

5-2016

## Fragmentation Studies of Lysine and Lysine Analog Containing Tetrapeptides

Zachariah Imran Hasan  
*College of William and Mary*

Follow this and additional works at: <https://scholarworks.wm.edu/honorsthesis>



Part of the [Analytical Chemistry Commons](#), [Biochemistry, Biophysics, and Structural Biology Commons](#), and the [Bioinformatics Commons](#)

---

### Recommended Citation

Hasan, Zachariah Imran, "Fragmentation Studies of Lysine and Lysine Analog Containing Tetrapeptides" (2016). *Undergraduate Honors Theses*. Paper 926.  
<https://scholarworks.wm.edu/honorsthesis/926>

This Honors Thesis is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of W&M ScholarWorks. For more information, please contact [scholarworks@wm.edu](mailto:scholarworks@wm.edu).

Fragmentation Studies of Lysine and Lysine Analog Containing Tetrapeptides

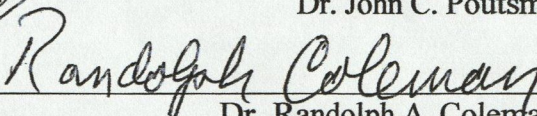
A thesis submitted in partial fulfillment of the requirement for the degree of  
Bachelor of Science in Chemistry from  
The College of William and Mary

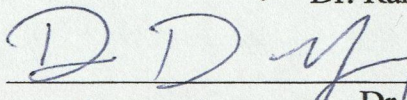
by

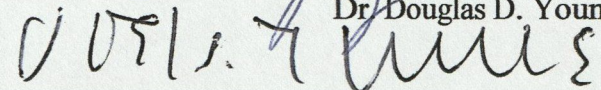
Zachariah Imran Hasan

Accepted for Honors

  
Chair Dr. John C. Poutsma

  
Dr. Randolph A. Coleman

  
Dr. Douglas D. Young

  
Dr. Joel S. Levine

Williamsburg, VA  
May 3, 2016

# Table of Contents

<b>List of Figures</b> .....	iii
<b>Acknowledgements</b> .....	v
<b>Abstract</b> .....	vi
<b>Chapter 1: Introduction</b> .....	1
1.1 MS-based Proteomics .....	1
1.2 Peptide Fragmentation .....	12
1.2.1 Fragmentation Mechanisms .....	12
1.2.2 Variations of b ions .....	14
1.2.3 Selective Cleavages .....	18
1.2.4 Macrocyclization.....	24
1.3 Tetrapeptide Fragmentation Studies .....	26
1.3.1 Lysine, Ornithine, DABA, DAPA .....	26
1.3.2 Tetrapeptides of Interest .....	27
<b>Chapter 2: Experimental Methods</b> .....	29
2.1 Peptide Synthesis .....	29
2.1.1 Example Synthesis of AAKA .....	31
2.2 ESI-Ion Trap Mass Spectrometry .....	35
2.2.1 Sample Preparation .....	35
<b>Chapter 3: Results and Discussion</b> .....	36
3.1 Alanine-Alanine-Alanine-X.....	36
3.1.1 AAAK .....	36
3.1.2 AAAO .....	39
3.1.3 AAAB .....	41
3.1.4 AAAZ .....	42
3.2 Alanine-Alanine-X-Alanine.....	44
3.2.1 AAKA .....	44
3.2.2 AAOA .....	46
3.2.3 AABA .....	47
3.2.4 AAZA .....	48

3.3 Alanine-X-Alanine-Alanine.....	50
3.3.1 AKAA .....	50
3.3.2 AOAA .....	51
3.3.3 ABAA .....	53
3.3.4 AZAA .....	54
3.4 X-Alanine-Alanine-Alanine.....	55
3.4.1 KAAA .....	55
3.4.2 OAAA .....	57
3.4.3 BAAA .....	58
3.4.4 ZAAA .....	59
3.5 Tyrosine-Alanine-Glycine-X .....	60
3.5.1 YAGK .....	61
3.5.2 YAGO .....	63
3.5.3 YAGB .....	64
3.5.4 YAGZ .....	65
3.6 Trends .....	67
<b>Chapter 4: Conclusions .....</b>	<b>69</b>
<b>References .....</b>	<b>71</b>
<b>Appendix .....</b>	<b>76</b>

## List of Figures

Figure 1.1 Gel vs. LC-MS based Proteomics .....	2
Figure 1.2 Bottom-up, Middle-down, Top-Down Proteomics .....	4
Figure 1.3 LC-ESI Apparatus .....	8
Figure 1.4 ESI Schematic .....	9
Figure 1.5 MS/MS CID Spectrum .....	10
Figure 1.6 Peptide Fragmentation Sites .....	12
Figure 1.7 CID vs. ECD or ETD.....	13
Figure 1.8 Mobile Proton Mechanism .....	14
Figure 1.9 $b_x$ , $y_z$ Ion Pathway .....	16
Figure 1.10 $b_n$ Ion Variations.....	17
Figure 1.11 Proline Effect.....	19
Figure 1.12 Histidine Effect.....	20
Figure 1.13 Glutamine Cleavage .....	21
Figure 1.14 Oxidized Cysteine Cleavage.....	21
Figure 1.15 Lysine and Arginine Effect .....	22
Figure 1.16 Ornithine Effect .....	23
Figure 1.17 YAGFL-NH <sub>2</sub> Macrocyclization .....	25
Figure 1.18 Lysine, Ornithine, DABA, DAPA.....	27
Figure 1.19 Tetrapeptide Macrocycle .....	28
Figure 2.1 Fmoc-Lys(Boc)-Wang Resin.....	29
Figure 2.2 Synthesis of AAKA.....	31
Figure 2.3 AAAX Tetrapeptides.....	33
Figure 2.4 AAXA Tetrapeptides.....	33
Figure 2.5 AXAA Tetrapeptides.....	34

Figure 2.6 XAAA Tetrapeptides.....	34
Figure 2.7 YAGX Tetrapeptides.....	34
Figure 2.8 ESI-Ion trap Schematic.....	35
Figure 3.1 CID Spectrum of AAAK.....	36
Table 3.1 CID Activation Data AAAX, AAXA, AXAA, XAAA.....	38
Figure 3.2 CID Spectrum of AAAO.....	39
Figure 3.3 CID Spectrum of AAAB.....	41
Figure 3.4 CID Spectrum of AAAZ.....	42
Figure 3.5 CID Spectrum of AAKA.....	44
Figure 3.6 CID Spectrum of AAOA.....	46
Figure 3.7 CID Spectrum of AABA.....	47
Figure 3.8 CID Spectrum of AAZA.....	48
Figure 3.9 CID Spectrum of AKAA.....	50
Figure 3.10 CID Spectrum of AOAA.....	51
Figure 3.11 CID Spectrum of ABAA.....	53
Figure 3.12 CID Spectrum of AZAA.....	54
Figure 3.13 CID Spectrum of KAAA.....	55
Figure 3.14 CID Spectrum of OAAA.....	57
Figure 3.15 CID Spectrum of BAAA.....	58
Figure 3.16 CID Spectrum of ZAAA.....	59
Figure 3.17 CID Spectrum of YAGK.....	61
Table 3.2 CID Activation Data YAGX.....	62
Figure 3.18 CID Spectrum of YAGO.....	63
Figure 3.19 CID Spectrum of YAGB.....	64
Figure 3.20 CID Spectrum of YAGZ.....	65

## **Acknowledgements**

I would like to thank Dr. John C. Poutsma for the opportunity to do research on such an interesting topic, his guidance throughout my honors project, and his mentorship in and out of the lab. I would like to thank my fellow Ionlab members, especially Hannah Smith, Anton Lachowicz, Matthew Wang, and Ethan Jones for their help when needed and for making my undergraduate research experience memorable. I would like to thank the W&M Charles Center for Honors Fellowship financial support to conduct summer research. Finally, special thanks to my friends and family, who have supported me throughout this process.

## Abstract

The fragmentation patterns of lysine and lysine-analog containing tetrapeptides were analyzed in this study using collision induced dissociation (CID) in an ESI-ion trap mass spectrometer. Understanding the fragmentation mechanisms of lysine-containing peptides is integral to improving bottom-up proteomics techniques and peptide sequencing and searching algorithms. Lysine and its non-protein amino acid (NPAA) analogs ornithine, DABA, and DAPA have been shown to affect fragmentation patterns based on their basicities in dipeptides and tripeptides. Studies have shown the occurrence of sequence scrambling due to macrocyclization of pentapeptides during fragmentation, which can result in inaccurate database matching. This study of the twenty tetrapeptides AAAX, AAXA, AXAA, XAAA, and YAGX (X = Lys, Orn, DABA, or DAPA) looked for macrocyclization leading to sequence scrambling and analyzed the effects of positional variance and of differing basicities between lysine and its analogs on tetrapeptide fragmentation patterns.

Fragmentation studies confirmed the occurrence of the ornithine effect, where there is selective cleavage C-terminal to an ornithine residue within the tetrapeptides. Macrocylic sequence scrambling was found to not occur in significant amounts for these tetrapeptides. The formation of  $b_n + H_2O$  ions was found to be most prevalent when the basic amino acid residues were at the C-terminus of the tetrapeptides. Positional variance and basicities of the lysine and its NPAA analogs affected the stabilities of the tetrapeptides, and influenced fragmentation patterns. Further investigations of lysine-containing peptides are necessary to better understand the fragmentation mechanisms at work and improve the robustness of proteomics experiments using mass spectrometry.



## **Chapter 1: Introduction**

Advances in molecular biology and genetics research leading to the sequencing of the human genome have sparked new investigations into understanding the proteomes of humans and other organisms [1-4]. Mass spectrometry has become a widely used technique in proteomics research for sequencing peptides and proteins [1-51]. Specifically, fragmentation of peptides in tandem mass spectrometry allows for mass spectra of peptide fragments to be obtained [1-6, 10-51]. Ultimately, these peptide mass spectra can be matched to specific proteins in protein structural databases, creating a powerful tool for future biochemistry and molecular biology research [1-5, 20-27]. Much of current research in mass spectrometric peptide sequencing focuses on further understanding of the underlying mechanisms of the peptide fragmentation process [1-4, 30-51]. For example, in the peptide fragmentation process, selective cleavages based on the chemical properties of amino acid residues in the peptide can occur [2, 30-47]. Further understanding of these mechanisms can lead to enhanced accuracy of peptide identification, and therefore a more robust proteomics experiment.

### **1.1 MS-based Proteomics**

Mass spectrometry based proteomics research utilizes a combination of biochemistry and analytical chemistry laboratory techniques. The first step in a proteomics experiment is usually to isolate the protein of interest from its source. Proper isolation of the protein at a high level of purity is vital to ensuring that the subsequent mass spectrometry experiments are as accurate as possible [6]. The protein purification process relies on a variety of laboratory techniques including subcellular fractionation, used to break open cells and organelles, and centrifugation, used to further separate fractions based on size and density [6]. Specifically, multiple steps of

differential centrifugation and density gradient centrifugation techniques are used to obtain fractions containing purified proteins [6]. Proteins within a fraction can be separated from one another using gel-electrophoresis techniques (most commonly 2D-PAGE) [3-4, 7]. Subsequently, the portion of the gel containing the protein of interest can be removed for further analysis [4, 7-8]. Alternatively, the proteins can be digested and then separated using liquid chromatography (LC) techniques such as LC, high performance liquid chromatography (HPLC) or nanoHPLC [4]. Finally, the proteins or digests are mass analyzed and fragmented via tandem mass spectrometry (MS/MS) and then identified or sequenced via protein databases [1-5, 20-27]

Figure 1.1 below adapted from Roepstorff, 2012 summarizes these two proteomics strategies.

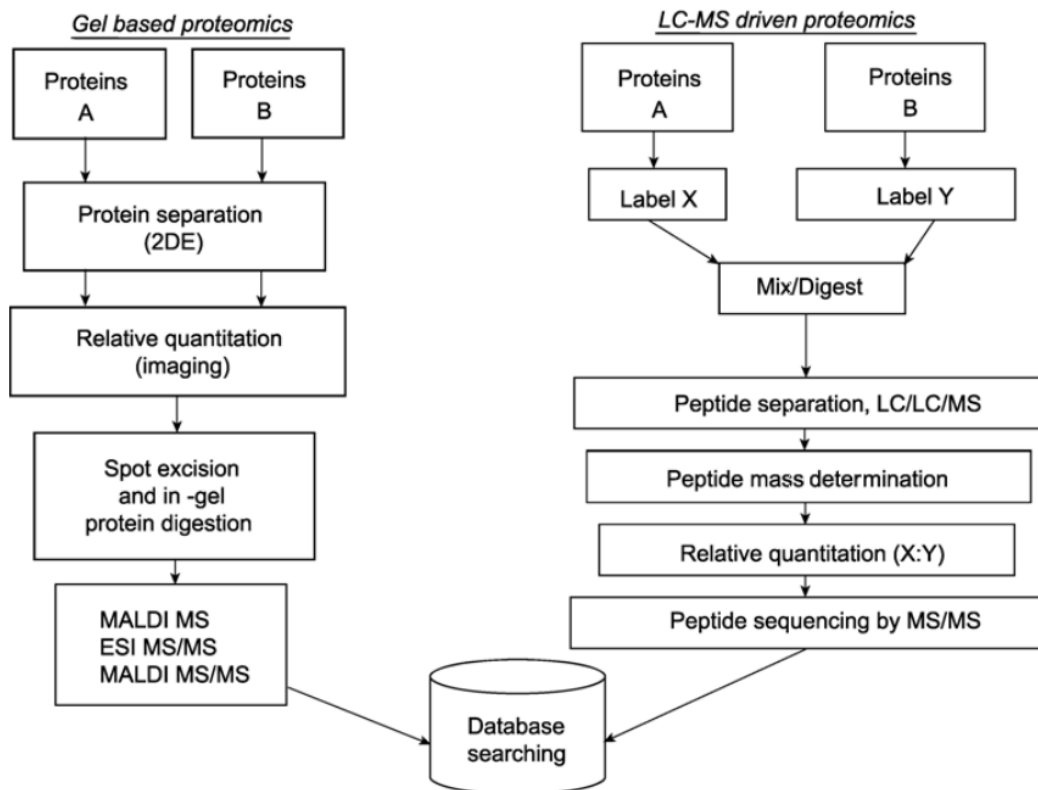


Figure 1.1 Summary of gel based proteomics and liquid chromatography coupled proteomics.

Adapted from [4].

MS-based proteomics research can be divided into profiling, functional, and structural proteomics [9-10]. Profiling proteomics aims to profile the proteins present in different samples and analyze their similarities and differences [9]. This can be used for a wide variety of applications including determining unknown organisms, cells, protein fractions, etc. by comparing to known samples, differentiating wild-type vs. mutant proteomes, differentiating between healthy and diseased states of cells, quantizing metabolites, etc. [9, 11-13]. Functional proteomics examines post-translational modifications, biological protein functions and interactions with other molecules [9-10]. Structural proteomics aims to analyze and identify all levels of protein structure (primary sequence, secondary structural features, native protein folding, protein complexes and subunits) [9].

There are three major strategies of MS-based proteomics experiments: top-down, middle-down and bottom-up proteomics (see Figure 1.2) [3, 5]. In these proteomics experiments, identification, molecular characterization, and sequence determination of proteins can be assessed based on their fragment ion mass spectra [1-6, 10-51]. In top-down proteomics, intact proteins are directly analyzed via tandem mass spectrometry techniques allowing for a variety of experiments of profiling, functional and structural proteomics [5, 9]. The main advantage of top down proteomics is that it allows for the identification and analysis of post-translational modifications (PTM) and PTM hierarchies [5, 14]. For example, Moradian et al. 2013 noted that top down proteomics has been successful in analyzing certain PTMs of histones [5]. However, the efficacy of top down proteomics is limited by poor protein fractionation and mass spectrum resolution (due to the large masses of intact proteins >10kDa), and difficulties in fragmentation of large proteins [3, 5, 14]. A top down method is also useful for conducting profiling proteomics experiments. More recently, advances in orbitrap mass analyzers and Fourier transform ion

cyclotron resonance (FT-ICR) mass spectrometry have increased the viability of top down proteomics experiments (increased mass resolution and dynamic range), but widespread experimentation via these techniques is still cost prohibitive [5, 15].

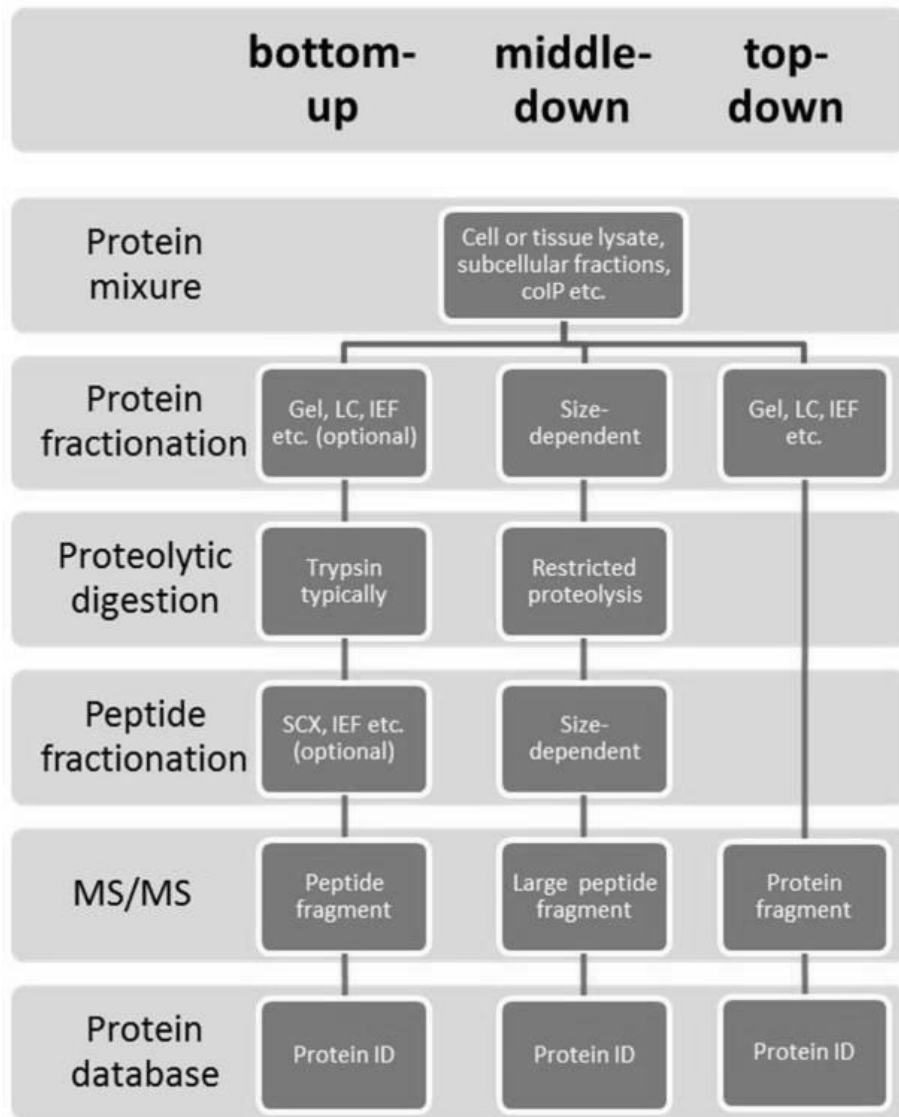


Figure 1.2 Summary of diagram of bottom up, middle down, and top down proteomics techniques. Adapted from [3].

In middle down and bottom up proteomics, the proteins are enzymatically digested prior to analysis with MS/MS [1-5, 16]. Middle down proteomics involves restricted protein digestion (for example using Glu-C or Asp-N enzymes) to create medium sized peptides (around 2kDa to 10kDa) [3, 5]. Middle down proteomics is advantageous because it retains some ability to study post translational modifications, while reducing the size of the peptides to increase instrumental resolution [3, 5]. The ability to analyze PTMs and obtain spectra with a higher resolution than top down approaches makes the middle down approach a strong choice for functional proteomics experiments [3, 5, 9].

Of all the forms of proteomics experiments, bottom up or “shotgun” proteomics remains widespread for profiling and structural experiments [3-4, 9]. As previously noted, the large size and complexity of biological proteins makes direct fragmentation in a mass spectrometer difficult [3, 5, 14, 16]. In bottom up proteomics, the protein of interest is enzymatically digested using a proteolytic enzyme such as trypsin to create smaller mass peptides (0.5kDa to 3kDa) that can more easily be analyzed by various mass spectrometry techniques [1-4, 7, 16-17]. Trypsin is most commonly used because of its high cleavage efficiency and high specificity for the C-terminus of lysine and arginine amino acid residues [16]. Additionally, tryptic digests create peptides of, on average, 9-10 amino acids long, which is very conducive to tandem mass spectrometric analysis [3, 16, 18]. However, the efficacy of using of trypsin as a proteolytic enzyme for bottom up proteomics was studied by Lowenthal, et al. in 2013 [16]. Their research suggests that there can be slight variability in the digestion patterns of trypsin, motivating further research on the effects of digestion conditions on the kinetics of proteolysis as used for bottom up proteomics [16]. Other physical digestion techniques such as microwave heating, acid

hydrolysis, and electrochemical oxidation have also been shown as a means of reproducibly digesting proteins into peptides of similar lengths to tryptic digests [3, 18-19]. Because of the widespread use of trypsin in bottom-up proteomics, analysis of the fragmentation patterns of peptides that contain C-terminal lysine or arginine residues are important for ensuring the reproducibility and accuracy of bottom-up proteomics protein characterizations.

