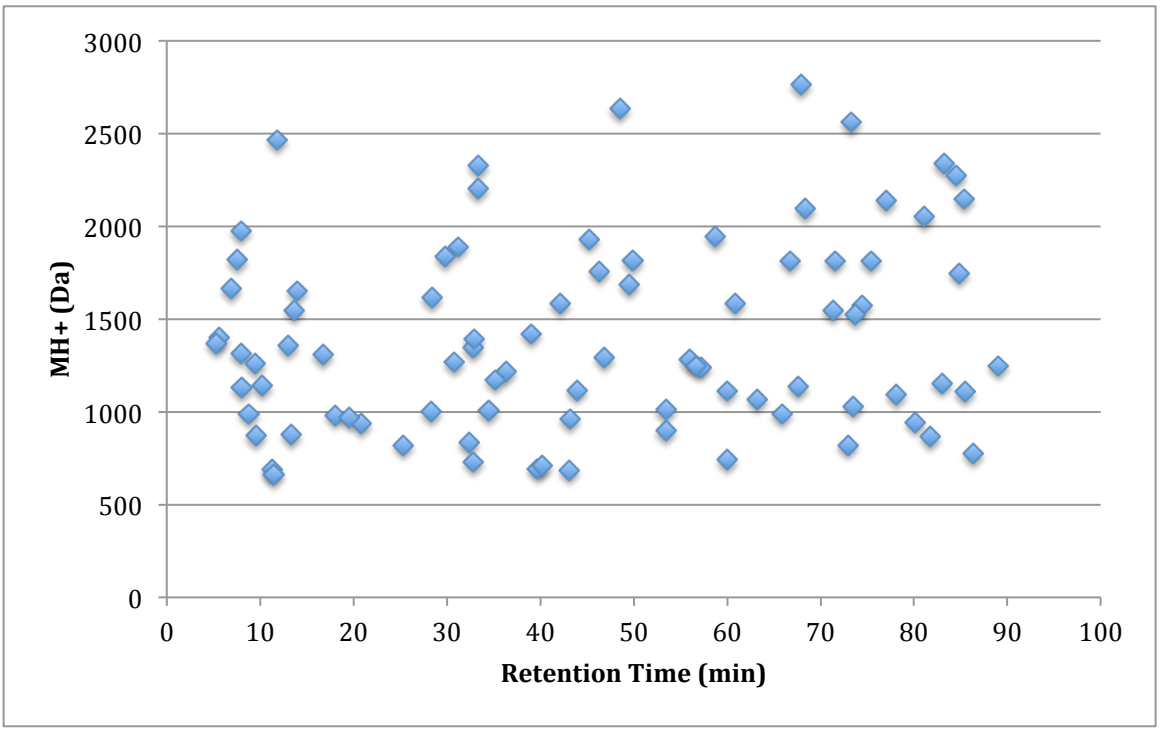

Figure 3.25 Retention Time (min) vs MH+ (Da) for superC18 column with 0.8 mL/min flow rate (90 minute gradient)

From this plot, it is graphically very clear that the total number of peptides matched to a protein has increased drastically. This time, the distribution of masses seems to be more spread out from 2500 Da to 500 Da, and is not isolated to the 2000 Da to 1000 Da range. The overall distribution in time that peptides are eluting remains consistently even throughout the time gradient.

Next, the sample was run using both CID and ETD as activation to induce fragmentation. For this run, SEQUEST was run using CID as the activation source but with calculations for b, y, c, and z fragments, with all other parameters the same. The chromatogram for this run is as shown below:

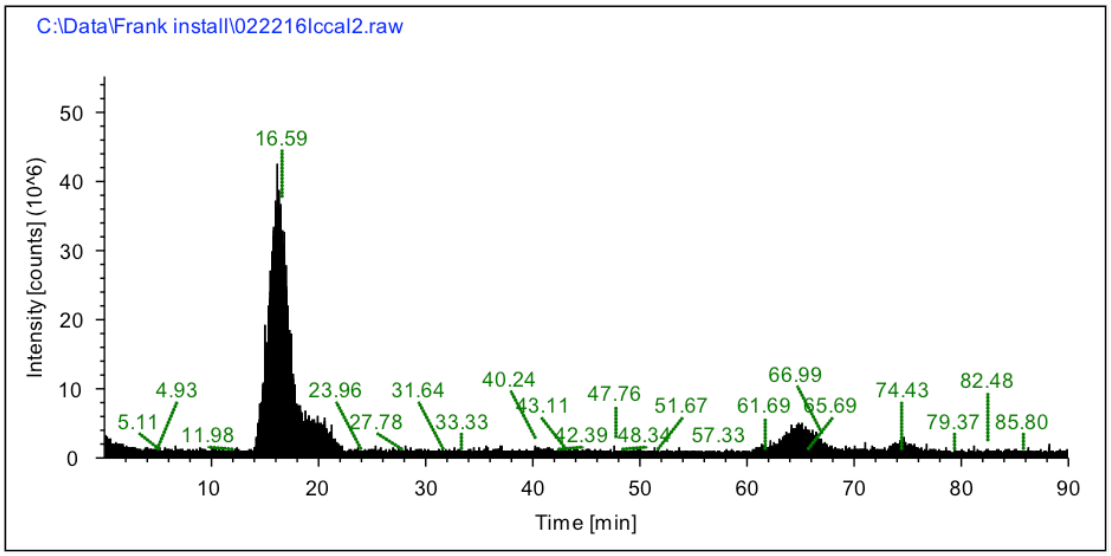

Figure 3.26 HPLC chromatogram using CID and ETD

SEQUEST only identified carbonic anhydrase 1 and 2 with coverage of 56.15% and 70.27% respectively. In total it found 239 peptides and matched 57 peptides to a protein. As the usage of ETD in conjunction with CID did not drastically improve the results, the data collection program was changed back to CID only, as the ETD device on the instrument takes additional time to set up/use and did not yield significantly improved results. The plot of retention time versus parent mass is shown below.

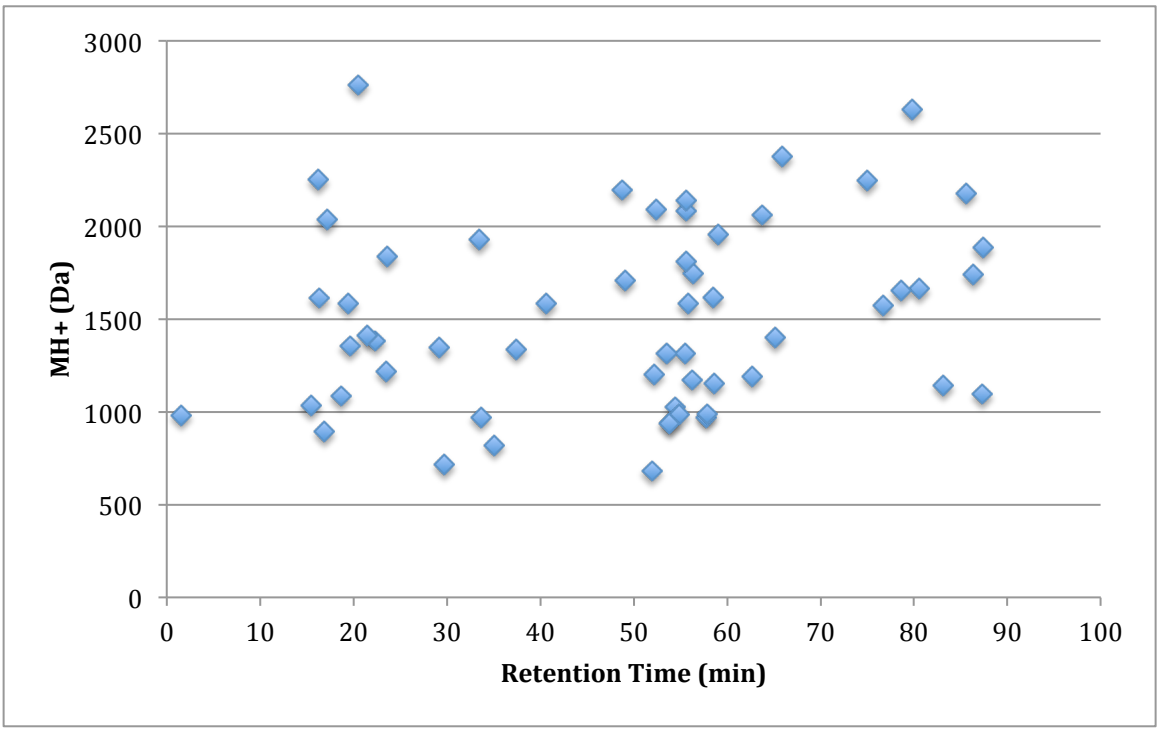

Figure 3.27 Retention time (min) vs MH+ (Da) using CID and ETD

In comparison to the previous run, this plot shows a marked decrease in peptides identified. There appears to be a slight concentration in peptides eluting from 50 to 60 minutes. There is also a gap between about 0 to 15 minutes where only 1 peptide was found.