High performance liquid chromatography (HPLC) is often used in bottom-up experiments to separate the peptides in a tryptic digest from one another before analysis with tandem mass spectrometry (MS/MS) [1, 3-5, 7-9]. MS/MS works by further separating the digests by selecting analytes of a specific mass, and fragmenting these digest peptides using a method such as collision-induced dissociation (CID) to obtain fragment ion spectra [1-5, 10-51]. The resulting fragmentation spectra can then be matched to theoretical peptide fragmentation spectra using computer searching algorithms such as SEQUEST, Mascot, MS-Fit, ProFound, MaxQuant, PEDRo, etc. [20-27]. These programs generate theoretical spectra that using well established fragmentation rules [20-27]. Once identified, the peptides can be matched to individual proteins by searching against a database containing known primary protein sequences [20-27]. Further research on the fundamentals of the peptide fragmentation process can improve the theoretical spectrum generation and matching accuracy of databases such as SEQUEST, Mascot, etc.

High performance liquid chromatography is essential to the bottom-up proteomics experiment because of its ability to separate digested peptides from one another before introduction into the mass spectrometer [1, 3-10]. This improves the resolution of the MS/MS spectra, and increases sequence coverage and sequencing accuracy [9]. HPLC is also

advantageous because it can be used in-line with electrospray ionization (ESI), a soft-ionization technique that is the most advantageous ionization source for studying biomolecules [1-5, 8-9, 16, 28-29]. In addition to HPLC-ESI-MS/MS, techniques such as multidimensional high performance liquid chromatography have been developed to achieve an improved separation resolution [9].

In their review paper on HPLC techniques for proteomic analyses, Goran Mitulovic and Karl Mechtler stress the importance of multidimensional techniques, optimized flow rate, and optimized column stationary phases for a robust HPLC separation [9]. The complexity and sheer number of peptides from the enzymatically digested proteins tend to exceed the peak capacity of most one-dimensional HPLC columns [9]. For this reason, adopting a multidimensional approach by adding additional columns, which will increase peak capacity for the system, proves beneficial to applications in proteomics research [9]. Further, the injection volumes and flow rates of the HPLC system should be lowered to nano-liter (nL) injection volumes and 200-300nL/min flow rates for optimized proteomic separation and analysis [9]. If nL sample injection volumes are used in the HPLC, nanospray-ESI is the best sample introduction method for the subsequent MS/MS [16]. Finally, column stationary phase compositions should be taken into consideration when conducting proteomics experiments. Immobilized metal-ion chromatography (IMAC) implemented prior to nanoHPLC, while using a titanium dioxide (TiO<sub>2</sub>) column (because of its high affinity to organic phosphates) makes for an excellent peptide separation scheme [9].

Once the peptides are appropriately separated, they must be introduced into the mass spectrometer. In early proteomics research, matrix assisted laser desorption ionization (MALDI)

was commonly used for sample introduction after gel-electrophoresis (generally 2D-PAGE) separation [9]. The MALDI source is often coupled with a time of flight (TOF) mass analyzer, which is capable of mass analyzing large molecular ions (up to approximately 300kDa) [9, 55]. The disadvantage of using MALDI-TOF for proteomics experiments is that despite high resolution, the peptide introduction is off-line, meaning the peptides must be spotted on a target for introduction into the MS [9]. Consequently, the sample preparation process is time consuming, and therefore the use of inline HPLC-ESI-MS/MS is a more efficient and appealing technique for proteomics researchers (see Figure 1.3). Additionally, ESI is a proven soft ionization sample introduction technique for analysis of biomolecules as it causes minimal disruption of the native state structures upon ionization [28-29]. ESI works by flowing sample solution through a charged capillary tube, creating a spray of charged droplets that undergo coulombic explosion and solvent evaporation resulting in molecular ion formation (see Figure 1.4) [28]. In 1989, John Fenn showed that ESI also can form multiply charged ions, allowing for analysis of very large biomolecules with mass spectrometry [28].

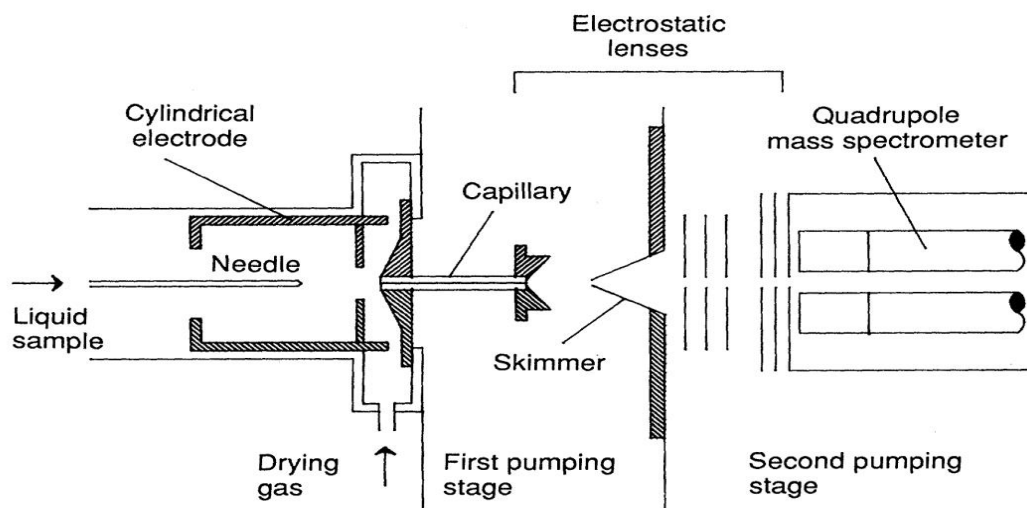


Figure 1.3 Example of an LC in line with an ESI source entering a quadrupole mass spectrometer (multiple quadrupoles would allow MS/MS). Obtained from [28].



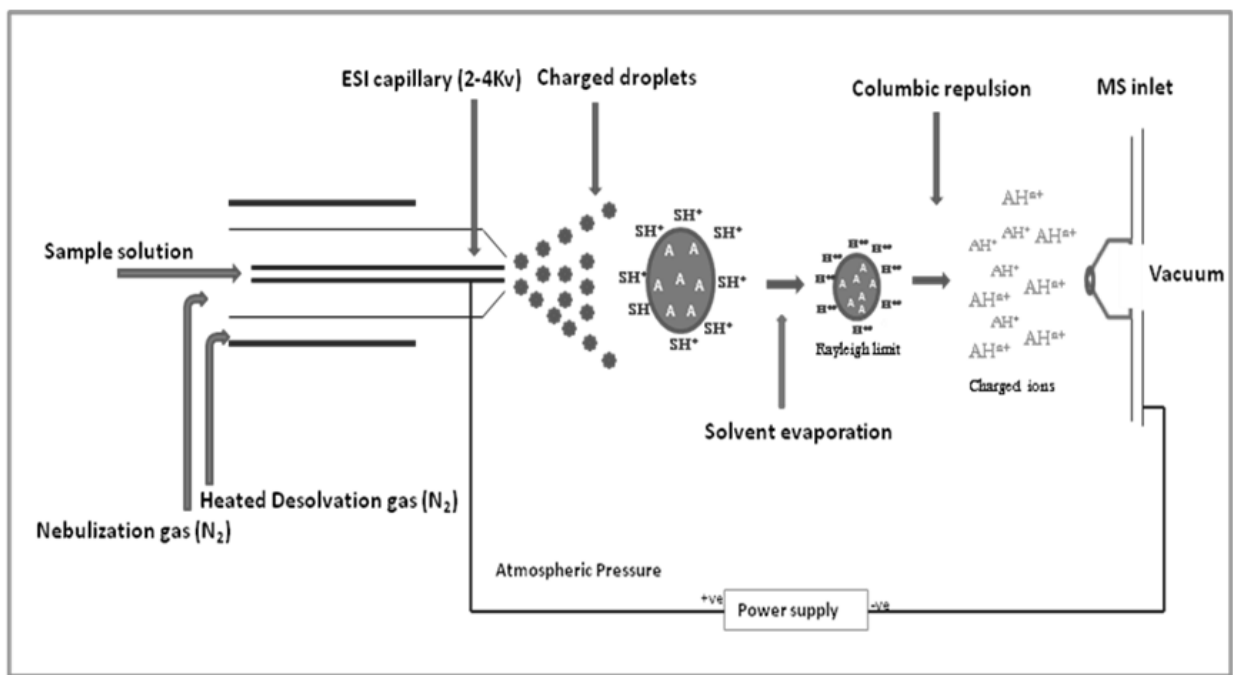


Figure 1.4 Schematic of ESI solvent droplet undergoing coulombic explosion and solvent evaporation to produce molecular ions of the analyte. Obtained from [\[http://www.chem.pitt.edu/sites/default/files/figure3.PNG\]](http://www.chem.pitt.edu/sites/default/files/figure3.PNG).

Tandem mass spectrometry (MS/MS) involves the mass selection (various mass analyzers can be used) and fragmentation of peptides, creating a mass spectrum of the peptide [1-5, 10-51]. The fragment ions that are analyzed originate from cleavages along the peptide backbone of a peptide [21, 30-47]. A common method of achieving this peptide bond cleavage is through collision induced dissociation (CID) activation, a hard-fragmentation technique [30]. In the MS/MS experiment, the peptides are separated in a first mass selection stage. Next, inert gas is leaked into the instrument and allowed to collide with the molecular (peptide) ions [17]. During these collisions, the kinetic energy of the inert gas molecules is converted into internal energy within the peptide molecular ions, causing bond breakage and fragmentation of peptides

[17]. These fragments can then be mass analyzed to generate a mass spectrum for the peptide that can be compared to sequence databases (see Figure 1.5).

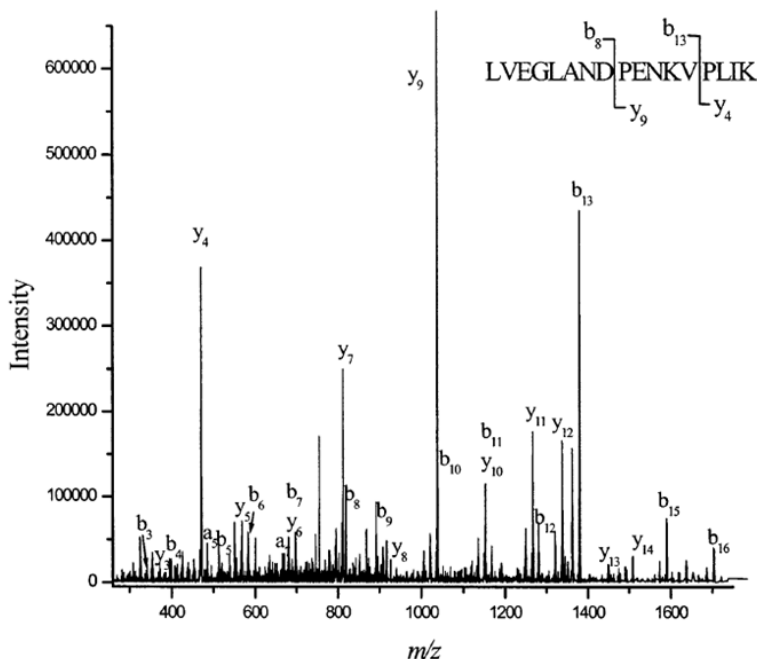


Figure 1.5 Example of a MS/MS CID spectrum of the peptide LVEGLANDPENKVPLIK using an ion trap mass spectrometer. Obtained from [41].

In addition to collision induced dissociation, other hard-ionization techniques are available for use in tandem mass spectrometry, including surface induced dissociation (SID), electron impact (EI), infrared multiphoton dissociation (IRMPD), electron capture dissociation (ECD), electron transfer dissociation (ETD) and electron induced dissociation (EID) [31]. SID works by colliding ions with a surface to cause dissociation, and is designed to cause higher energy ion collisions than CID [32]. As such, SID can be used to probe quaternary structure of protein complexes in the gas phase [32]. Mosely et al. noted that ECD was particularly useful for identifying functional group modifications within protein samples such as phosphorylation,

sulphation, acetylation or glycosylation [31]. EID and ECD produce particularly good data that can complement CID data, but using CID produces spectra in a more consistent manner [31]. Thus CID would make a slightly better dissociation technique for creating a uniform database of fragmented peptide spectra and sequences.

After fragmentation spectra are obtained, a searching algorithm and database of predicted peptide fragmentation spectra and known sequences such as SEQUEST, Mascot, etc. is used to identify the experimental sample [20-27]. According to Dong et al., the searching algorithms generally incorporate categories of data for predictions, including de novo sequencing, protein sequence database searches, sequence tag approaches, and a spectral library search [20]. For example, the searching algorithm SEQUEST uses the spectral library search and a protein sequence database search [20]. The protein sequence database data is based on predicted peptide fragments due to theoretical enzymatic digestion [20]. However, there are many potential complications of the peptide fragmentation process that need to be fully studied to ensure accurate theoretical spectra for comparison are generated [20, 23]. Using protein sequence database matches in combination with spectral data matches likely produces the most accurate and robust results [23]. In order to create high accuracy databases, research into the methods of enzymatic digestion and into complications in the peptide fragmentation process is vital [1-4, 16, 23, 30].

## 1.2 Peptide Fragmentation

### 1.2.1 Fragmentation Mechanisms

Peptide fragmentation pathways in MS/MS are studied to better understand the effects amino acid residues within the peptide sequences have on the fragmentation patterns [2, 30-47]. When ionized by ESI, singly charged peptides are protonated at the most basic site within the peptide, which is usually the N-terminus of the peptide. However, when basic amino acid residues such as lysine, arginine, or histidine are present, the peptide will be protonated on the side chains of these amino acids. When fragmented, peptides would be expected to show random cleavage along multiple sites in the peptide backbone as shown in Figure 1.6 below [30]. In actuality, selective cleavages preferentially forming specific product ions have been documented extensively in the literature [2, 30-47]. Understanding these fragmentation pathways and their mechanisms are critical to improving matching with databases of MS/MS spectra [30].

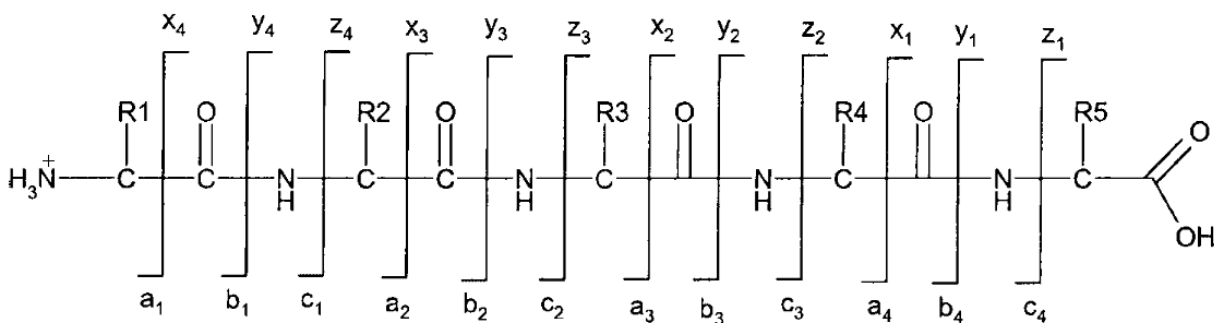


Figure 1.6 Potential fragmentation sites along the backbone of a pentapeptide and nomenclature of product ions formed. Obtained from [30].

As seen above in Figure 1.6, fragmentation along the backbone can form product ions referred to as a, x, b, y, c, and z ions. The types of product ions that are formed depends on the

dissociative technique used [3, 5, 14, 21, 30]. For example, using CID to fragment the peptide will result in the formation of y and b ions (with some a ions) [3, 21,30]. Using ECD or ETD to fragment the peptides will result in the formation of only c and z ions via different fragmentation mechanisms (see Figure 1.7) [3, 5]. When designing a proteomics experiment, the choice of which activation technique to use and its resulting fragmentation mechanisms must be carefully considered.

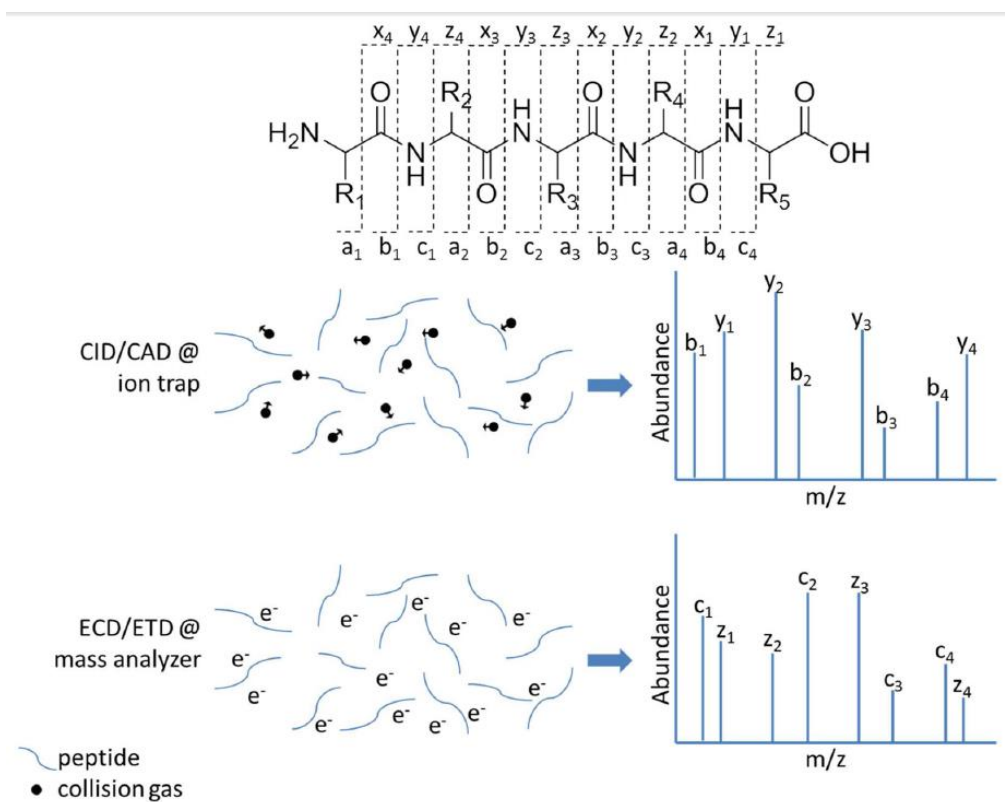


Figure 1.7 Schematic depicting the differing product ion formation when using CID vs ECD or ETD on a pentapeptide. Adapted from [3].

Under CID conditions, the mobile proton model is the most widely accepted model for describing peptide fragmentation mechanisms (see Figure 1.8) [30, 33]. In this model, the proton

creating the positive charge on the ionized peptide is mobilized by the applied CID energy, and can be relocated at any amine or oxygen site within the peptide (including amino acid side chains) [30, 33]. However, the mobility of this proton can be obstructed by basic amino acid residues, thus resulting in selective ion formation [30, 33]. In Figure 1.8, an example mobile proton mechanism involving the charge-directed nucleophilic attack induced by the proton located on the amide oxygen or amide nitrogen is depicted [33]. Protons at these locations would result in b and y ion formation under CID fragmentation conditions [33].

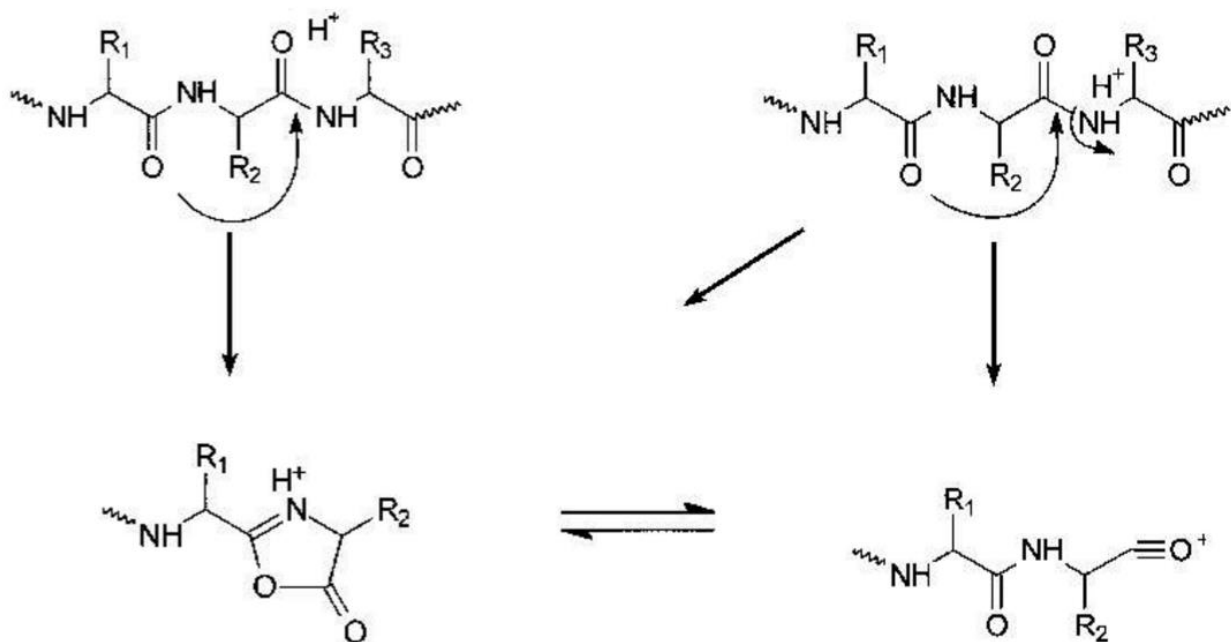


Figure 1.8 Mobile proton mechanism depicting formation of b ions (depicted) and corresponding y ion (not shown). Adapted from [33].

### 1.2.2 Variations of b ions

Under CID of peptides, b ions are formed in various isomers [30, 34-36]. The mobile proton mechanism in Figure 1.8 shows the formation of a linear acylium (right) or an oxazolone

derivative (left) b ion [33]. Using infrared multiphoton dissociation (IRMPD) studies it has been experimentally shown that b ions will form the more stable oxazolone over the linear acylium ion [34]. The mechanism for the  $b_x$ - $y_z$  pathway forming an oxazolone b ion involves nucleophilic attack by an oxygen (N-terminal) on the carbonyl carbon of the neighboring (C-terminal) amide bond when the mobile proton is on the oxygen of the carbonyl carbon (see Figure 1.9) [30]. Alternatively, it has also been shown that cyclic peptide b ions can be formed by nucleophilic attack of the N-terminus on the carbonyl carbon of the amide bond where the mobile proton is on the carbonyl oxygen (see Figure 1.10). [30, 34]. In the case of  $b_2$  ion formation a diketopiperazine structure is formed, and can be protonated via proton transfer from the y ion (Figure 1.11) [30].

The formation of diketopiperazine b ions over oxazolone b ions has been suggested to be governed by the length of the peptide chain and the identities of the first three amino acid residues [37]. There is also evidence in the literature that a mixture of the two structures can be formed with histidine containing peptides (this was noted during IRMPD studies of histidine-alanine<sub>x</sub>  $b_2$  ions and CID studies) [36, 38-39]. Computational modeling has shown that the side chains of arginine and lysine containing peptides can stabilize oxazolone b ion structures via intramolecular hydrogen bonding [40]. However, this modeling also showed that the most stable arginine b ion was a cyclic peptide structure (the energies of oxazolone and cyclic peptide structures for lysine were fairly similar) [40]. Finally, hydrogen deuterium exchange (HDX) experiments on doubly protonated tryptic peptides with lysine or arginine at the C terminus showed that the protonated oxazolone b ions were preferentially formed over other cyclic peptide b ions [35].

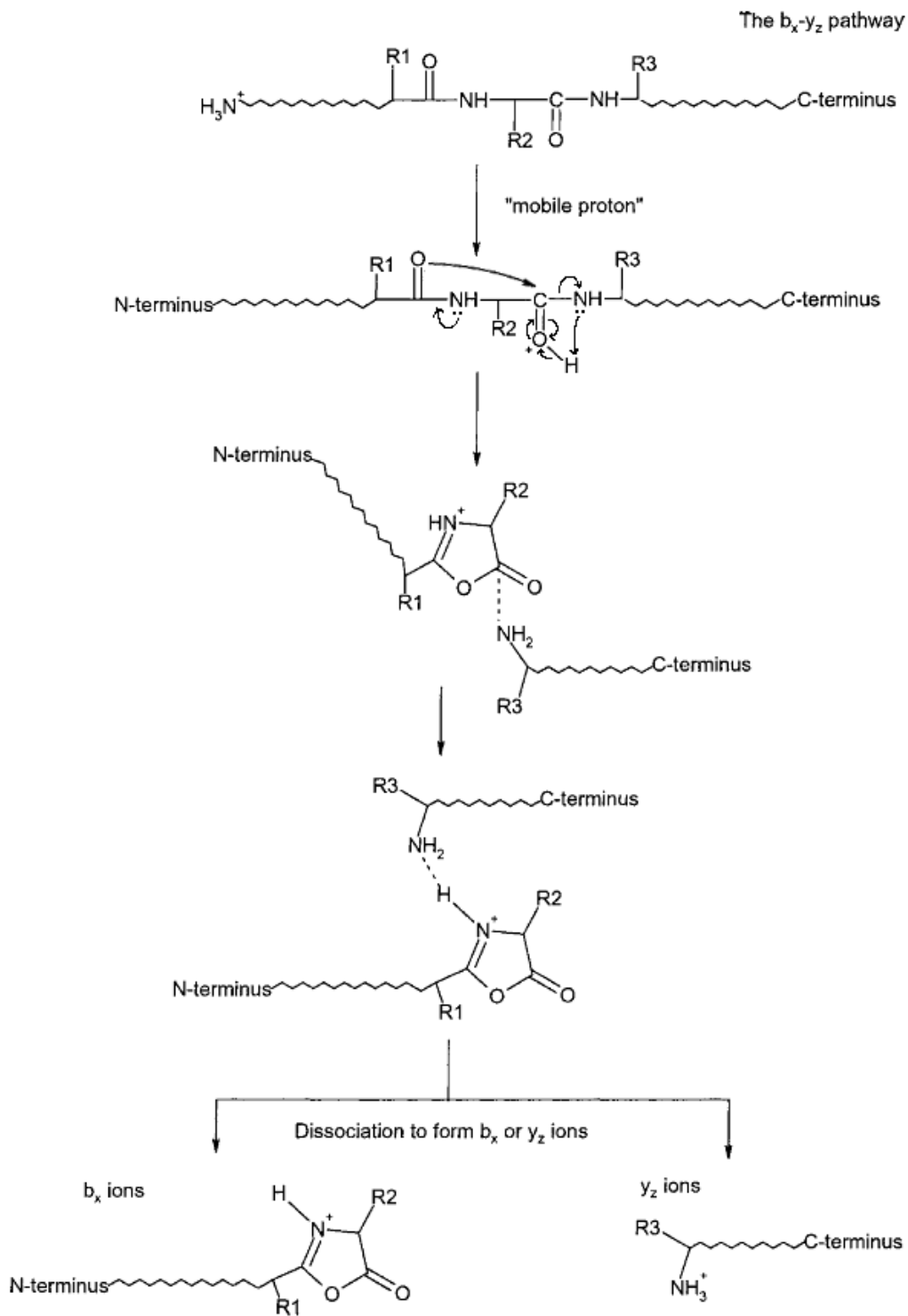


Figure 1.9 Mobile proton mechanism of  $b_x$ - $y_z$  pathway forming oxazolone derivative b ion and y ion formation. The ion-neutral complex can dissociate and the proton can transfer to either the b or y ion. Adapted from [30].



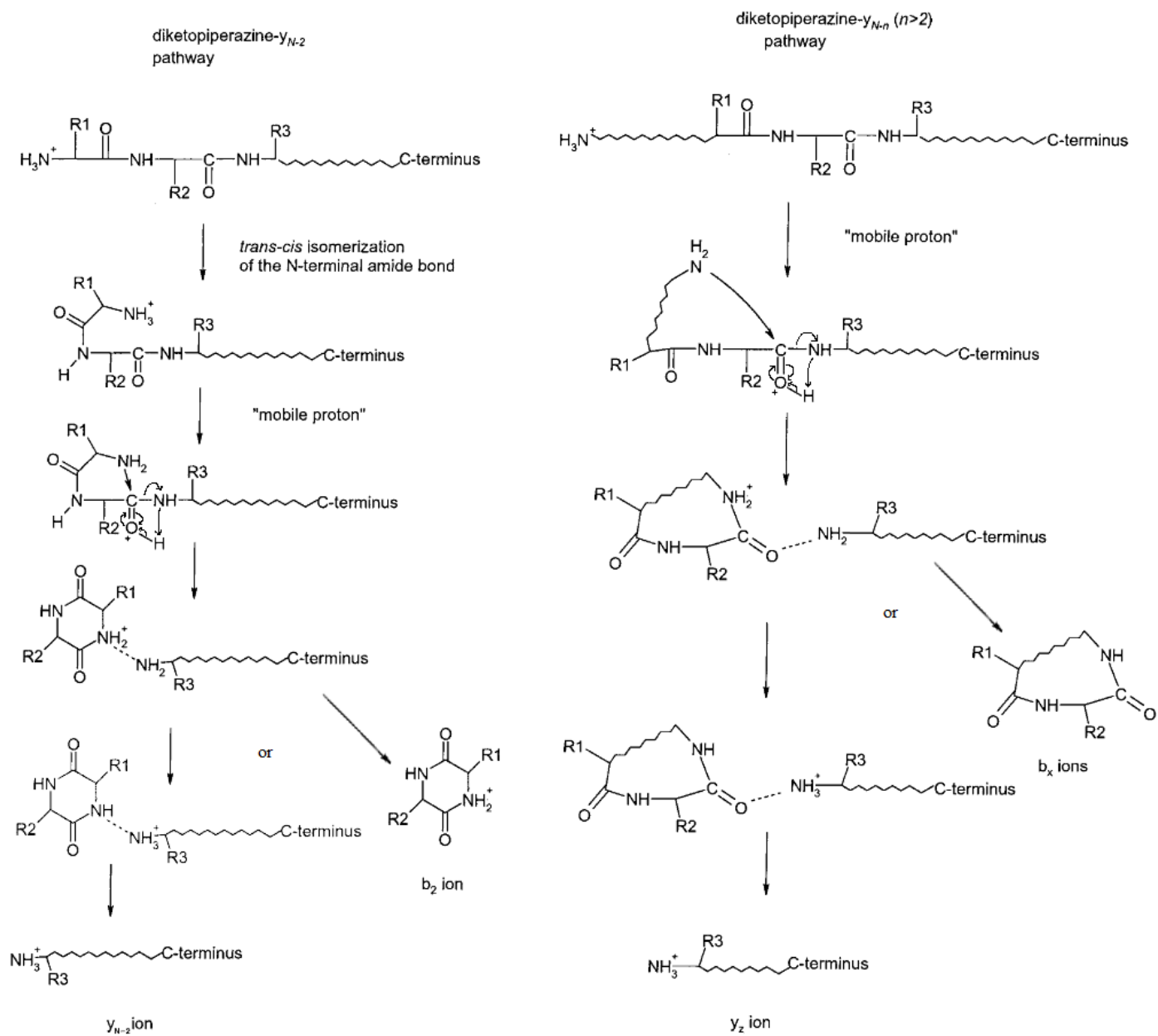


Figure 1.10 Mobile proton mechanism of  $b_2$  ion diketopiperazine formation (left) and cyclic peptide  $b_x$  ion formation (right) via N-terminus nucleophilic attack. The ion-neutral complexes can dissociate and the proton can transfer to either the b or y ion. Adapted from [30].

### 1.2.3 Selective Cleavages

There are many reports of selective cleavage effects of amino acid residues throughout the literature [1-4, 30-47]. An important selective cleavage that has been noted by multiple researchers is the proline effect [2, 30, 41-43]. The proline effect involves an energetic preference for N-terminal cleavage of proline residues within the peptide during fragmentation forming  $\gamma$  ions [2, 41-42]. While the mechanism of the proline effect is still being studied, it is believed that the dissociation occurs in a two-step process involving the amide bond cleavage and proton transfer, and decomposition of the intermediate to form a  $\gamma$  ion fragment (see Figure 1.11) [2].

The proline effect is believed to be partially attributed to the ring structure of proline and the increased basicity of the prolyl-amide site within the proline containing peptide [2]. The proline effect was also found to be influenced by the adjacent N-terminal residue, where the effect was strongly observed for Val, His, Asp, Ile, and Leu, while there was a decrease in effect with adjacent Gly or Pro [41]. In addition to the proline effect, a more marked selective cleavage was observed for pipecolic acid, a six membered ring analog of proline [2]. This pipecolic acid effect results in C-terminal cleavage of the peptide bond, and forms  $b$  ions [2]. The pipecolic acid effect is believed to not be governed by the basicity of the residue, but instead is likely dependent on the increased flexibility of the ring structure as compared to proline [2].

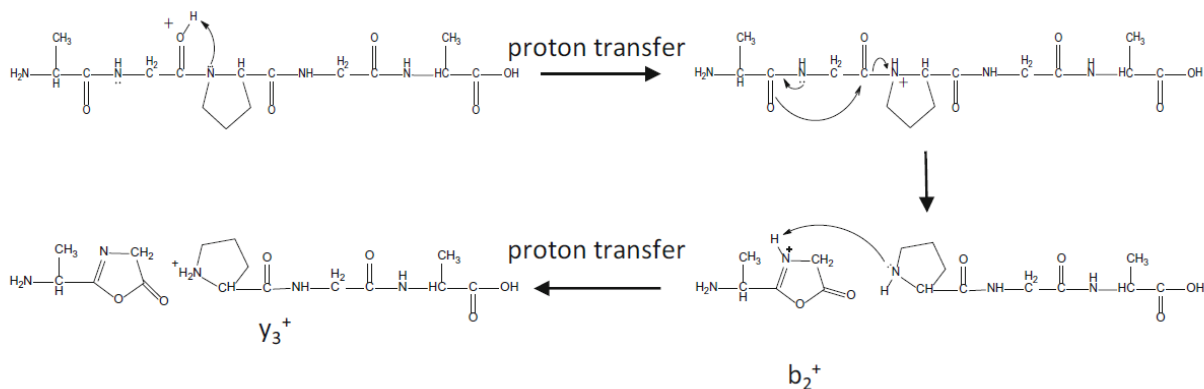


Figure 1.11 Proposed mechanism by Raulfs et al. for b and y ion formation based on the proline effect Obtained from [2].

The histidine effect involves preferential cleavage at the C-terminal side of the histidine residue, and is mediated by the protonated histidine side chain [30, 33]. Histidine has a high proton affinity (PA) and is therefore likely to be protonated in the gas phase [30]. The histidine effect also produces a unique bicyclic b ion fragment (see Figure 1.12) [30]. Similarly, doubly protonated peptides containing histidine showed a C-terminal preferential cleavage resulting in the same b ion formation [30, 44]. Finally, Huang et al. in 2002 found cleavage at an aspartic acid residue was enhanced by a peptide that contained internal basic residues such as histidine [45]. They determined an order of basicity for amino acids as arginine (highest) > lysine > histidine (lowest) correlates to a higher degree of cleavage at the aspartic acid residue [45].

Histidine effect

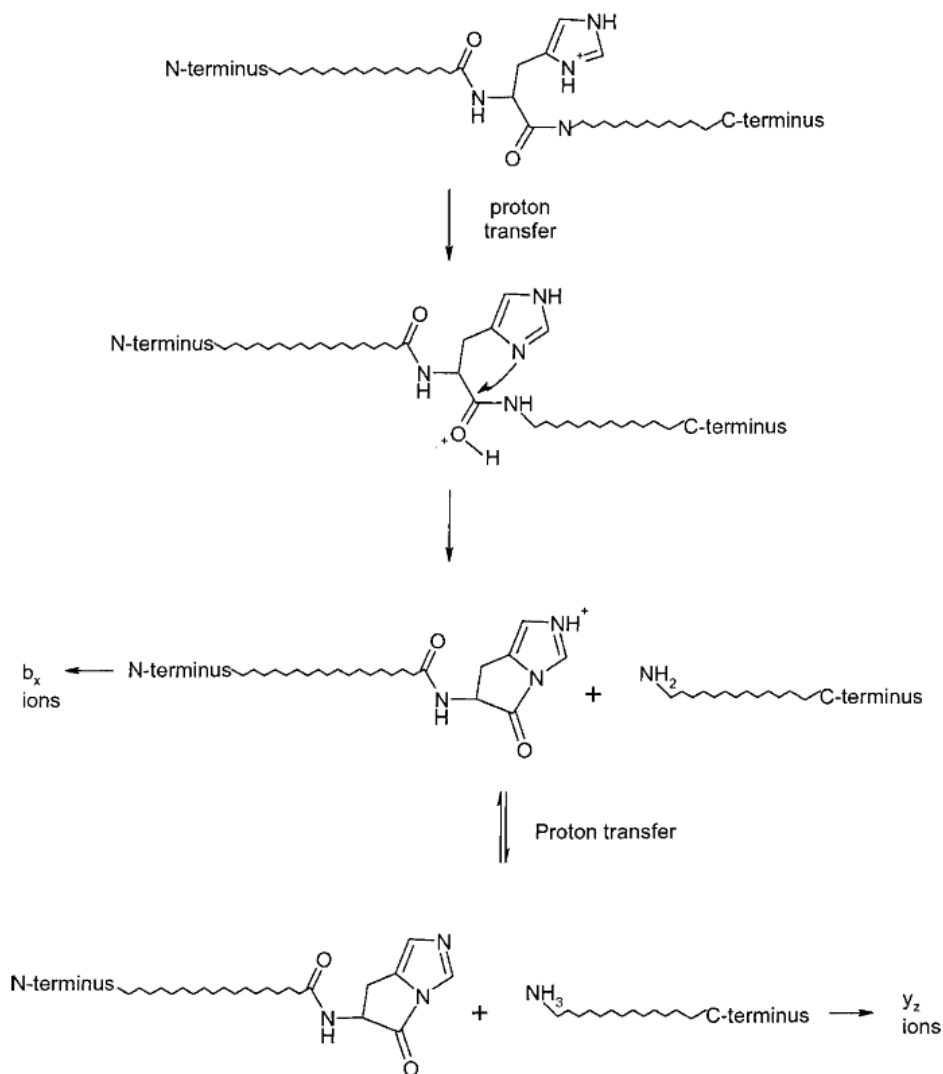


Figure 1.12 Mechanism for the histidine effect and formation of a novel bicyclic  $b$  ion. Obtained from [30].

There is evidence for glutamine (Gln) and asparagine (Asn) selective cleavage of the amide bond C-terminal to the amino acid residues [30]. These cleavage effects result in the formation of another novel  $b$  ion in the form of a cyclic isoimide (see Figure 1.13) [30]. A cysteine effect has also been shown to cause preferential C-terminal cleavage [46]. Additionally, selective C-terminal cleavages of oxidized cysteine residues (Cys-SO<sub>2</sub>H and Cys-SO<sub>3</sub>H) require

less energy to fragment than unmodified cysteine, providing a means of identifying and locating oxidized cysteine residues within peptides (see Figure 1.14) [46].

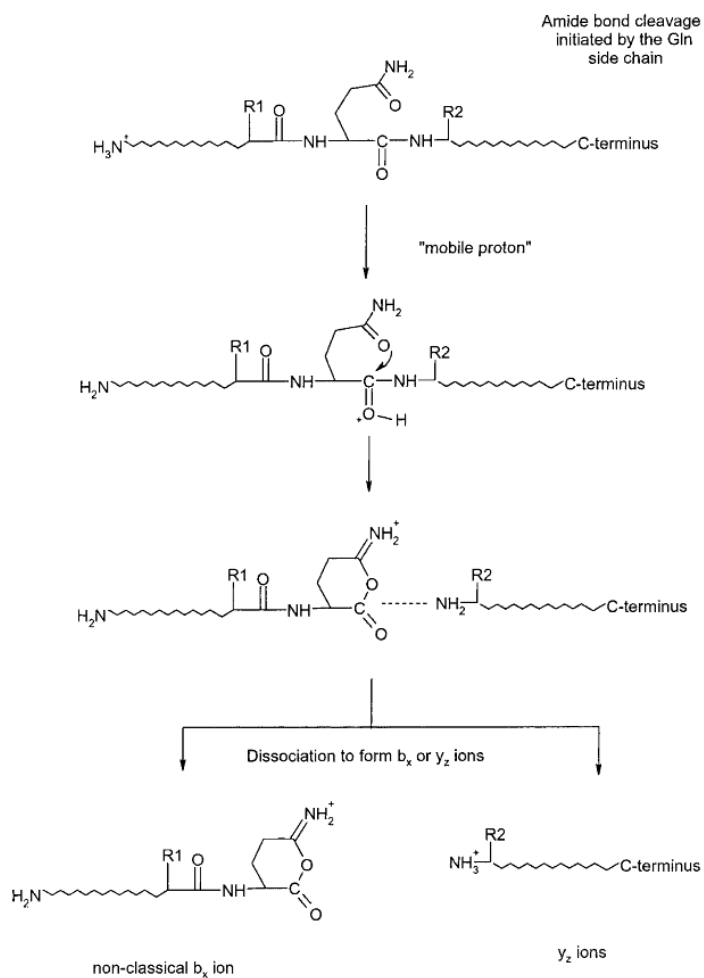


Figure 1.13 Mechanism of cyclic isoimide b ion formation with a Gln residue. Obtained from [30].