At this time an attempt was made to replace the ESI source with a nano electrospray source, however, the flow rate of the HPLC that was required was much too high to interface properly with the nanospray source. After several unsuccessful attempts to decrease the HPLC flow rate to a low enough value to still work with the nanospray, the source was switched back to an electrospray.

Given the lack of success obtaining successful data collection using the gradient previously developed by members of this lab, it was decided to start with a new gradient based on gradients from the literature. This 63.3 minute gradient is as follows:

7 min – 0% Pump B
50 min – 0%-50% Pump B
53 min – 100% Pump B
58 min – 100% Pump B
58.1 min – 0% Pump B
63 min – 0% Pump B

A graphical representation of this gradient curve is shown below.

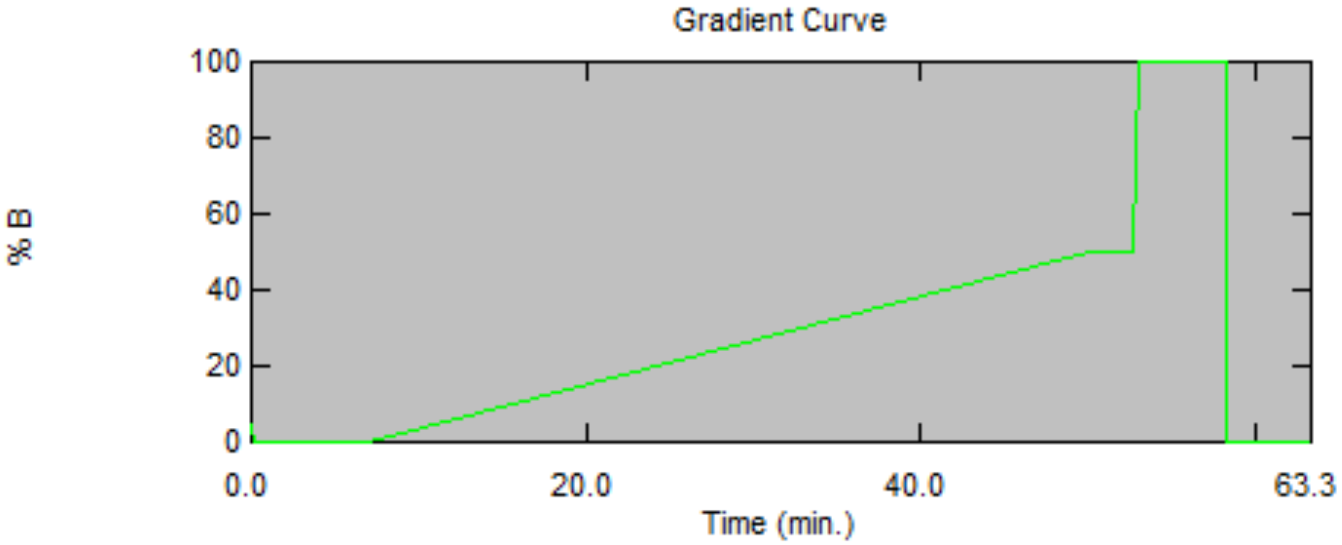

Figure 3.28 HPLC 63.3 minute gradient

The sample was then run using this new, shorter gradient with an injection size of 10 microliters and flow rate of 0.2 mL/min. The chromatogram for this run is as follows.