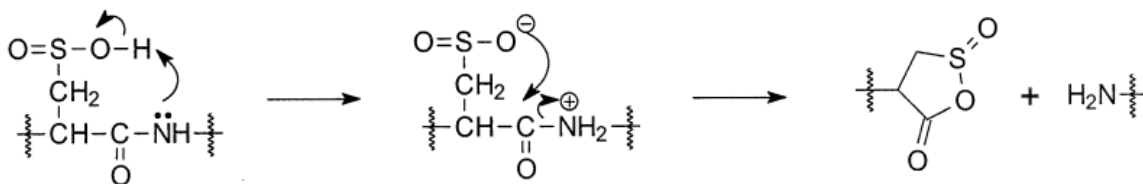


Figure 1.14 Mechanism of oxidized cysteine (Cys-SO<sub>2</sub>H) preferential C-terminal cleavage. Obtained from [46].

There are also interesting fragmentation pathways involving arginine and lysine side chains [30]. The side chain amino group of lysine can attack the C-terminal adjacent carbonyl of the peptide bond causing C-terminal cleavage [30]. A novel b ion in the form of a protonated 7 membered caprolactam derivative is formed in this cleavage (see Figure 1.15) [30]. This form was proposed because no loss of CO was found in MS/MS/MS studies of b ions from Lys containing peptides (an oxazolone would lose CO) [30]. Arginine residues also induce C-terminal cleavage via nucleophilic attack of the Arg side chain guanidinium group [30]. The resulting b ions form in a six membered ring as shown in Figure 1.15 [30].

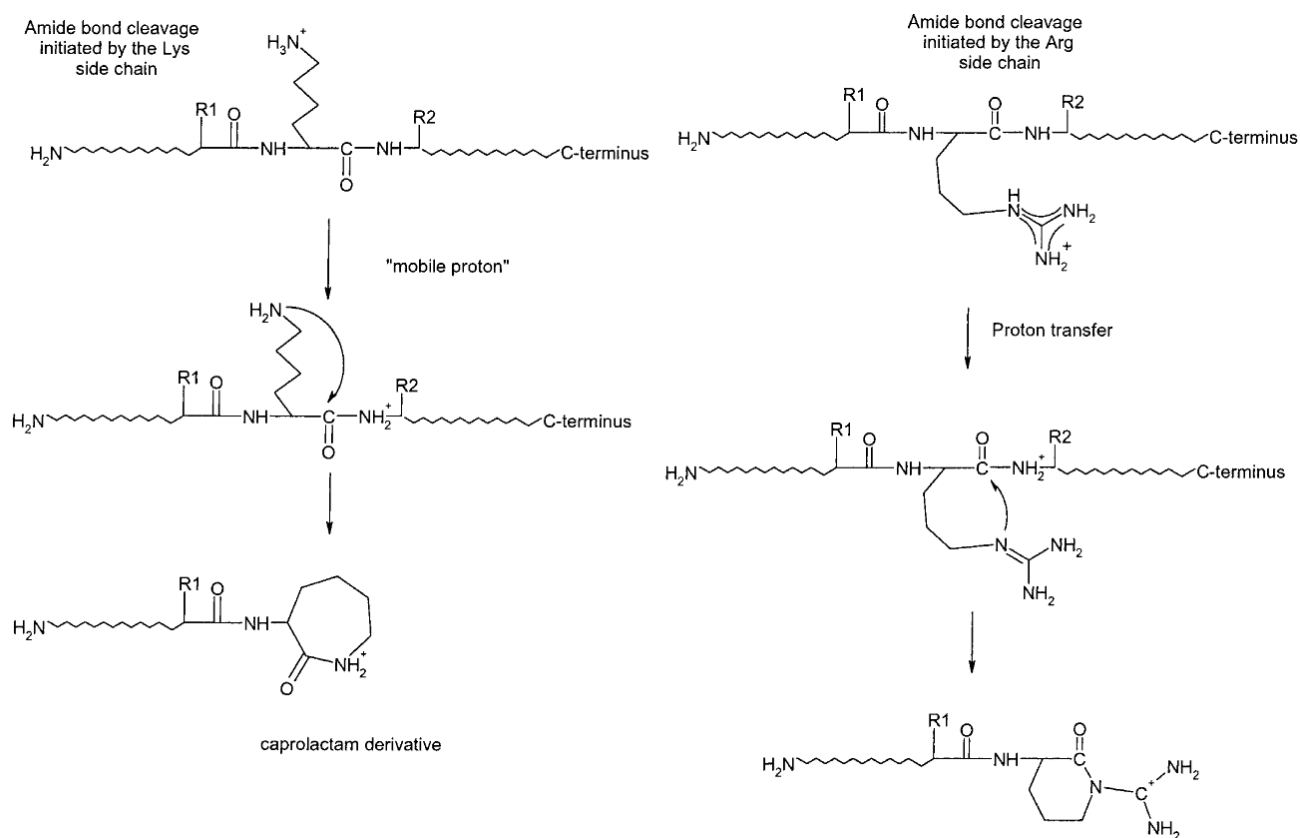


Figure 1.15 Mechanism for cleavage via Lys side chain to form caprolactam derivative b ion (left) and mechanism for cleavage via Arg side chain to form six membered ring b ion (right). Obtained from [30].

The ornithine effect involves preferential C-terminal cleavage of the amide bond in peptide fragmentation [47]. Similar to lysine cleavage, the ornithine side chain amino group acts as a nucleophile and attacks the adjacent amide bond carbonyl carbon to form a 6 membered lactam ring b ion (Figure 1.16) [47]. This preferential cleavage is even seen with the ornithine residue at the C-terminus of the peptide, forming a b ion and loss of H<sub>2</sub>O as a y ion [47]. McGee and McLuckey noted that the ornithine effect is more energetically favored than proline selective cleavages [47].

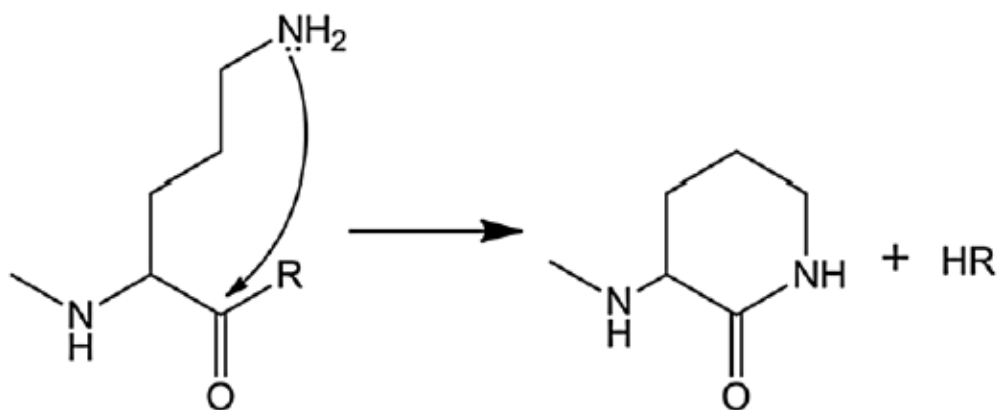


Figure 1.16 Mechanism of the ornithine effect showing nucleophilic attack of the ornithine side chain to create a 6 membered lactam derivative b ion. Obtained from [47].

### 1.2.4 Macrocyclization

It has been determined that peptide primary sequence scrambling can occur during CID of peptides in tandem mass spectrometry [48-50]. As the peptides are activated via CID, nucleophilic attack of the N-terminus on a carbonyl carbon at the C-terminus can result in a protonated macrocyclic intermediate [48-50]. This macrocyclic intermediate can then cleave at any amide bond location within the backbone, resulting in a new peptide with a scrambled sequence from the original [48-50]. For example, the pentapeptide YAGFL-NH<sub>2</sub> was studied and sequence scrambling was observed in the MS/MS spectrum [48-49]. The proposed mechanism for this scrambling first involves the formation of the largest possible b ion oxazolone fragment (YAGFL<sub>oxa</sub>), then nucleophilic attack by the N-terminus nitrogen on the oxazolone carbonyl group to form a protonated macrocycle (see Figure 1.17) [48-49]. The ring of this protonated macrocyclic intermediate of YAGFL<sub>oxa</sub> can then open via proton transfer, and scramble into any of the sequences AGFLY<sub>oxa</sub>, GFLYA<sub>oxa</sub>, FLYAG<sub>oxa</sub>, LYAGF<sub>oxa</sub>, or YAGFL<sub>oxa</sub> [48-49].

All of the above findings of selective peptide fragmentation and macrocyclic rearrangements are important to understand so that sequencing databases can accurately predict theoretical fragmentation spectra. The lysine and ornithine effects on selective cleavage and macrocyclization are of particular relevance to this thesis. Additionally, two other lysine analogs will be studied and analyzed for similar selective cleavage effects (see Section 1.3). With repeated research on selective cleavage in the fragmentation process, mechanisms can be accurately incorporated into peptide spectral and sequencing databases.



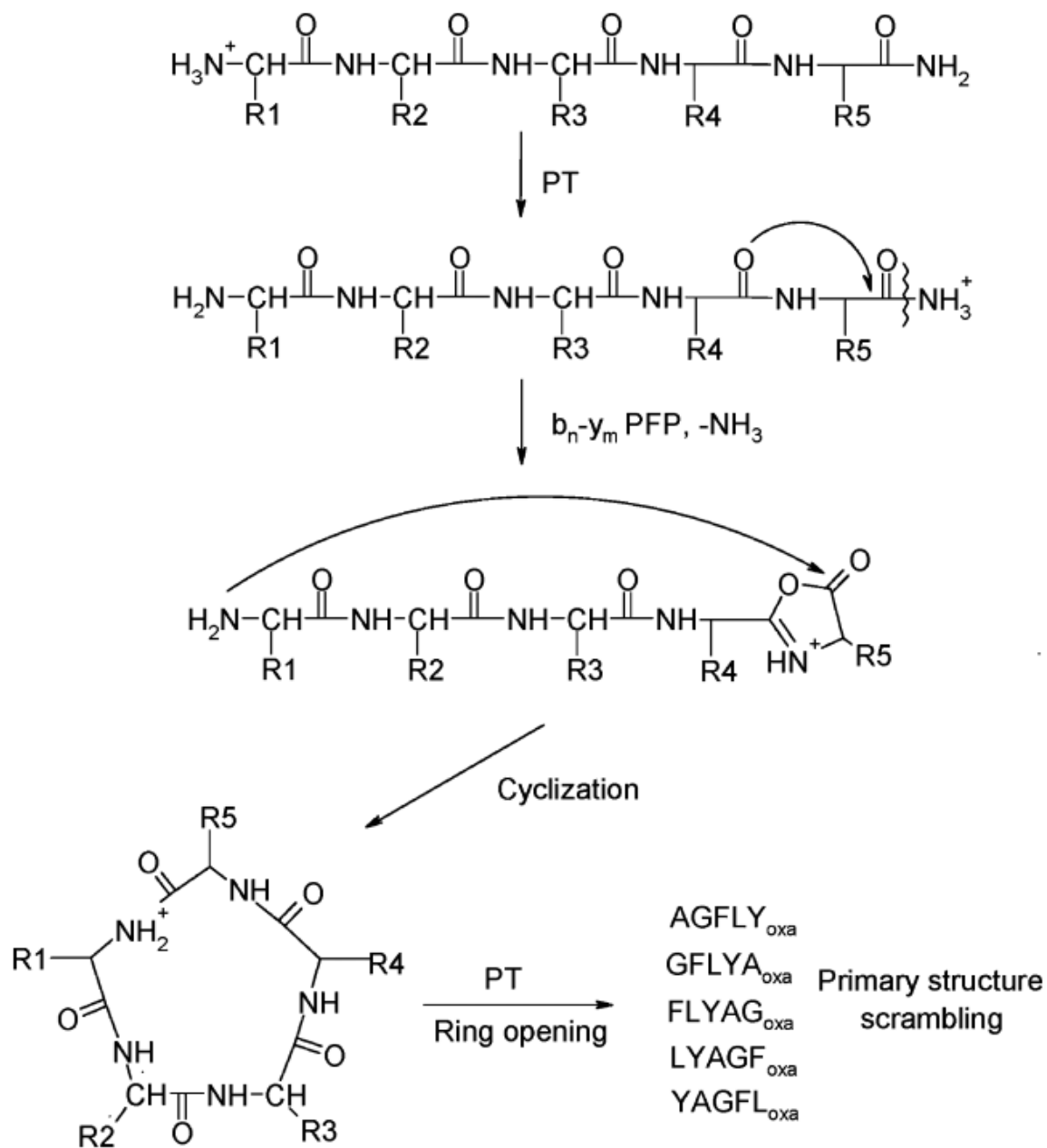


Figure 1.17: Proposed mechanism for macrocycle formation and primary sequence scrambling of the peptide YAGFL-NH<sub>2</sub>. Obtained from [48].

## 1.3 Tetrapeptide Fragmentation Studies

### 1.3.1 Lysine, Ornithine, DABA, and DAPA

Lysine (lys or K) and its analogs ornithine (orn or O), 2,4-diaminobutanoic acid (DABA or B) acid and 2,3-diaminopropanoic acid (DAPA or Z) are the amino acid residues studied in this tetrapeptide fragmentation experimentation (see Figure 1.18). Note these amino acids all differ by the number of carbons in their respective side chains. Lysine is the amino acid residue of this set that is biologically encoded for in humans (protein amino acids or PAAs). Its analogs ornithine, DABA, and DAPA are not encoded by the human genome and are described as non-protein amino acids or NPAAAs. Studying the fragmentation patterns of peptides containing lysine analogs may provide insight into the fragmentation mechanisms of lysine containing peptides.

The proton affinities (which can be used to obtain gas-phase basicity) of lysine and its NPAA analogs obtained using the extended kinetic method are  $1004.2 \pm 8.0$  ,  $1001.1 \pm 6.6$  ,  $975.8 \pm 7.3$ , and  $950.2 \pm 7.1$  kJ/mol for lysine, ornithine, DABA, and DAPA, respectively [51]. The differing proton affinities will affect the fragmentation patterns of peptides containing these residues. The characteristics of the NPAA fragmentation patterns can give greater insights into gas phase peptide fragmentation and the characteristics of lysine containing peptide fragmentation. Studying the mechanisms of peptide fragmentation containing these species can improve peptide sequencing databases and the robustness of a bottom up proteomics experiment involving a tryptic digest.

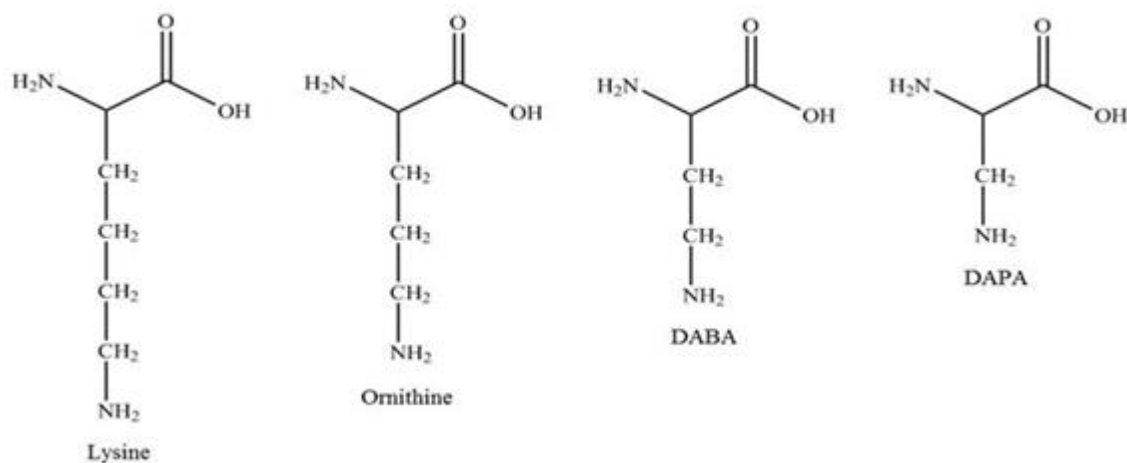


Figure 1.18 Lysine and its NPAA analogs Ornithine, DABA, and DAPA

### 1.3.2 Tetrapeptides of Interest

Previous Ionlab work has looked at lysine and lysine analog containing dipeptides and tripeptides. For this thesis tetrapeptides containing Lysine and its NPAA analogs were synthesized and studied via MS/MS CID studies in an ESI- Ion trap mass spectrometer. The goals of this study were to confirm the selective cleavage effects of ornithine and lysine in tetrapeptides, determine if similar effects for DABA or DAPA are present, and to search for potential macrocyclization mediated peptide sequence scrambling in tetrapeptides.

The first set of peptides analyzed were the set of 16 tetrapeptides AAAX, AAXA, AXAA and XAAA where A is alanine and X represents lysine, ornithine, DABA or DAPA. Alanine was chosen as the “filler” amino acid to complete the tetrapeptide because of its small size, non-polar and neutral side chain, and cost effectiveness. The next set of 4 peptides studied was YAGK, YAGO, and YAGB and YAGZ. These were chosen because they are similar tetrapeptide

versions of the pentapeptides in the CID fragmentation studies that showed the occurrence of macrocyclization [48]. Below is a proposed macrocyclic structure of a tetrapeptide (XAAA) showing the potential sequence scrambling based ring opening at any of the 4 peptide amide bonds (Figure 1.19).

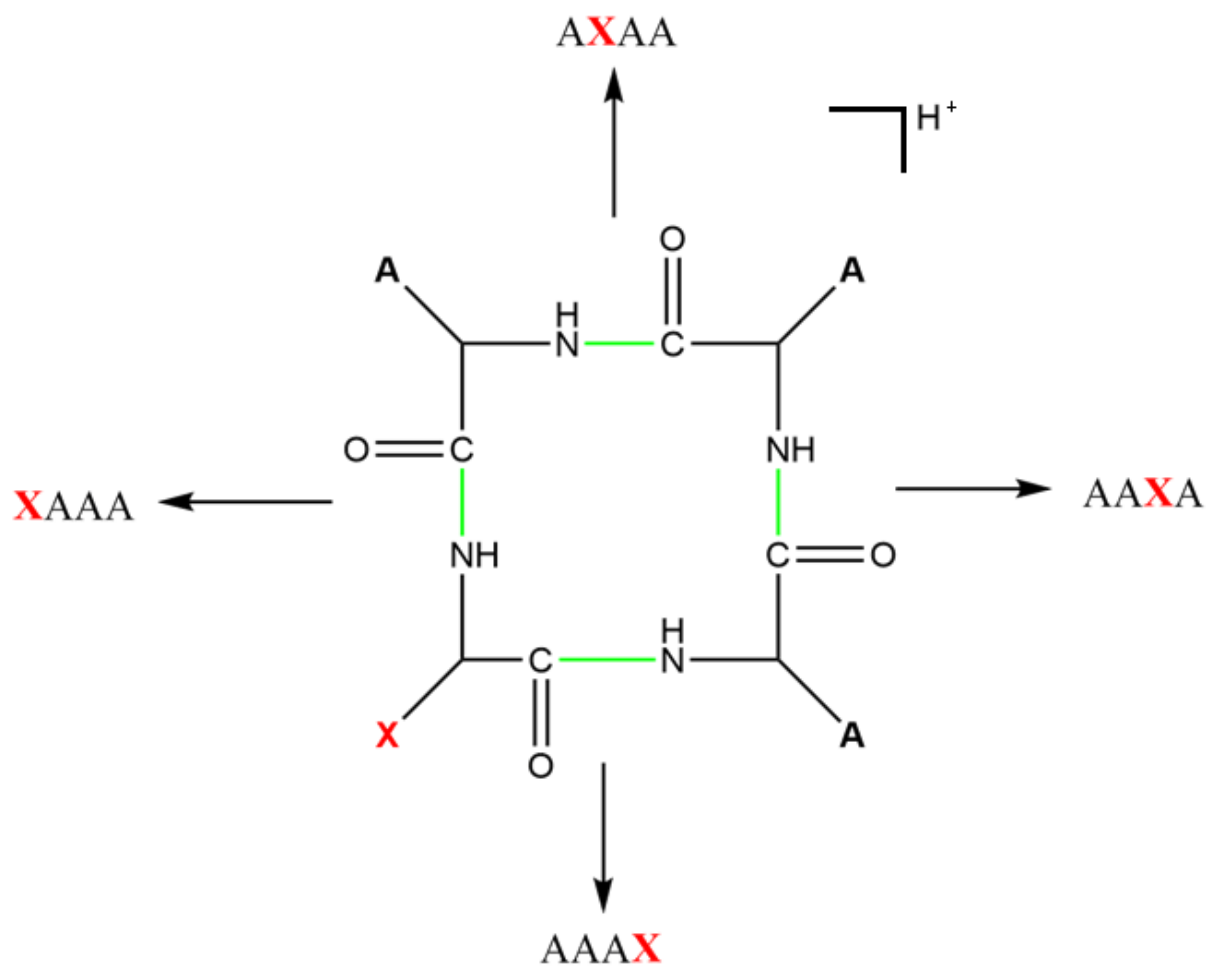


Figure 1.19 Proposed macrocyclic intermediate structure example for the XAAA peptide.

Cleavage at three of the four peptide bonds in the macrocycle would result in sequence scrambling to AXAA, AAXA, or AAAX.

## Chapter 2: Experimental Methods

### 2.1 Peptide Synthesis

Peptide synthesis was conducted using the Fmoc solid phase peptide synthesis (SPPS) method. SPPS uses an insoluble support resin to anchor an amino acid that can be systematically added upon, creating a peptide chain [52-54]. The resin is covalently bound to the C-terminus of this first amino acid residue, and thus peptides are assembled from C to N terminus. The Fmoc SPPS strategy has been shown to effectively synthesize peptide chains of up to 50 amino acids in length [52-54]. In our syntheses, a Wang resin (4-Benzyloxybenzyl alcohol, polystyrene polymer bound) was used as a solid support (see Figure 2.1). In SPPS, the Fmoc (9-fluorenylmethoxycarbonyl) is used as a protecting group on the chemically reactive N-terminus of the amino acids [52-54]. Additionally, reactive amino acid side chains are protected using Boc (t-butyloxycarbonyl), t-butyl (tBu) groups, or a variety of other organic molecules [52].

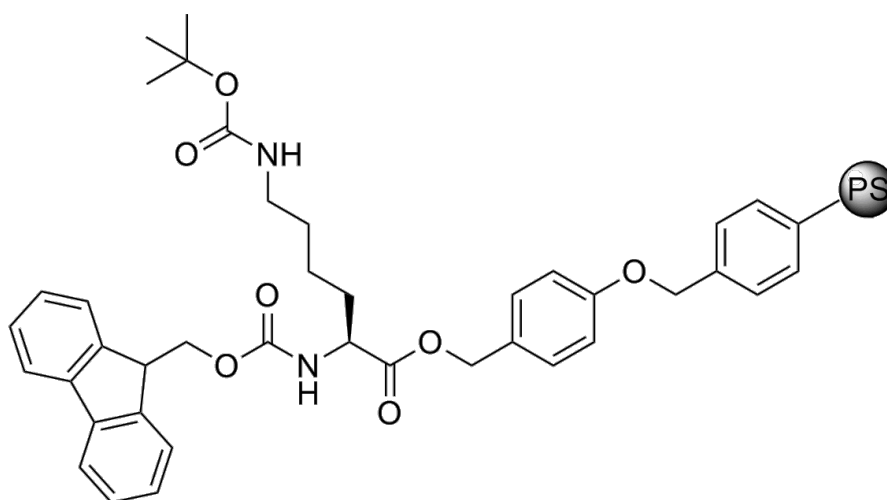


Figure 2.1 Fmoc-Lys(Boc)-Wang resin showing the Lys C terminus bound to the Wang resin, Fmoc protecting group on the Lys N terminus, and Boc protecting group on the Lys side chain. Obtained from [<http://www.matrix-innovation.com/wp-content/uploads/2013/08/Fmoc-L-LysBoc-Wang-resin.png>].

Our Fmoc SPPS strategy begins by adding a pre-loaded amino acid bound Wang resin (such as shown in Figure 2.1) into a 12mL plastic syringe with a guard that only allows liquids to permeate out of the nozzle. Next, a 50:50 dichloromethane (DCM): N,N-dimethylformamide (DMF) solution is used to swell the Wang resin in order to increase resin loading capacity and amino acid accessibility. The Fmoc protecting group is then removed via two 20:80 Piperidine:DMF solution wash steps. The Fmoc is removed from the synthesis syringe using a series of DCM and DMF wash steps. Next, an Fmoc protected amino acid (Fmoc-amino acid-OH) to be added to the peptide chain is dissolved in DMF with the coupling reagents HCTU (2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate) and DIEA (N,N-diisopropylethylamine), which are used to increase the rate of peptide bond formation. The HCTU (electron withdrawing group) activates the C-terminus carboxyl group of the added Fmoc amino acid for nucleophilic attack by the unprotected N terminus of the amino acid bound to the Wang resin resulting in peptide bond formation (and loss of H<sub>2</sub>O) [52]. The DIEA promotes the synthesis to proceed with no racemization [52]. Finally, any unreacted reactants are washed out of the syringe via a series of DMF and DCM washes. The above process is repeated until the desired peptide chain is synthesized.

After synthesis, the peptide must be deprotected and removed from the Wang resin. This is accomplished using a series of 20:80 Piperidine:DMF wash steps to remove Fmoc, and a 95% trifluoroacetic acid (TFA), 2.5% triisopropylsilane (TIS), and 2.5% H<sub>2</sub>O solution to cleave the peptide from the Wang resin. The cleavage solution will also remove any side chain protecting groups on the peptide. The cleaved solution is then diluted with anhydrous diethyl ether, to promote peptide precipitation. Finally, the precipitated peptides are centrifuged to remove solvent, and allowed to air dry to obtain a solid peptide sample.

### 2.1.1 Example synthesis of Alanine-Alanine-Lysine-Alanine (AAKA)

An example synthesis of the tetrapeptide AAKA using the Fmoc SPPS method will be described below. Fmoc protected amino acids, Wang resins, and HCTU were purchased from ChemPep. Piperidine, TFA, DIEA, and triisopropylsilane were purchased from Sigma-Aldrich. Anhydrous diethyl ether, DCM and DMF were purchased from Fischer Scientific. Approximately 5mL of solution was used for each wash step, and was conducted for the denoted time on a Vortex-Genie 2 shaker. The syringe nozzles were capped with NMR tube cap septums during the shaking process. This procedure, outlined in Figure 2.2 below, was used for synthesis of all 20 tetrapeptides in this study.

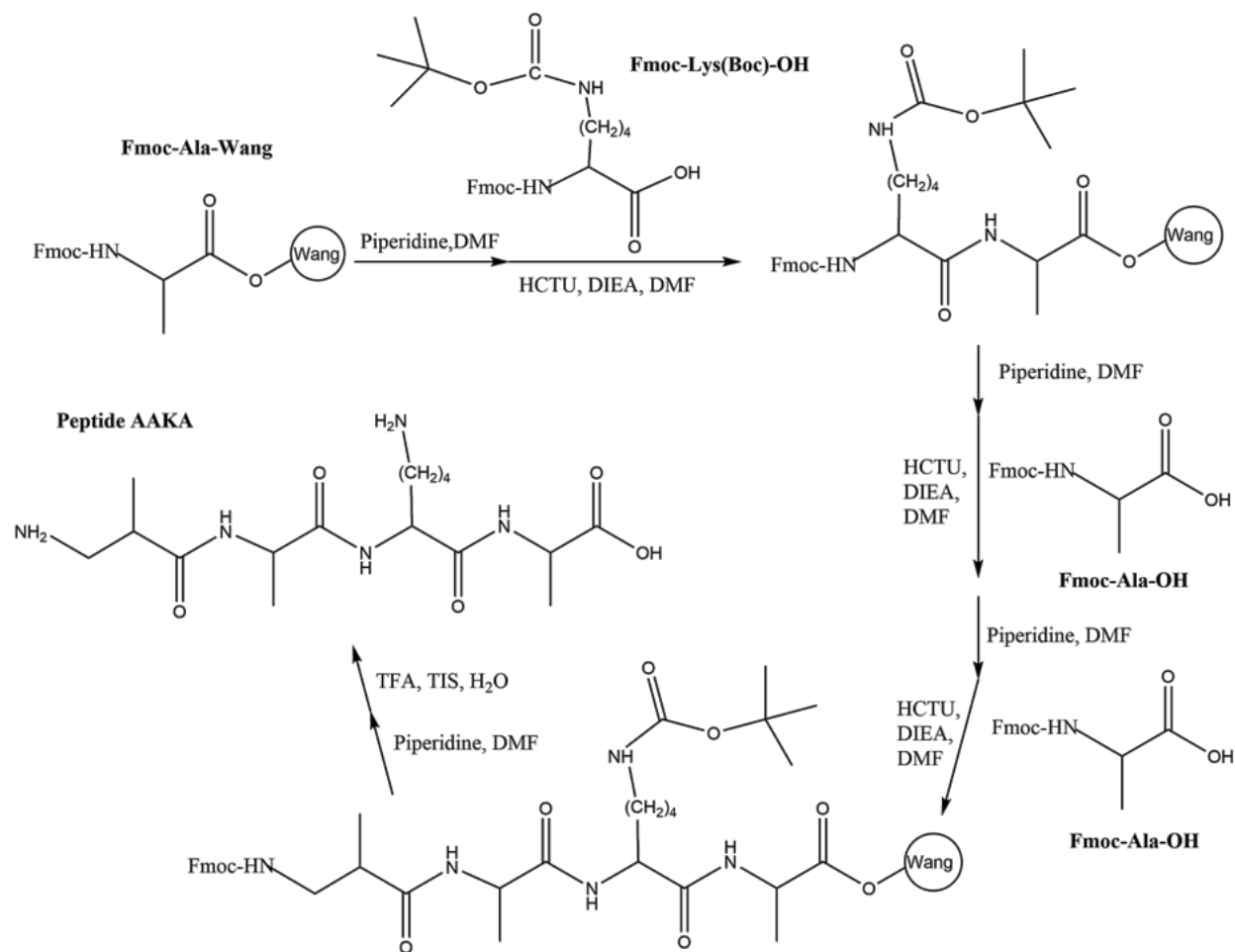


Figure 2.2 Example synthesis summary of AAKA.

First, 0.2g of Fmoc-Ala-Wang resin was added to a 12mL syringe. The resin was then swelled for 30 minutes using a 50:50 DCM:DMF solution followed by two 1 min 5mL DMF wash steps. Fmoc deprotection was accomplished via one 5 min and a subsequent 30 min wash steps with 5mL of 20:80 Piperidine:DMF. Two 1min DMF washes, four 1min DCM washes, and two 1 min DMF washes removed the Fmoc protecting group and prepared the resin for addition of the next amino acid. Next, a solution of Fmoc-Lys(Boc)-OH (amount dependent on Ala-Wang resin loading capacity) with HCTU and DIEA dissolved in DMF was added to the syringe and shaken for 1 hour. Subsequently, two 1 min DMF steps and four 1 min DCM wash steps removed excess unreacted reagents. This process of conducting Fmoc-Ala-OH additions was repeated twice, creating a Fmoc-AAK(Boc)A-Wang protected tetrapeptide.

The Fmoc-AAK(Boc)A-Wang was swelled with 50:50 DCM:DMF and deprotected with 20:80 Piperidine:DMF as outlined above. Next, two 1 min DMF wash steps and four 1 min DCM wash steps were conducted. The syringes were then dried with suction. A 10mL cleavage solution of 95% trifluoroacetic acid (TFA), 2.5% triisopropylsilane, and 2.5% H<sub>2</sub>O was added to the syringe and placed on the shaker for 2 hours. Finally, the contents of the syringe, with the peptide now in solution, were injected into a 100mL round bottom flask and 30mL of cold anhydrous diethyl ether was added (giving a total volume of 40mL). This flask was capped and stored in the freezer for 24-48 hours to allow for precipitation of the solid peptide.

Centrifugation was used to remove the precipitated peptides from the 40mL of ether solution. Previous centrifugation techniques used in the lab group called for forty 3 minute runs centrifuging 1mL at a time in an epindorf tube to remove all ether and obtain the solid peptide. I adjusted the procedure by centrifuging 10mL at a time in larger test tubes (4 times) for 5 minutes, then taking 2-4 mL of the ether decant to break up the pellet at the base of the test tube.



Finally 2-4 1mL 2 min centrifuge runs were used to transfer the peptide to an epindorf tube. This resulted in no noticeable loss of product, but saved over 90 minutes of centrifuge time per tetrapeptide. After air drying, the peptides were stored in the freezer to prevent degradation.

The tetrapeptides AAAK, AAAO, AAAB, AAAZ, AAKA, AAOA, AABA, AAZA, AKAA, AOAA, ABAA, AZAA, KAAA, OAAA, BAAA, ZAAA, YAGK, YAGO, YAGB, and YAGZ were synthesized (see Figures 2.3-2.7).

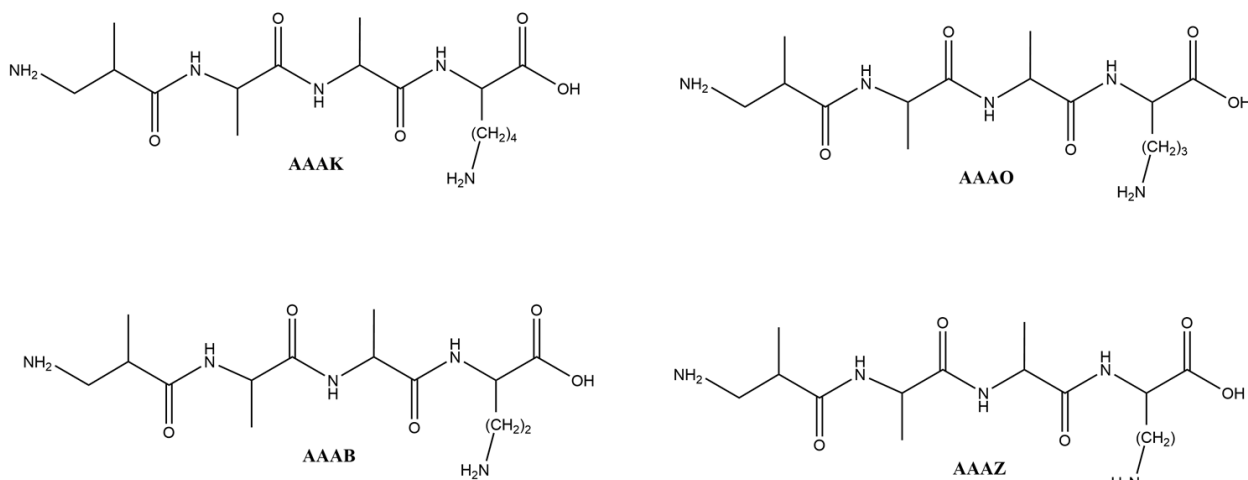


Figure 2.3 Tetrapeptides AAAK, AAAO, AAAB, and AAAZ

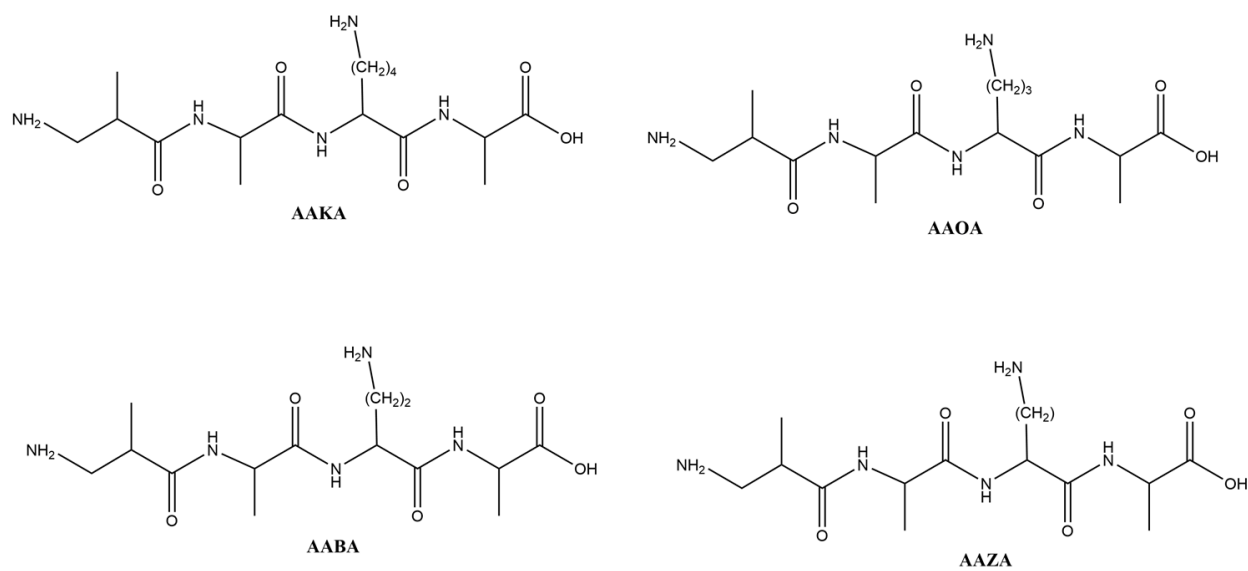


Figure 2.4 Tetrapeptides AAKA, AAOA, AABA, and AAZA

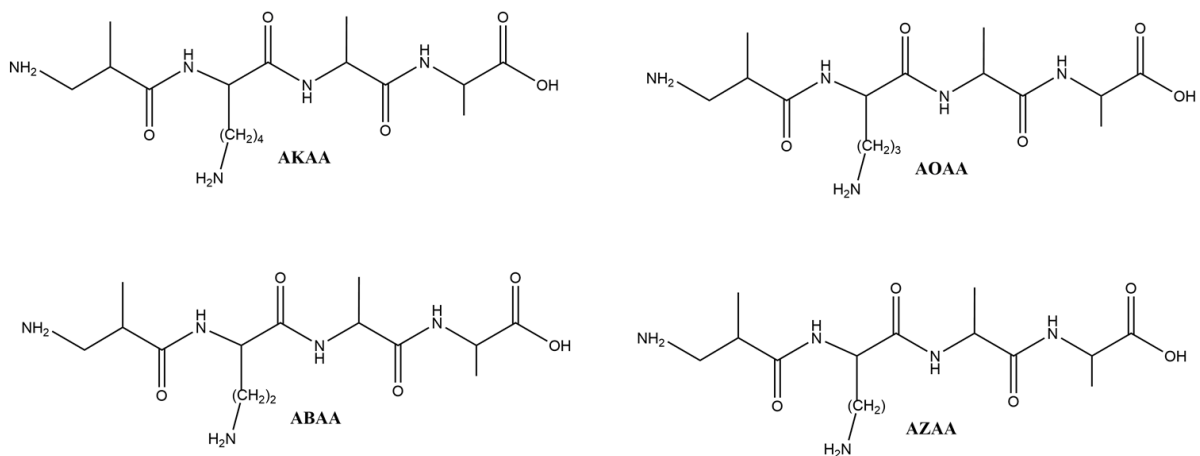


Figure 2.5 Tetrapeptides AKAA, AOAA, ABAA, and AZAA

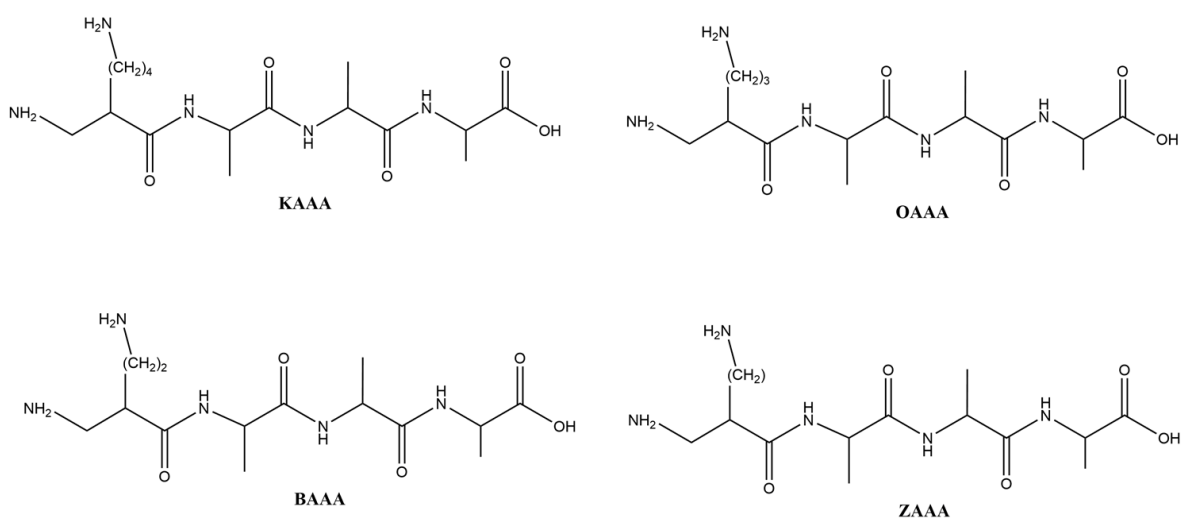


Figure 2.6 Tetrapeptides KAAA, OAAA, BAAA, and ZAAA

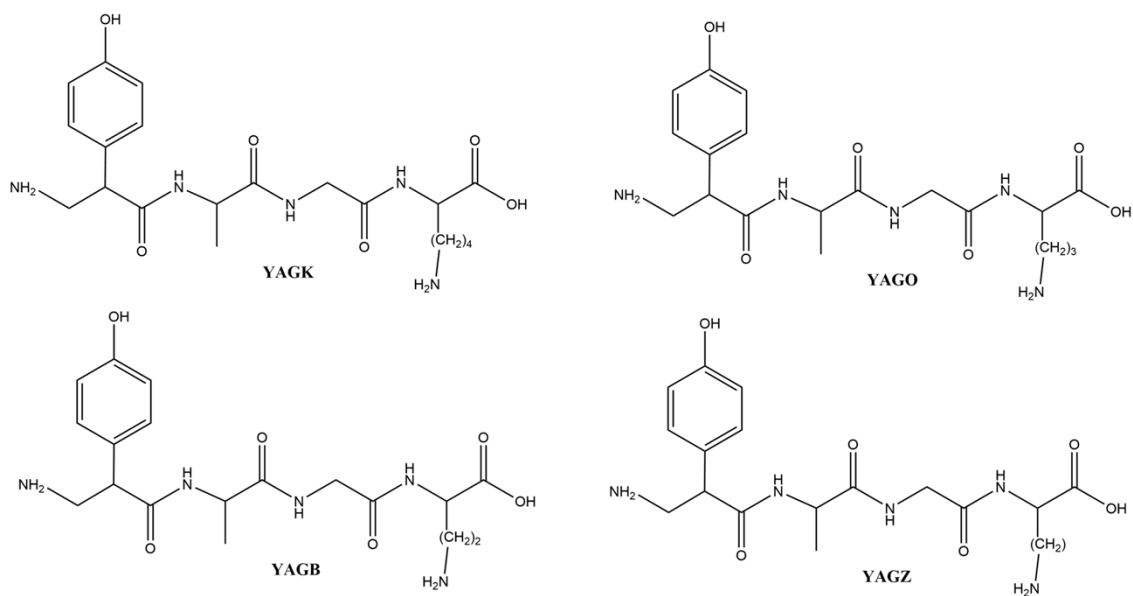


Figure 2.7 Tetrapeptides YAGK, YAGO, YAGB, YAGZ

## 2.2 ESI-Ion Trap Mass Spectrometry

Fragmentation studies of the 20 synthesized peptides were carried out in a Thermo LCQ Deca quadrupole ion trap mass spectrometer with an ESI sample introduction source (see Figure 2.8). Parent ion masses of the tetrapeptides were selected using the ion trap successfully with an isolation width of 3  $m/z$ . MS/MS studies were conducted in the ion trap using CID with He gas to obtain tetrapeptide fragmentation mass spectra. Activation ( $Q$ ) was set to 0.250, and activation time was 30ms for all runs. Data was collected at 0, 20, 25, 28, 30, 32, and 34 % CID. Additionally, an activation parameter scan (parent ion count vs. 0-100% CID) was conducted for each tetrapeptide. In some instances, product ions of the MS/MS fragmentation were isolated and fragmented with CID (MS/MS/MS).