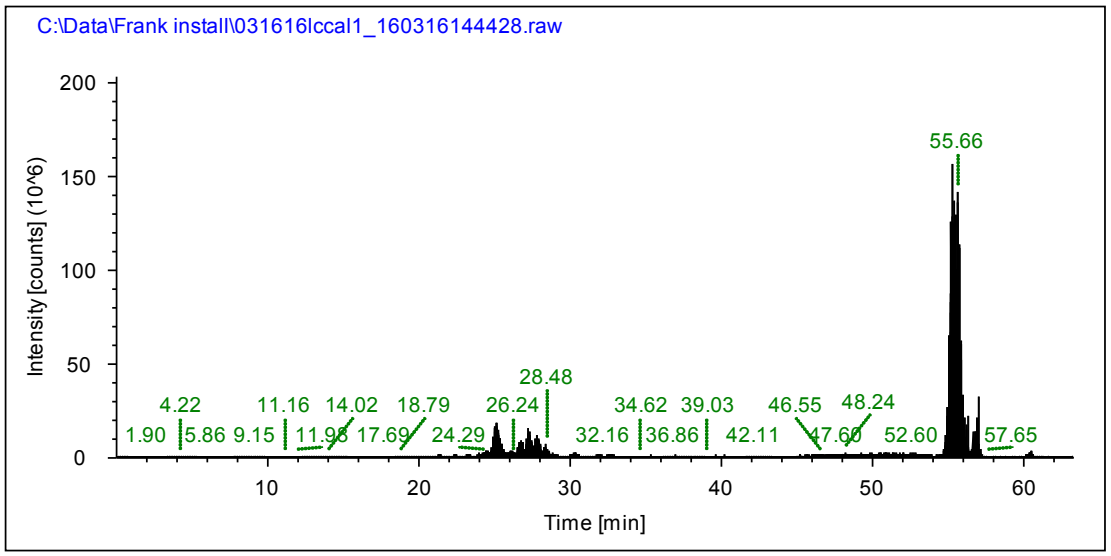

Figure 3.29 HPLC chromatogram from 63.3 minute gradient and 10 microliter injection

This time, SEQUEST was able to identify all six proteins known to be in the sample. The table below summarizes the data for these proteins. In total, SEQUEST found 191 peptides, all of which were matched to a protein. In addition, SEQUEST did identify all stabile isotope labeled peptides found in table 2.1.

Protein	**Coverage**	**# Peptides**	**MW (kDA)**
Carbonic anhydrase 1	65.38%	26	28.7
Carbonic anhydrase 2	55.98%	26	29.1
NAD(P)H dehydrogenase	53.48%	32	30.7
C-reactive protein (chain 19-224)	44.23%	14	23.2
Peptidyl-prolyl cis-trans isomerase A	64.32%	29	8.21
Catalase (chain 5-227)	66.73%	64	7.38

Table 3.1 Coverage values for the 6 proteins identified using 63.5 minute gradient.

The plot relating parent mass to retention time of the peptides matched to proteins is shown below.

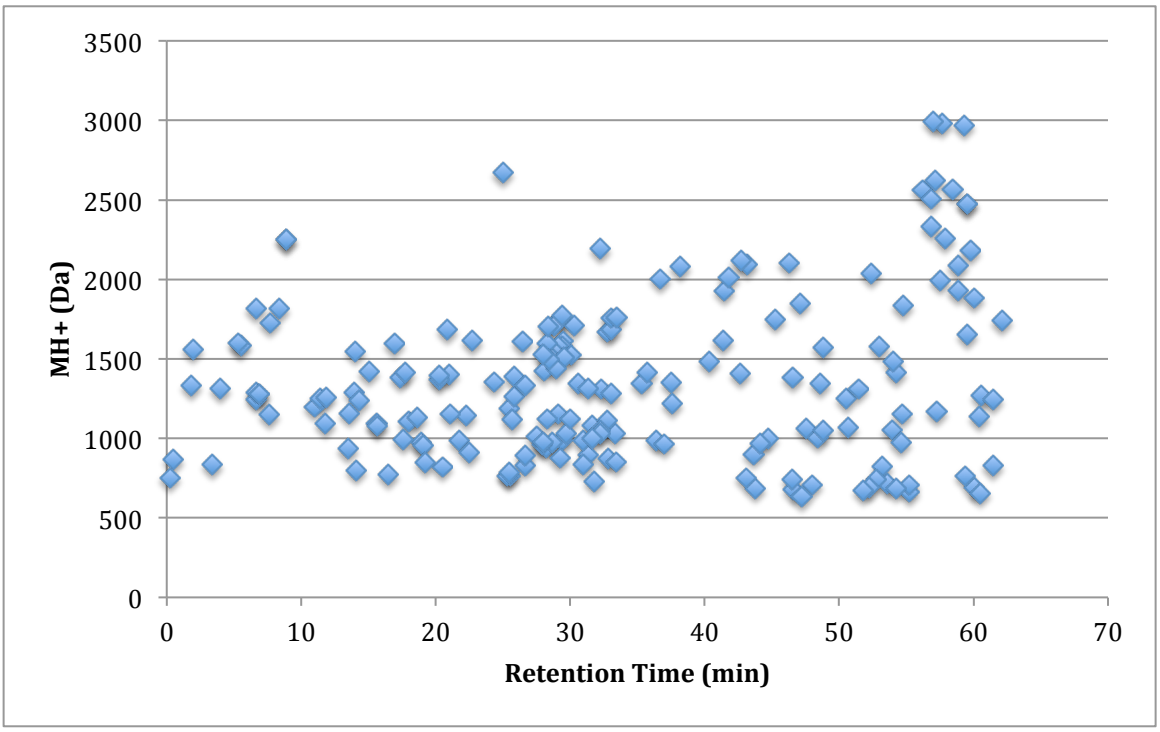

Figure 3.30 Retention Time (min) vs MH+ (Da) using 63.3 minute gradient and 10 microliter injection