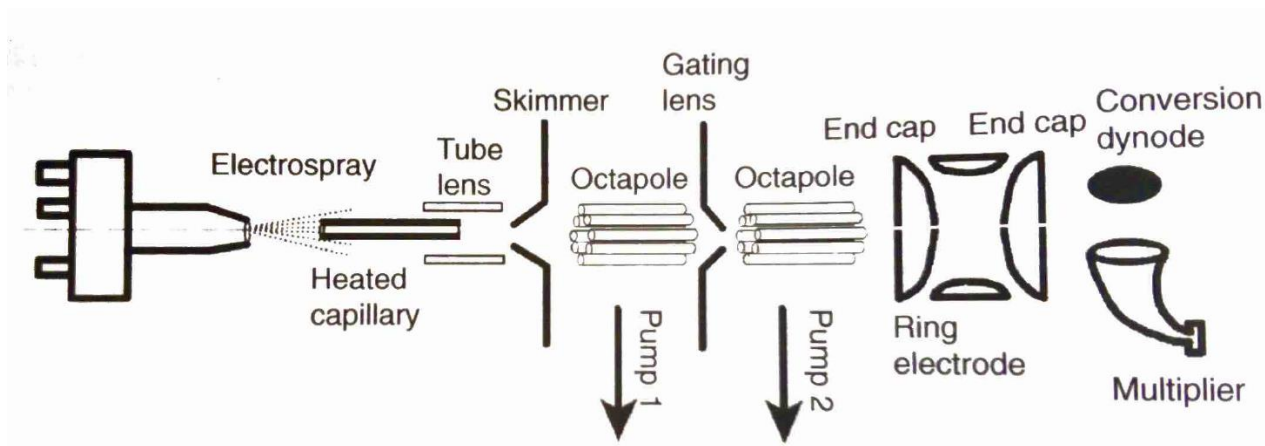


Figure 2.8 ESI ion trap mass spectrometer setup. Obtained from [55].

### 2.2.1 Sample Preparation and Runs

Tetrapeptide samples were prepared for MS/MS analysis by dissolving weighed amounts of solid peptide in 50:50 MeOH:H<sub>2</sub>O solution. Various dilutions were performed to obtain solutions on the order of approximately  $1 \times 10^{-5}$  M. The peptide solutions were protonated with the addition of 1% formic acid solution. Sample introduction into the ESI source was through a 500  $\mu$ L Hamilton gastight syringe at a flow rate of 12  $\mu$ L/min. Flush runs of 49.5% MeOH: 49.5% H<sub>2</sub>O: 1% formic acid solution were conducted before and after each tetrapeptide run.

## Chapter 3: Results and Discussion

The CID fragmentation spectra of the 20 synthesized tetrapeptides are presented and discussed below. Major product ion peaks have been identified and labeled on the spectra. For each of the tetrapeptides studies, an activation parameter scan (CID 0% - 100%) was conducted and data for % CID vs. parent normalized ion count was obtained (see Table 3.1 and 3.2).

Graphs of these activation parameter scans are provided in the Appendix.

### 3.1 Alanine-Alanine-Alanine-X

#### 3.1.1 AAAK

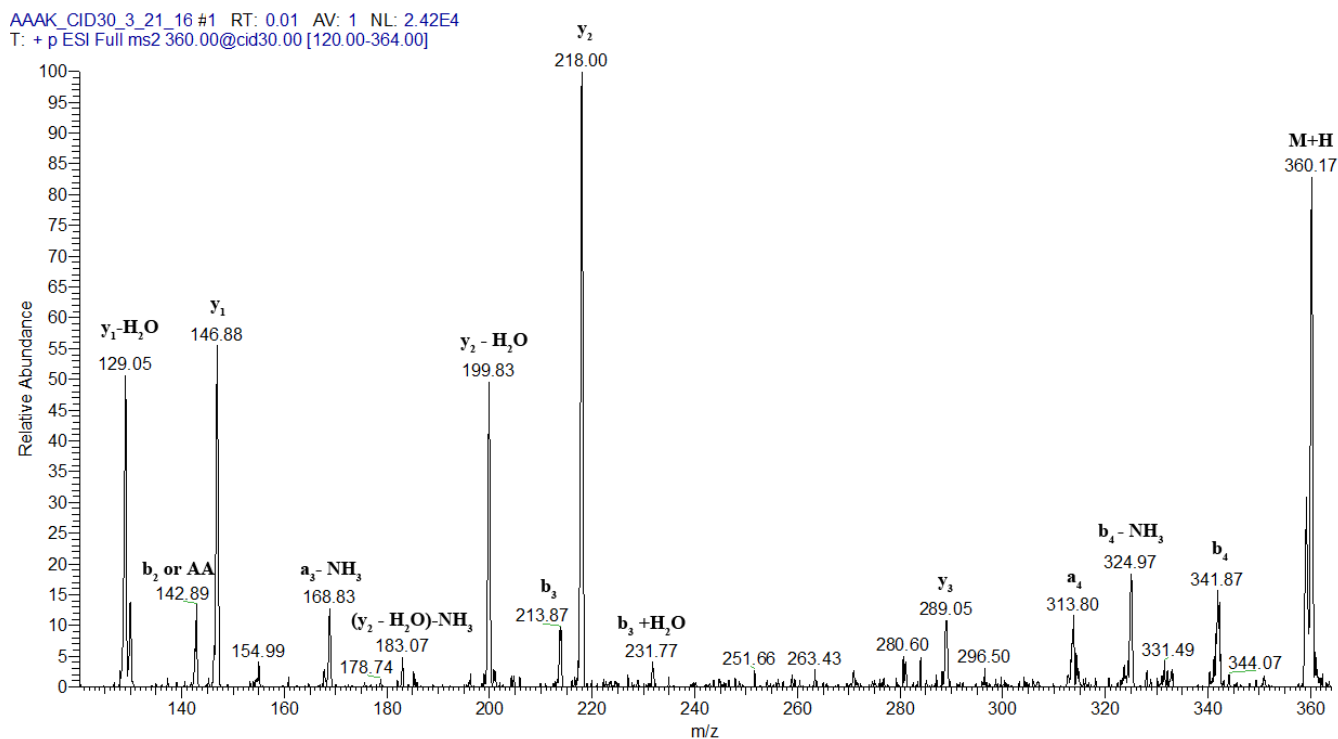


Figure 3.1 AAKK MS/MS Spectrum 30% CID

Figure 3.1 shows the MS/MS spectrum at 30% CID of AAKK with a parent mass, M+H, of 360. The base peak after fragmentation is the y<sub>2</sub> ion at m/z of 218. The formation of a b<sub>4</sub> ion at 342 represents a loss of water from the parent mass. The b<sub>4</sub> ion can subsequently lose NH<sub>3</sub> from

the lysine side chain giving a peak at 325. For AAAK, formation of  $y_n$  ions seems preferential to formation of  $b_n$  or  $a_n$  ions based off of the relative intensities of product ions. There also is significant  $a_3$ -NH<sub>3</sub> formation at a m/z of 169. This  $a_3$ -NH<sub>3</sub> ion is present for all AAAX and most AXAA (section 3.3) tetrapeptides, but is less commonly found in the fragmentation of AAXA (section 3.2) and XAAA (section 3.4) tetrapeptides.

Histerodt et al. 2008 found in a study of 200 dipeptides that CID fragmentation could yield b ions plus water when there was a basic PAA residue (Lys, Arg or His) at the N terminus of the dipeptide [56]. Based on the fragmentation, this would also form an intense  $y - H_2O$  complementary ion to the  $b + H_2O$  [56]. They noted that  $b + H_2O$  was not formed in significant amounts for any of the other dipeptide combinations [56]. The fragmentation spectrum of AAAK shows significant formation of  $b_3 + H_2O$  at a m/z of 232 and corresponding  $y_1 - H_2O$  ion at a m/z of 129. This suggests that for tetrapeptides  $b + H_2O$  formation is not limited to having a basic amino acid residue at the N terminus. This trend of  $b + H_2O$  ion formation is also observed for AAAO (section 3.1.2), YAGK (section 3.5.1), YAGO (section 3.5.2) and KAAA (section 3.4.1). Finally, the fact that the intensity of  $y_1 - H_2O$  product is greater than the  $b_3 + H_2O$  product for AAAK indicates that proton transfer preferentially favors the y ion upon peptide cleavage by this mechanism. This is likely due to the high proton affinity and gas phase basicity of lysine [51].

Macrocyclization is likely not occurring for AAAK. There are peaks on the spectrum that could represent macrocyclization, but the potential cyclized sequence product ions are isobaric with common peaks in the fragmentation spectrum. For example, a m/z of 200 could represent a macrocyclized scrambled  $b_2$  product peak for KAAA or AKAA. But a m/z of 200 can also be the  $y_2 - H_2O$  of AAAK, which is the same ion as a  $b_2$  from KAAA or AKAA. Therefore there is no

possibility of telling these two structures apart, and the presence of macrocyclization in AAAK cannot be supported.

AAAK had the second largest amount of CID (42%) required to fragment to 50 percent of initial ion count as compared to the other AAAX tetrapeptides (see Table 3.1 below). This is because there is the potential for intramolecular stabilization via hydrogen bonding between the lysine side chain and an oxygen or nitrogen atom in the peptide, which increases the overall stability of AAAK. To further probe the thermodynamics of AAAK, computational modeling with methods such as density functional theory calculations will be performed (outside the scope of this project).

<b>Tetrapeptide</b>	<b>% CID</b>
AAAK	42%
AAAO	35%
AAAB	48%
AAAZ	33%
AAKA	37%
AAOA	34%
AABA	33%
AAZA	31%
Akaa	37%
AOAA	31%
ABAA	33%
AZAA	32%
KAAA	37%
OAAA	31%
BAAA	35%
ZAAA	33%

Table 3.1 Approximate % CID required to fragment each peptide to 50 percent of parent ion count for AAAX, AAXA, AXAA, and XAAA tetrapeptides (see Appendix for graphs).

### 3.1.2 AAAO

AAAO\_CID30\_3\_21\_16 #1 RT: 0.00 AV: 1 NL: 3.36E6  
T: + p ESI Full ms2 346.00@cid30.00 [100.00-352.00]

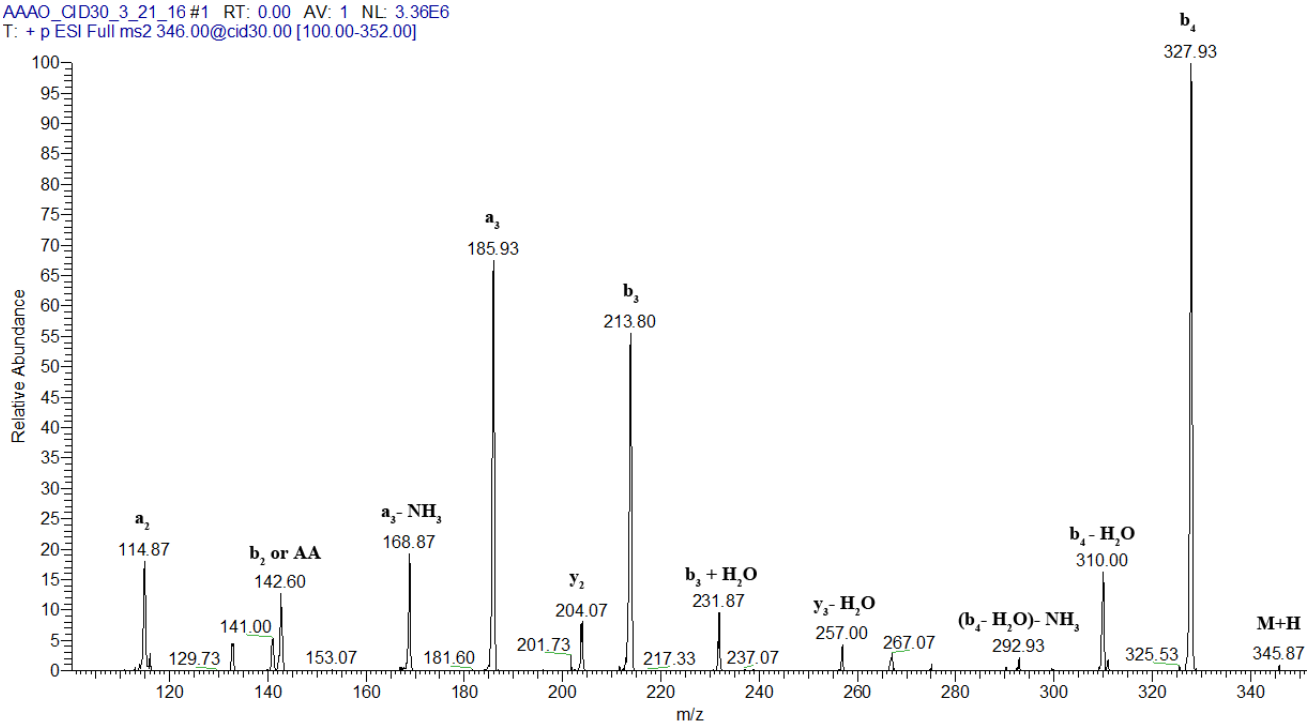


Figure 3.2 AAAO MS/MS Spectrum 30% CID

The base peak for the fragmentation of AAAO ( $M+H = 346$ ) is  $b_4$  ion formation (loss of  $H_2O$ ) at a  $m/z$  of 328. The  $b_4$  ion can lose water to form the peak at a  $m/z$  of 310, and subsequently lose  $NH_3$  to form a product peak at a  $m/z$  of 293. The loss of  $NH_3$  from the  $b_4$  ion is observed for AAK and AAAO, but not AAAB or AAAZ. This indicates that the  $NH_3$  is lost from the side chain of lysine or ornithine (not the N-terminus of the peptide) and that the  $b_4-NH_3$  formation is governed by the basicity of the amino acid residue at the C-terminus (DABA and DAPA lower basicity than lysine and ornithine). The intensities of  $b_n$  and  $a_n$  product ion formation are greater than  $y_n$  ion formation, which is opposite from the observed trend for AAK. The  $b_4$  ion formation is indicative of the ornithine effect: preferential C-terminal cleavage, in this case to lose water after ornithine side chain attack on the C-terminus forming a stable 6 membered lactam ring  $b_4$  ion. While the ornithine effect does predominate, there are still

a large number of other product ions formed when ornithine is at the C-terminus as compared to the CID spectra of AAOA, AOAA, and OAAA.

A significant  $b_3 + H_2O$  peak at a  $m/z$  of 232 was observed indicating that  $b + H_2O$  ions can be obtained with ornithine at the C-terminus of a tetrapeptide. The study by Histerodt et al. 2008 did not look at NPAAAs, however this finding for AAAO is further evidence that  $b + H_2O$  ions can be formed with a basic amino acid residue which is not located at the N-terminus. This effect is likely apparent for ornithine (in AAAO) as well as lysine (in AAAK) because of their similar basicities [51].

Additionally, there is  $a_3 - NH_3$  formation that is common to all AAAX and most AXAA tetrapeptides. The  $a_3$  ion consists of two alanine residues and one ornithine (or lysine, DABA, DAPA for other AAAX peptides) residue at the C-terminus. It seems that loss of  $NH_3$  from an  $a_n$  ion could be preferential when the basic amino acid residue is at the C-terminus. Although it is still present in AAXA and XAAA tetrapeptides, it is less universal to all the tetrapeptides within each set.

Similar to AAAK, there is no direct evidence for macrocyclization of AAAO. There are peaks in the fragmentation spectrum of AAAO that correspond to sequence scrambled fragments, but they cannot be differentiated with CID studies. For example, the  $b_3 + H_2O$  peak at a  $m/z$  of 232 is also the  $y_3$  peak for OAAA (again, these are the same ion). Finally, AAAO required the second least amount of % CID (%35) to fragment 50 percent of parent ion count (see Table 3.1). This could be because the ornithine effect is energetically favorable. Further studies with density functional theory calculations are recommended to assess AAAO thermodynamics.



### 3.1.3 AAAB

AAAB\_CID30\_3\_21\_16 #1 RT: 0.01 AV: 1 NL: 7.68E5  
T: + p ESI Full ms2 332.00@cid30.00 [90.00-338.00]

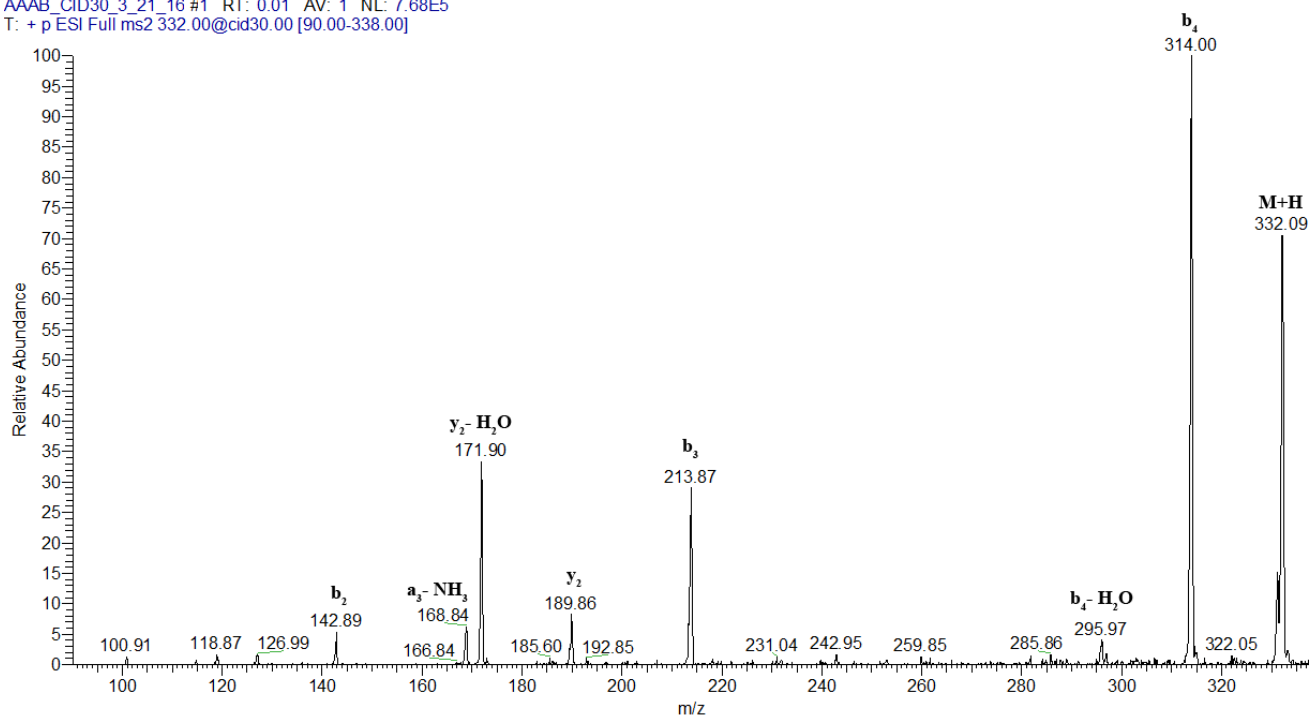


Figure 3.3 AAAB MS/MS Spectrum 30% CID

Similar to AAAO, the base peak for AAAB ( $M+H = 332$ ) fragmentation is the  $b_4$  ion at a  $m/z$  of 314. This  $b_4$  ion can subsequently lose  $H_2O$  to form a peak at a  $m/z$  of 296, but at a significantly smaller quantity than observed for AAAK or AAAO. No loss of  $NH_3$  from the  $b_4$  ion is observed. This suggests that the side chain length and/or basicity of Lys, Orn, DABA and DAPA plays a role in the ability for the tetrapeptide  $b_4$  ions to lose  $H_2O$  or  $NH_3$ . There is a peak at a  $m/z$  of 232 present that represents  $b_3 + H_2O$ , but it is very low intensity because of DABA's low basicity. There is also a strong  $y_2 - H_2O$  peak, which would suggest a  $b_2 + H_2O$  peak should be present at a  $m/z$  of 161. Because a peak at a  $m/z$  of 161 is not observed, there is preferential proton transfer to the  $y_2 - H_2O$  fragment due to the proton affinity of DABA (relative to alanine). An  $a_3 - NH_3$  fragment is also observed at a  $m/z$  of 169. Also, there are fewer fragment peaks observed for AAAB than for AAAO or AAAK.