Here there is again a significant increase in total peptides identified and matched to a protein. Though peptides are eluting from this gradient throughout the 60 minute run, there is a concentration of peptides that appear around 30 minutes and around 60 minutes. The range of masses of the peptides has also increased, and now seems to average about 3000 Da to 500 Da.

Following the success of this run, solvent was allowed to run through the column and multiple blanks were run for the full 63 minute gradient until SEQUEST analyses run on the blanks were no longer identifying any proteins. This took a total of 3 blank 63-minute runs, with the pumps allowed to run in addition to these runs for about 2 hours.

The final run completed was a 2 microliter injection with flow rate of 0.2 mL/min, using the successful 63-minute gradient above. The chromatogram for this run is shown below.

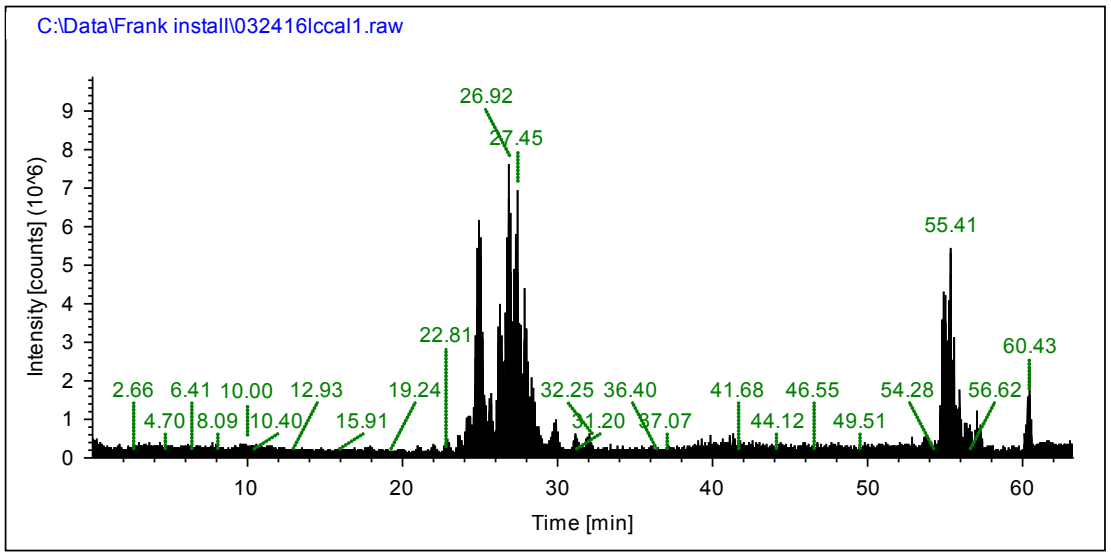

Figure 3.31 HPLC chromatogram using 63.3 minute gradient and 2 microliter injection

From this run, SEQUEST only identified 3 proteins, carbonic anhydrase 2, NAD(P)H dehydrogenase, Peptidyl-prolyl cis-trans isomerase A with coverage values of 62.93%, 49.08%, and 64.32%, with a total of 188 peptides identified and 95 peptides matched to a protein. This means that while a smaller injection size may be sufficient, 2 microliters is too small, and further research must be done to optimize the size of the injection to proteins identified and their coverage. The plot of retention time versus parent mass is shown below.