Macrocyclization is not significant for AAAB. However, the peak at a  $m/z$  of 172 could represent the scrambled  $b_2$  of BAAA or ABAA or  $y_2 - H_2O$ , which are the same ion and cannot be differentiated. AAAB surprisingly required the highest CID energy (48%) to fragment to 50 percent of parent mass of the AAAX tetrapeptides (see Table 3.1). This indicates that the AAAB tetrapeptide stability, due to hydrogen bonding between the DABA side chain and oxygen or nitrogen on the peptide, is greater than that of AAAK. This increased stability could be because the side chain length of DABA is the optimal length and has the optimal orientation for hydrogen bonding. HDX experiments, which will be performed in the future, could elucidate this AAAB tetrapeptide hydrogen bonding stability. To further probe the energetics of AAAB, density functional theory computations are needed.

### 3.1.4 AAAZ

AAAZ\_CID30\_3\_21\_16 #1 RT: 0.00 AV: 1 NL: 1.24E6  
T: + p ESI Full ms2 318.00@cid30.00 [100.00-324.00]

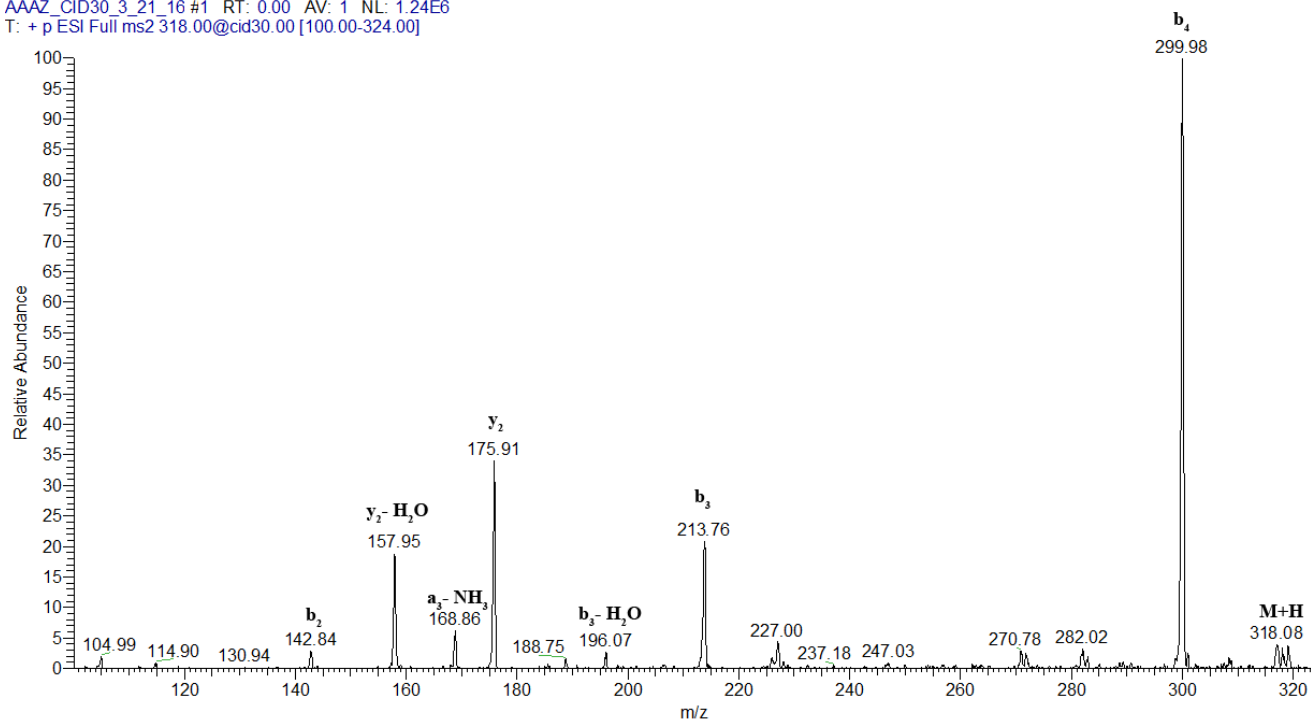


Figure 3.4 AAAZ MS/MS Spectrum 30% CID

The base peak for AAAZ ( $M+H = 318$ ) is the  $b_4$  ion at a  $m/z$  of 300. There is evidence of loss of  $H_2O$  to form a  $b_4-H_2O$  ion at a  $m/z$  of 282, but at a low relative intensity. Of the AAAX peptides, AAAZ shows no  $b + H_2O$  peak formation because DAPA has the lowest basicity as compared to lysine, ornithine, and DABA. However, there is a peak at a  $m/z$  of 158 that would indicate  $y_2-H_2O$  formation, and thus a complementary  $b + H_2O$  ion could be formed, but proton transfer favors the  $y_2-H_2O$  formation due to DABA's basicity. Additionally, an  $a_3-NH_3$  product ion at a  $m/z$  of 169 is observed.

There is no significant evidence for macrocyclization in AAAZ. Again there is a peak at a  $m/z$  of 158 that could represent the scrambled  $b_2$  of ZAAA or AZAA or  $y_2 - H_2O$ , but they are the same ion. AAAZ shows fewer fragments formed as compared to AAAK and AAAO, but it required the least amount of CID to fragment to 50 percent of parent ion count at %33 CID (see Table 3.1). This suggests strong preferential formation of the  $b_4$  ion, which is fairly stable. To further probe the energetics of AAAZ, density functional theory computations are planned.

## 3.2 Alanine-Alanine-X-Alanine

### 3.2.1 AAKA

AAKA\_CID30\_3\_17\_16 #1 RT: 0.00 AV: 1 NL: 2.41E5  
T: + p ESI Full ms2 360.00@cid30.00 [96.00-364.00]

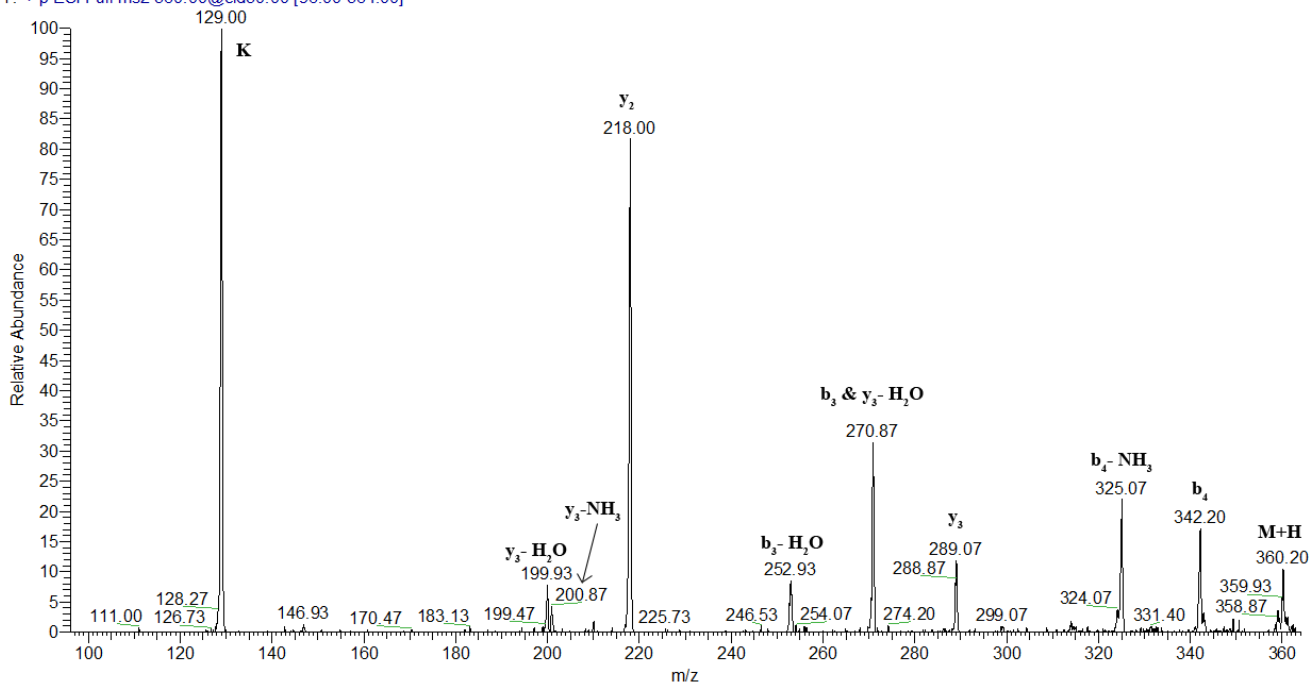


Figure 3.5 AAKA MS/MS Spectrum 30% CID

The base peak for the fragmentation of AAKA ( $M+H = 360$ ), is a K monomer internal fragment at a m/z of 129. The formation of monomeric internal fragments in AAXA and AXAA is observed at high relative abundances for lysine and ornithine containing tetrapeptides, at much lower relative abundances for DABA containing tetrapeptides, and is not observed for DAPA containing tetrapeptides. AAKA shows significant  $b_4$  formation and loss of  $NH_3$  and  $H_2O$  from  $b_4$  to form peaks at a m/z of 325 or 324 ( $NH_3$  is more intense). This is in contrast to what was observed for the AAAX peptides. AAKA may show the formation of  $b_n + H_2O$ , but it is difficult to tell because, for example,  $b_3 + H_2O$  shares a peak at a m/z of 289 with the AAKA  $y_3$  ion (they are the same).

The so called “lysine effect” which hypothesized preferential cleavage C-terminus to lysine residues may be present for AAKA. There is significant  $b_3$  ion formation for AAKA at a  $m/z$  of 271 (likely forming a stable caprolactam derivative  $b$  ion), but it is only at about 33% relative abundance to the base peak. Additionally, this peak can also represent  $y_3-H_2O$  (which is the same ion). The strong intensity of K monomer formation could indicate lysine is causing preferential C-terminal cleavage, but is not preferentially forming a  $b$  ion (cleavage at the both peptide bonds surrounding the lysine residue is occurring to form protonated K monomer instead).

Macrocyclization is not significant for AAKA. There are no peaks present which could be interpreted as a product ion of a scrambled sequence. AAKA requires approximately 37% CID to fragment to 50 percent of parent ion count, which is the largest of the AAXA tetrapeptides. Additionally, the AAXA tetrapeptides follow the trend  $AAKA > AAOA > AABA > AAZA$  showing a decreasing amount of CID required to fragment as the mass of the parent peptide decreases (see Table 3.1). To further probe the energetics of AAKA, density functional theory computations are planned.

### 3.2.2 AAOA

AAOA\_CID30\_3\_17\_16 #1 RT: 0.00 AV: 1 NL: 1.41E6  
T: + p ESI Full ms2 346.00@cid30.00 [100.00-350.00]

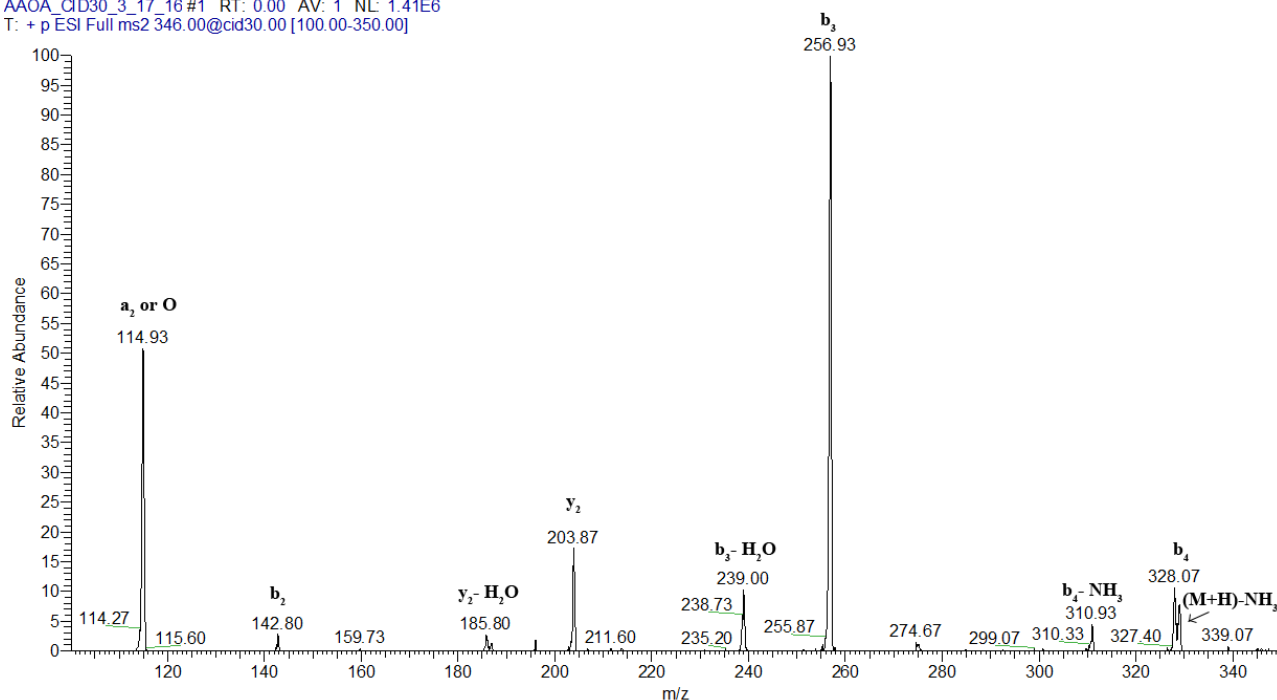


Figure 3.6 AAOA MS/MS Spectrum 30% CID

CID fragmentation of AAOA ( $M+H = 342$ ) shows a base peak at a  $m/z$  of 257 which is the  $b_3$  ion. This is confirmation of the ornithine effect showing C-terminal preferential cleavage after an ornithine residue. There is the formation of  $b_3 + H_2O$ , but not in significant amounts. The  $b_4$  and  $b_3$  ions can both lose  $H_2O$  and  $NH_3$ , but subsequent loss to form a  $(b-H_2O) - NH_3$  ion is not observed. This in contrast to the AAO tetrapeptide, which had  $b_4 - H_2O$  and  $(b_4 - H_2O) - NH_3$  peaks. There is also evidence of an ornithine monomer at a  $m/z$  of 115. However, this peak could also represent an  $a_2$  ion. This peak could be further analyzed with a technique such as IRMPD to determine its composition, or if both species are present, their relative abundances [36, 38].

There is no significant evidence of macrocyclization present for AAOA. The peak at a  $m/z$  of 186 could be the  $y_2 - H_2O$  ion or could represent the scrambled sequence  $b_2$  of AAOA or

OAAA, which are the same ion. However, based on the intensity of the peak, the formation of this ion would not be a significant detriment to peptide sequencing databases. AAOA follows the trend of AAKA > AAOA > AABA > AAZA in terms of percent CID required to fragment the peptide to 50 percent of parent ion peak requiring 34% CID (see Table 3.1). The energetics of AAOA will be further explored using computational studies.

### 3.2.3 AABA

AABA\_CID30\_3\_17\_16 #1 RT: 0.01 AV: 1 NL: 2.49E6  
T: + p ESI Full ms2 332.00@cid30.00 [90.00-336.00]

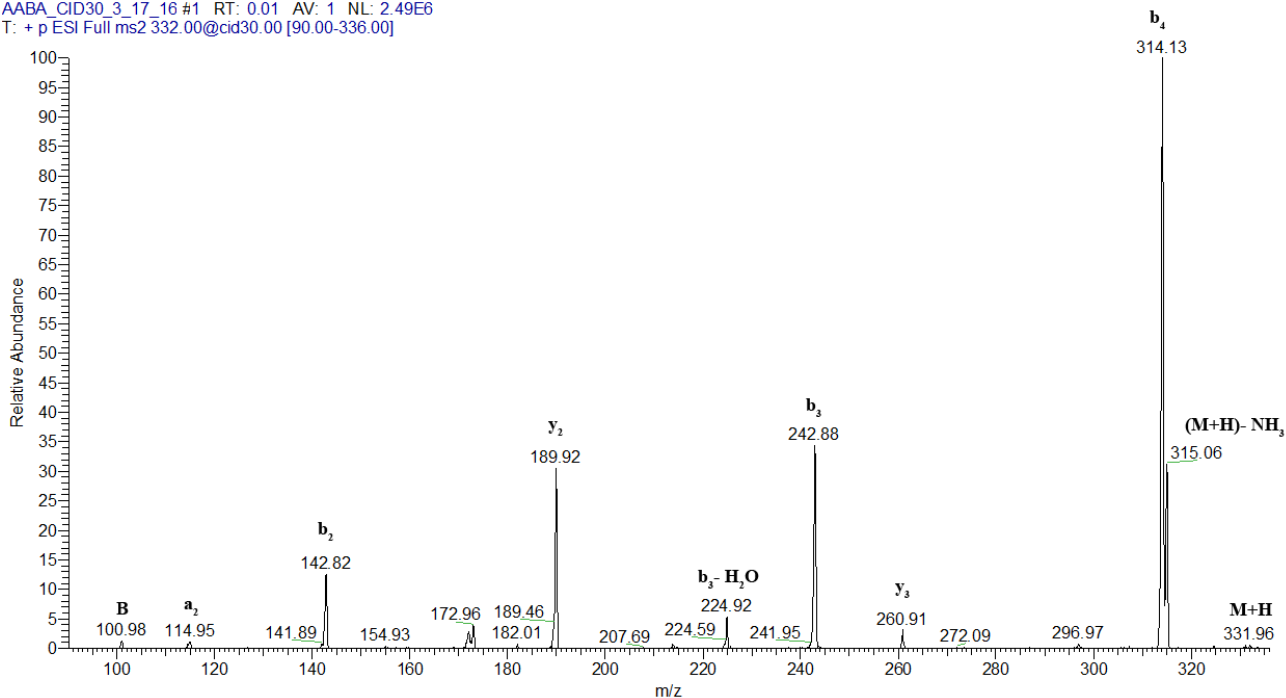


Figure 3.7 AABA MS/MS Spectrum 30% CID

The base peak in the fragmentation spectrum of AABA ( $M+H = 332$ ) is the  $b_4$  ion. There is also a significant amount of parent  $(M+H) - NH_3$  present. This strong  $b_4$  ion formation (which is also noted for AAZA in section 3.2.4 below) was not observed in AAKA or AAOA. DABA and DAPA are less basic than Lys or Orn, which affects the energetics of water leaving from the C-terminus to form a  $b_4$  ion. Surprisingly, the  $b_4$  ion does not lose  $H_2O$  or  $NH_3$  in any significant amount. Similarly, this trend is noticed for AAZA below (section 3.2.4). This suggests that the

lys and lys NPAA analog side chain lengths and position in the tetrapeptides plays a significant role in the loss of  $\text{NH}_3$  from b ions. There is little evidence for  $\text{b}_3 + \text{H}_2\text{O}$  ion formation in this spectrum; however the  $\text{y}_3$  peak at a m/z of 261 is isobaric with  $\text{b}_3 + \text{H}_2\text{O}$ . Finally, there is almost no  $\text{a}_n$  ion formation

There is no significant evidence for macrocyclization in AABA. AABA required 33% CID to fragment to obtain 50 percent of parent ion count, which is lower than that of AAKA and AAOA (see Table 3.1). Studying the energetics of AABA using computational calculations may provide insight into the reasons for strong  $\text{b}_4$  ion formation in AABA as opposed to what is observed for AAKA or AAOA.