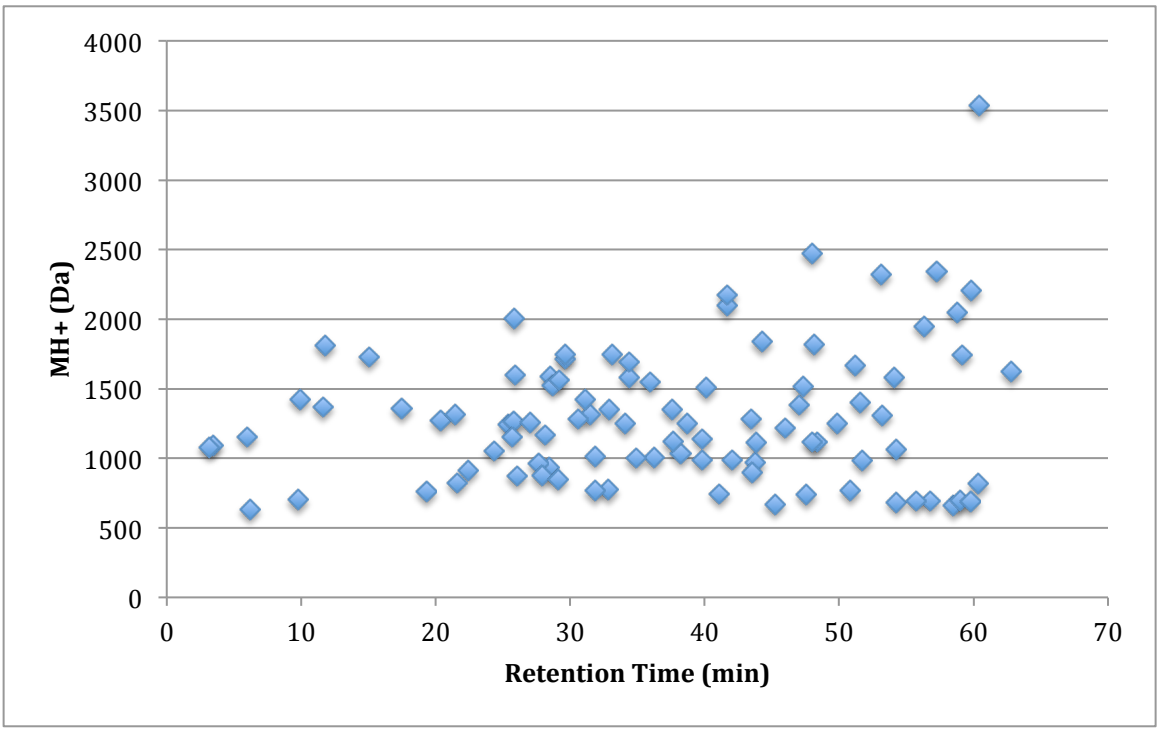

Figure 3.32 Retention Time (min) vs MH+ (Da) using 63.5 minute gradient and 2 microliter injection

This plot appears similar to the previous run, though has a marked decrease in data points. This shows that the size of the injection was insufficient to obtain enough peptides to fully identify all six proteins in the sample.

3.3 Conclusions and Future Work

The fragmentation patterns of both singly- and doubly-charged APipAAR, APipAAK, APAAR, APAAK, AAPipAR, AAPipAK, AAPAR, and AAPAK were studied. Several interesting results are noted by comparing these spectra with each other as well as the anticipated results based on the proline and pipecolic acid effects.

First, overall when placed 1 residues away from either an arginine or lysine residue, the pipecolic acid effect seems to have less of an effect. This pattern holds for all of the peptides analyzed containing a pipecolic acid residue. This would suggest that the mechanism by which

the pipecolic acid effect occurs must be being affected by the close proximity of the lysine/arginine residue.

Next, the position of the pipecolic acid in relation to the arginine residue affects the production of doubly charged product ions. In the spectrum for AAPipAR, there were doubly charged product ions whereas in APipAAR there were not. This means that changing the position of the pipecolic acid likely has an impact on the proton affinity of the arginine residue.

Doubly charged product ions were noted for AAPAR, APAAK, APAAR, but not for AAPAK. Similarly to the above result with pipecolic acid, this means that changing the position of the proline residue in relation to lysine has an impact on whether or not the second proton remains on the lysine residue.

The proline effect was consistent in all singly charged peptides except APAAK, APAAR, and AAPAR, though this result is not strong because the main product ion for both APAAK and APAARK was neutral loss of water. In the doubly charged species of APAAK, APAAR, AAPAR, and AAPAK the proline effect is noted. This shows that the doubly charged species does not necessarily hinder the proline effect.

Finally, peptides containing pipecolic acid generally were less likely to produce doubly charged product ions as compared to peptides containing proline. Three out of four doubly charged proline peptides produced doubly charged products whereas only one out of four pipecolic acid peptides produced doubly charged products.

Future studies in peptide fragmentation of doubly charged proline/pipecolic acid containing species should be done on hexapeptides with further variation of the location of the pipecolic acid or proline residue in relation to lysine and arginine, to see if the location based patterns noted above still apply with increasing length of the peptide. In addition to this, studies

should be performed on the XAAAR and XAAAK (X= proline, pipecolic acid) in order to see if a separation of three alanine residues still allows for doubly charged product ions in the proline/lysine containing peptide and if the lack of doubly charged product ion is noted in the pipecolic acid/arginine peptide.

In addition, the mass spectrometry proteomics based experiments provided an optimized methodology for performing future research on the labs' newly acquired HPLC/LTQ instrument. A gradient of acetonitrile/H_2O that increased with respect to concentration of acetonitrile was more effective when the gradient increased in a more linear fashion. Increasing flow rate of the HPLC from 0.2 mL/min to 0.8 mL/min initially improved results, however the final gradient was used successfully at a flow rate of 0.2 mL/min. In this initial experiment, usage of ETD in conjunction with CID did not notably improve the results.

Future work in proteomics will use this optimized technique to study bacterial *E. coli* cells infected with various bacteriophages. In addition, the HPLC/LTQ may be modified using a splitter to potentially allow for usage of the nano-electrospray source, to see if that improves protein identification results from the Qual/Quant standard. Further research must be performed in order to determine optimal initial sample injection.

References

[1] Jensen, L.J., Gupta, R., Blom, N., Devos, D., Tamames, J., Kesmir, C., Nielsen, H., Stærfeldt, H.H., Rapacki, K., Workman, C. and Andersen, C.A.F., 2002. Prediction of human protein function from post-translational modifications and localization features. *Journal of molecular biology*, *319*(5), pp.1257-1265.

[2] Richardson, J.S., 1981. The anatomy and taxonomy of protein structure. *Advances in protein chemistry*, *34*, pp.167-339.

[3] Raulfs, M.D.M., Breci, L., Bernier, M., Hamdy, O.M., Janiga, A., Wysocki, V. and Poutsma, J.C., 2014. Investigations of the mechanism of the "proline effect" in tandem mass spectrometry experiments: The "pipecolic acid effect". *Journal of The American Society for Mass Spectrometry*, *25*(10), pp.1705-1715.

[4] PDB DOI: 10.2210/pdb4fqf/pdb. Lang, B.S., Gorren, A.C., Oberdorfer, G., Wenzl, M.V., Furdui, C.M., Poole, L.B., Mayer, B. and Gruber, K., 2012. Vascular bioactivation of nitroglycerin by aldehyde dehydrogenase-2 reaction intermediates revealed by crystallography and mass spectrometry. *Journal of Biological Chemistry*, *287*(45), pp.38124-38134.

[5] Ng, P.C. and Henikoff, S., 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*, *31*(13), pp.3812-3814.

[6] Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M. and Obradovic, Z., 2002. Intrinsic disorder and protein function. *Biochemistry*, *41*(21), pp.6573-6582.

[7] Stuart, M.J. and Nagel, R.L., 2004. Sickle-cell disease. *The Lancet*, *364*(9442), pp.1343-1360.

[8] Feist, P. and Hummon, A.B., 2015. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *International journal of molecular sciences*, *16*(2), pp.3537-3563.

[9] Rozek, W. and Ciborowski, P.S., 2008. Proteomics and Genomics. *Neuroimmune Pharmacology* (pp. 725-741). Springer US.

[10] Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.C. and Yates III, J.R., 2013. Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews*, *113*(4), pp.2343-2394.

[11] Cellulaire, B., 2002. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics*, *2*, pp.3-10.

[12] Carter, P. and Wells, J.A., 1988. Dissecting the catalytic triad of a serine protease. *Nature*, *332*(6164), pp.564-568.

[13] Nold, M.J., Wesdemiotis, C., Yalcin, T. and Harrison, A.G., 1997. Amide bond dissociation in protonated peptides. Structures of the N-terminal ionic and neutral fragments. *International Journal of Mass Spectrometry and Ion Processes*, *164*(1), pp.137-153.

[14] Hustoft, H.K., Malerod, H., Wilson, S.R., Reubsaet, L., Lundanes, E. and Greibrokk, T., 2012. A critical review of trypsin digestion for LC-MS based proteomics. *Integrative Proteomics*, *1*, pp.73-82.

[15] Parks, B.A., Jiang, L., Thomas, P.M., Wenger, C.D., Roth, M.J., Boyne, M.T., Burke, P.V., Kwast, K.E. and Kelleher, N.L., 2007. Top-down proteomics on a chromatographic time scale using linear ion trap Fourier transform hybrid mass spectrometers. *Analytical chemistry*, *79*(21), pp.7984-7991.

[16] Armirotti, A. and Damonte, G., 2010. Achievements and perspectives of top-down proteomics. *Proteomics*, *10*(20), pp.3566-3576.
[17] Lam, H., Deutsch, E.W., Eddes, J.S., Eng, J.K., King, N., Stein, S.E. and Aebersold, R., 2007. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics*, *7*(5), pp.655-667.
[18] Veenstra, M.Z.T.D., Top-down Proteomics.
[19] Nesvizhskii, A.I., 2010. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of proteomics*, *73*(11), pp.2092-2123.
[20] Weiller, G.F., Djordjevic, M.J., Caraux, G., Chen, H. and Weinman, J.J., 2001. A specialised proteomic database for comparing matrix-assisted laser desorption/ionization-time of flight mass spectrometry data of tryptic peptides with corresponding sequence database segments. *Proteomics*, *1*(12), pp.1489-1494.
[21] Huang, Y., Wysocki, V.H., Tabb, D.L. and Yates, J.R., 2002. The influence of histidine on cleavage C-terminal to acidic residues in doubly protonated tryptic peptides. *International Journal of Mass Spectrometry*, *219*(1), pp.233-244.
[22] Bleiholder, C., Suhai, S., Harrison, A.G. and Paizs, B., 2011. Towards understanding the tandem mass spectra of protonated oligopeptides. 2: The proline effect in collision-induced dissociation of protonated Ala-Ala-Xxx-Pro-Ala (Xxx= Ala, Ser, Leu, Val, Phe, and Trp). *Journal of the American Society for Mass Spectrometry*, *22*(6), pp.1032-1039.
[23] Breci, L.A., Tabb, D.L., Yates, J.R. and Wysocki, V.H., 2003. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Analytical chemistry*, *75*(9), pp.1963-1971.
[24] Qian, W.J., Liu, T., Monroe, M.E., Strittmatter, E.F., Jacobs, J.M., Kangas, L.J., Petritis, K., Camp, D.G. and Smith, R.D., 2005. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. *Journal of proteome research*, *4*(1), pp.53-62.
[25] MacCoss, M.J., Wu, C.C. and Yates, J.R., 2002. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Analytical chemistry*, *74*(21), pp.5593-5599.
[26] Roepstorff, P.; Kohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed Mass Spectrom*, 1984, 11, 601.
[27] Wysocki, V.H., Tsaprailis, G., Smith, L.L. and Breci, L.A., 2000. Mobile and localized protons: a framework for understanding peptide dissociation. *Journal of Mass Spectrometry*, *35*(12), pp.1399-1406.
[28] Boyd, R. and Somogyi, Á., 2010. The mobile proton hypothesis in fragmentation of protonated peptides: a perspective. *Journal of the American Society for Mass Spectrometry*, *21*(8), pp.1275-1278.
[29] Roepstorff, P.; Kohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed Mass Spectrom*, 1984, 11, 601.
[30] Paizs, B. and Suhai, S., 2005. Fragmentation pathways of protonated peptides. *Mass spectrometry reviews*, *24*(4), pp.508-548.
[31] Yalcin, T., Khouw, C., Csizmadia, I.G., Peterson, M.R. and Harrison, A.G., 1995. Why are B ions stable species in peptide spectra?. *Journal of the American Society for Mass Spectrometry*, *6*(12), pp.1165-1174.

[32] Loo, J.A., Edmonds, C.G. and Smith, R.D., 1993. Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization. *Analytical chemistry*, *65*(4), pp.425-438.
[33] Rodgers, K.J., Samardzic, K. and Main, B.J., 2015. Toxic Nonprotein Amino Acids.
[34] Merrifield, R. B.; Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* 1963, 85, 2149-2154.
[35] Hood, C.A., Fuentes, G., Patel, H., Page, K., Menakuru, M. and Park, J.H., 2008. Fast conventional Fmoc solid-phase peptide synthesis with HCTU. *Journal of Peptide Science*, *14*(1), pp.97-101.
[36] Chan, W. C.; White, P. D.; Fmoc Solid Phase Peptide Synthesis: A Practical Approach. New York: Oxford University Press Inc.; 2000.
[37] Wu Z, Fenselau C. Proton affinity of arginine measured by the kinetic approach. Rapid communications in mass spectrometry. 1992 Jun 1;6(6):403-5.
[38] Wu, Z., Fenselau, C. and Graham Cooks, R., 1994. Gas-phase basicities and proton affinities of lysine and histidine measured from the dissociation of proton-bound dimers. *Rapid communications in mass spectrometry*, *8*(9), pp.777-780.