### 3.2.4 AAZA

AAZA\_CID30\_3\_17\_16 #1 RT: 0.00 AV: 1 NL: 9.55E6  
T: + p ESI Full ms2 318.00@cid30.00 [100.00-326.00]

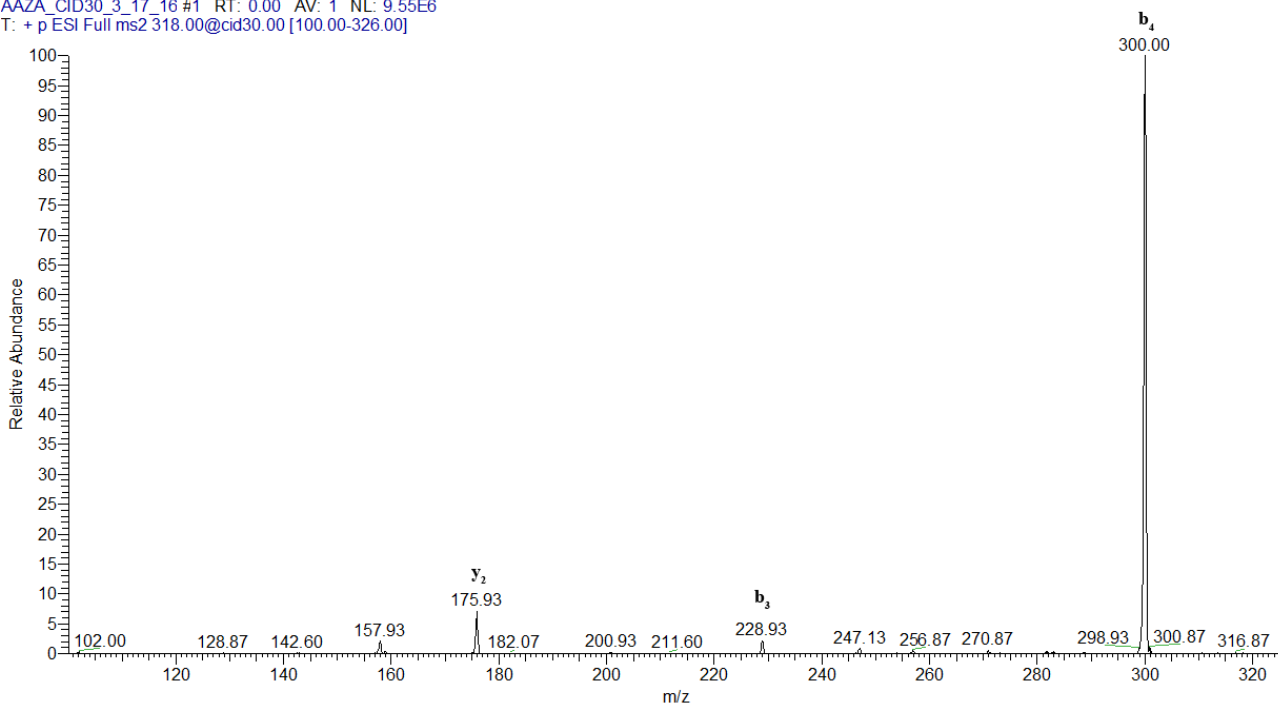


Figure 3.8 AAZA MS/MS Spectrum 30% CID



The base peak in the spectrum for AAZA ( $M+H = 318$ ) is the  $b_4$  ion at a  $m/z$  of 300. This peak is significantly more intense than any other peak in the spectrum. There is a  $y_2$  product ion at an  $m/z$  of 176 has a relative abundance of roughly 8% compared to  $b_4$ , and a  $b_3$  product ion at an  $m/z$  of 229 which has a relative abundance of about 3% compared to  $b_4$ . There is likely no formation of  $b + H_2O$  ions. There is a peak at a  $m/z$  of 247 that could be either  $b_3 + H_2O$  or a  $y_3$  peak. However, its low intensity compared to the base peak would make it irrelevant for peptide sequencing databases. There is almost no  $a_n$  ion formation.

There is no significant evidence for macrocyclization in AAZA. The AAZA tetrapeptide does not form many fragments as compared to the others in the AAXA set. AAZA requires the least amount (31%) of CID to fragment to 50 percent of parent ion count. This is in keeping with the trend noted for the AAXA peptides where  $AKA > AOA > ABA > AAZA$  for amount of CID required to fragment to 50 percent of parent ion count (see Table 3.1). Having a DAPA residue in the middle of a tetrapeptide results in a low number of fragment ions being formed (also see AZAA in section 3.3.4). Computational studies on the energetics of AAZA may help to elucidate why this trend occurs.

### 3.3 Alanine-X-Alanine-Alanine

#### 3.3.1 AKAA

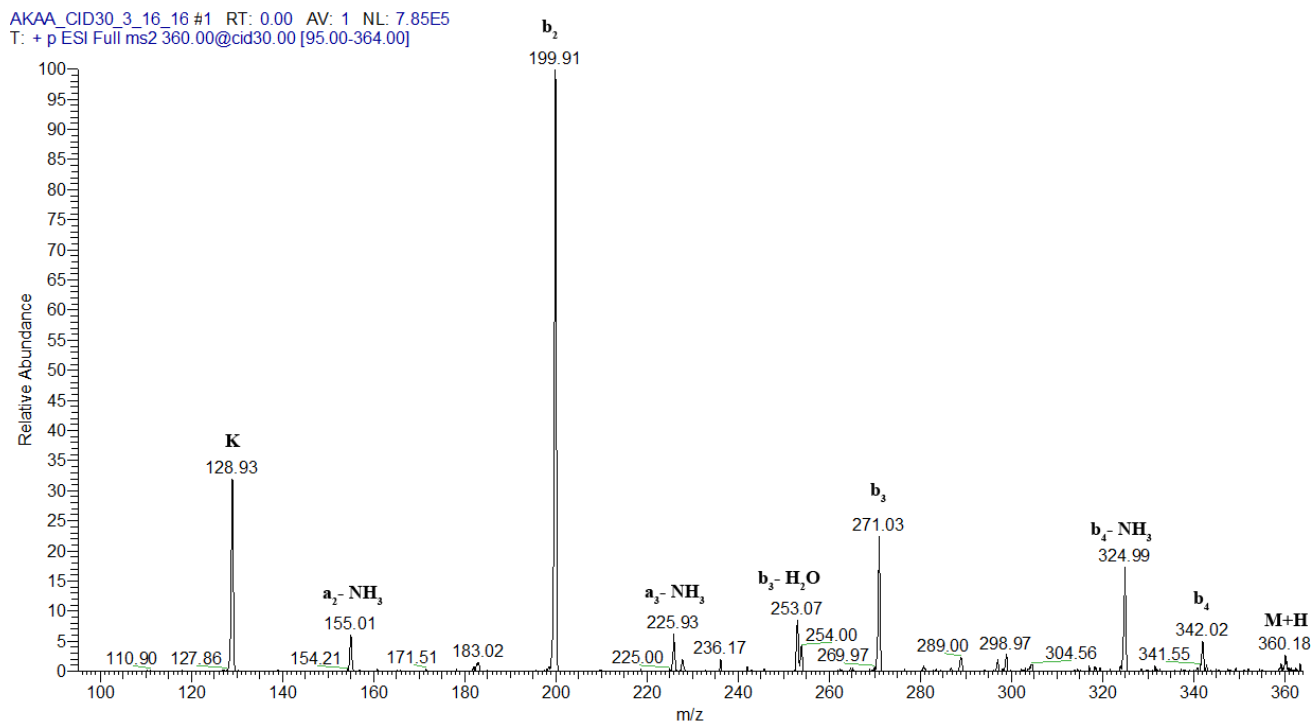


Figure 3.9 AKAA MS/MS Spectrum 30% CID

The base peak for fragmentation of AKAA ( $M+H = 360$ ) is a  $b_2$  ion at a  $m/z$  of 200. Based on the fragmentation pathways presented by Paizs and Suhai, this  $b_2$  ion is likely a 7 membered caprolactam ring derivative formed from a lysine effect [30]. There is loss of  $NH_3$  from the  $b_4$  ion, but no indication of loss of  $H_2O$ . There is a significant amount of K monomer formation for AKAA at a  $m/z$  of 129. This further indicates a preference for K monomer formation when the lysine residue is in the middle of the tetrapeptide (AKAA and AAKA). In contrast to the AAXA peptides, there is significant formation of  $a_n$  ions for AKAA. Loss of  $NH_3$  from the  $a_n$  ions of AKAA is prevalent, which is not unusual seeing as the a ion fragment

contains the lysine side chain (likely the site of  $\text{NH}_3$  loss). There is some indication of  $\text{b}_3 + \text{H}_2\text{O}$  formation at an  $m/z$  of 298, but this peak is isobaric with ACAA  $y_2$  ion.

There is little evidence for macrocyclization in the fragmentation spectrum of ACAA. The peak at an  $m/z$  of 129 which is most likely a K monomer could also be the  $\text{b}_1$  of KAAA. The ACAA tetrapeptide required approximately 37% CID to fragment to 50 percent of the parent ion count (see Table 3.1). The energetics of ACAA will be further studied using computational methods.

### 3.3.2 AOAA

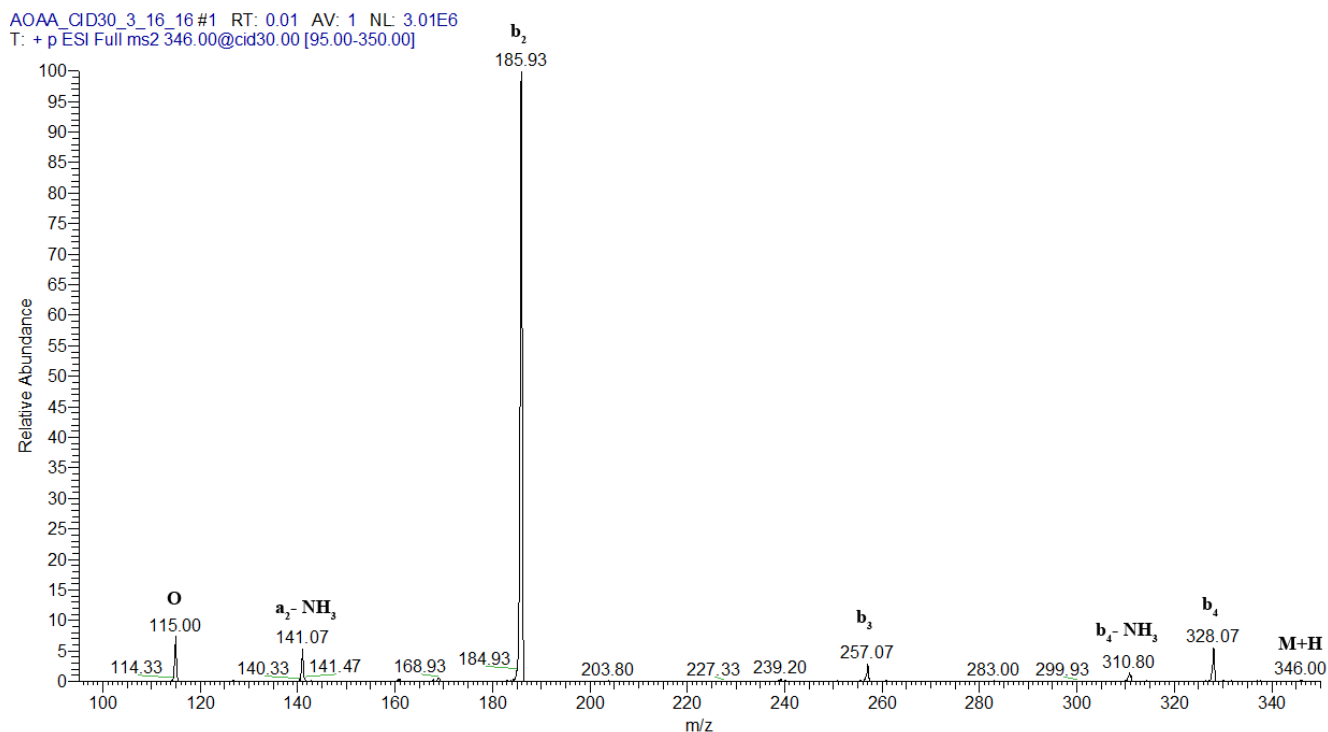


Figure 3.10 AOAA MS/MS Spectrum 30% CID

The base peak of the fragmentation spectrum for AOAA ( $M+H = 346$ ) is the  $b_2$  ion at a  $m/z$  of 186. This peak is indicative of the ornithine effect causing selective C-terminal cleavage via nucleophilic attack of the ornithine side chain on the adjacent carbonyl group to form the  $b_2$  ion in AOAA. Therefore, this  $b_2$  ion contains a 6 membered lactam derivative at its C-terminus. The  $b_4$  ion loses  $NH_3$  but does not lose  $H_2O$ . There is no  $y$  ion formation for AOAA. There is possibly  $b + H_2O$  ion formation, but at insignificant intensities. There is formation of an  $a_2 - NH_3$  (at a  $m/z$  of 141) which is found in the AKAA and ABAA spectra as well. There is a significant formation of an O monomer at an  $m/z$  of 115 (similar to monomer formation found in AKAA, AAKA and AAOA) indicating that monomer formation is preferred when the residue in middle of the tetrapeptide is lysine or ornithine as opposed to DABA or DAPA.

There is no significant evidence for macrocyclization in AOAA. AAOA required approximately 31% CID to fragment to 50 percent of parent ion count (see Table 3.1). With a similar trend as the AAAX set of tetrapeptides, this is less than the amount required to fragment ABAA and AKAA but more than that of AZAA. This lower % CID required may be due to the energetic preference of the ornithine effect occurring to form a  $b_2$  ion in AOAA. To further probe the energetics of AOAA, computational studies will be performed.

### 3.3.3 ABAA

ABAA\_CID30\_3\_16\_16 #1 RT: 0.01 AV: 1 NL: 6.08E5  
T: + p ESI Full ms2 332.00@cid30.00 [90.00-336.00]

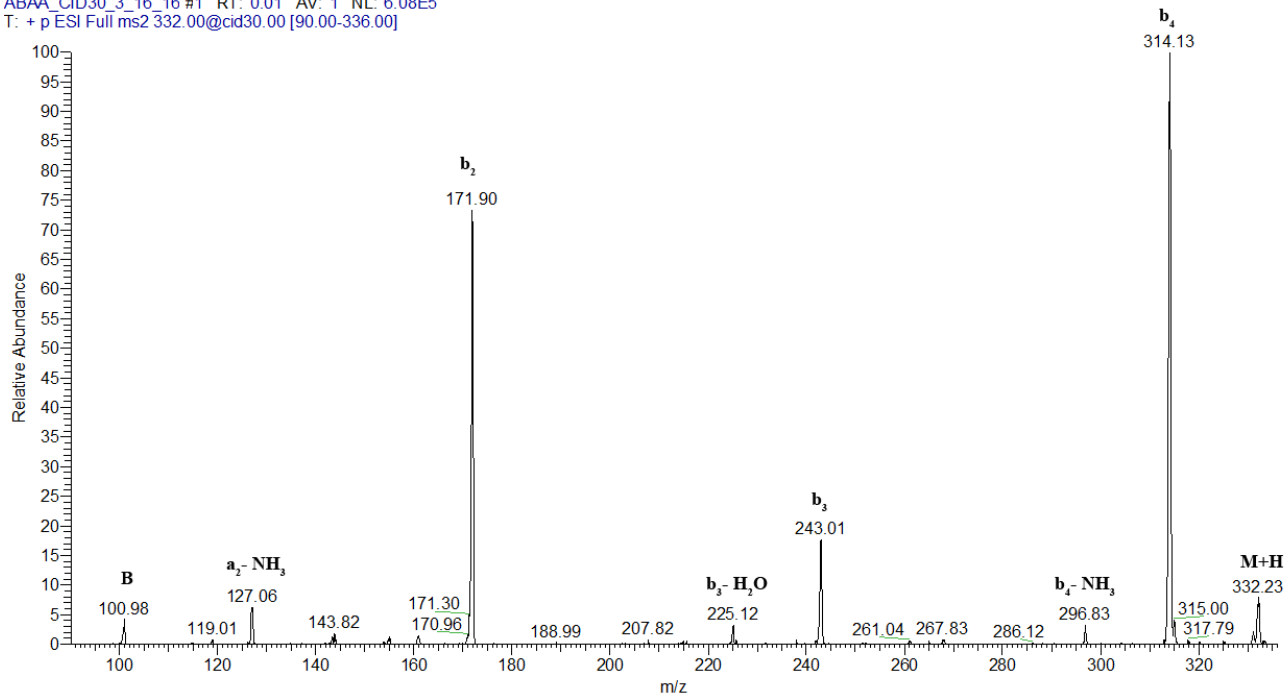


Figure 3.11 ABAA MS/MS Spectrum 30% CID

The base peak in the fragmentation spectrum for ABAA ( $M+H = 332$ ) is the  $b_4$  ion at a  $m/z$  of 314. There is a peak at 297 indicating  $b_4 - NH_3$  but no loss of  $H_2O$  is observed. Additionally, there is an intense peak at an  $m/z$  of 172 which is the formation of a  $b_2$  ion. It is possible that this  $b_2$  ion represents a selective cleavage indicative of a “DABA effect.” This would entail the DABA side chain attack of the adjacent c-terminal carbonyl to form a 5 membered lactam derivative b ion. However, this effect is not as intense for AABA, and nonexistent for BAAA (see section 3.4.3). It is clear that the DABA residue influences fragmentation intensities as opposed to random cleavage, potentially based on its location in the tetrapeptide chain. There is some indication of  $b_3 + H_2O$  formation at an  $m/z$  of 261, but this peak is isobaric with ABAA  $y_2$  ion. There is no significant evidence for macrocyclization in ABAA. ABAA required approximately 33% CID to fragment to 50 percent of parent ion count,

which is higher than that of AOAA (see Table 3.1). The energetics of ABAA could be further studied with computational studies

### 3.3.4 AZAA

AZAA\_CID30\_3\_16\_16 #1 RT: 0.00 AV: 1 NL: 1.42E7  
T: + p ESI Full ms2 318.00@cid30.00 [85.00-326.00]

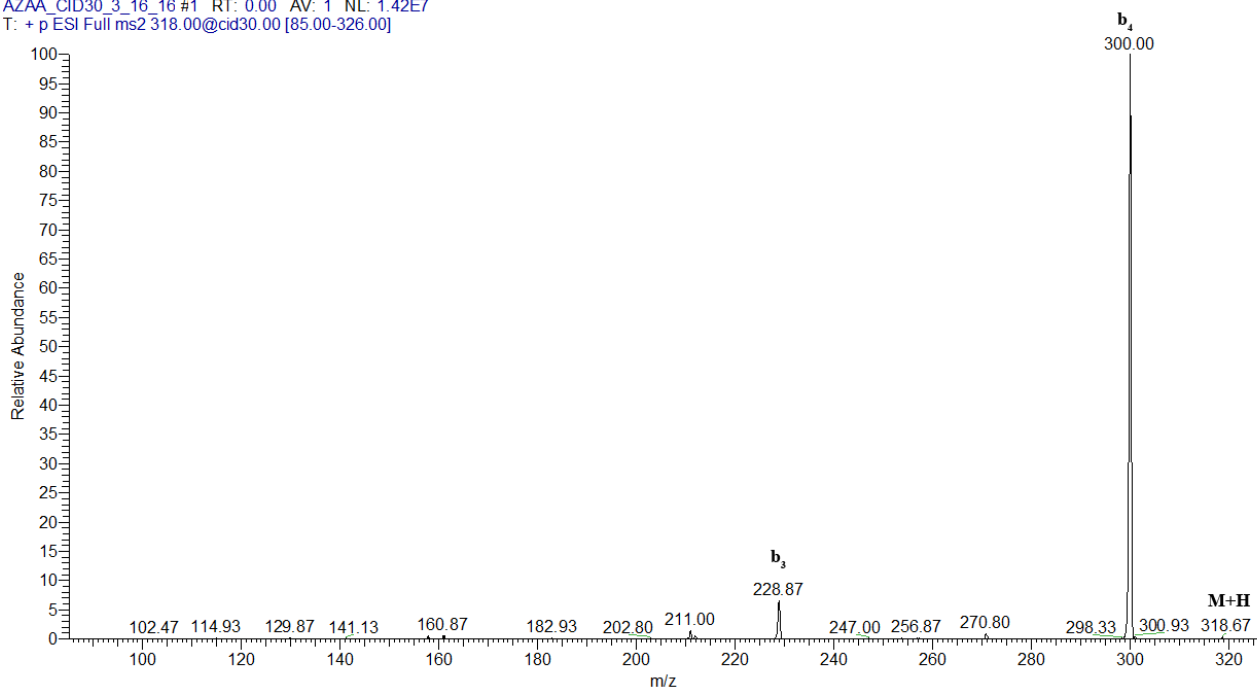


Figure 3.12 AZAA MS/MS Spectrum 30% CID

The fragmentation spectrum for AZAA ( $M+H = 318$ ) shows a base peak of  $b_4$  ion formation at a  $m/z$  of 300. This  $b_4$  ion is likely very stable due to its high relative abundance. There is no indication of  $b_4$  losing  $NH_3$  or  $H_2O$  in significant amounts. The only other major peak on this spectrum is the  $b_3$  ion at a  $m/z$  of 229. There is some indication of  $b_3 + H_2O$  formation at an  $m/z$  of 247, but this peak is isobaric with AZAA  $y_2$  ion. There is no significant evidence for macrocyclization in AZAA. AZAA required approximately 32% CID to fragment to 50 percent of parent ion count, which is the lowest of the AXAA tetrapeptides (see Table 3.1). The energetics of AZAA could be further studied with computational studies to elucidate why so few fragments are formed.

## 3.4 X-Alanine-Alanine-Alanine

### 3.4.1 KAAA

KAAA\_CID30\_2\_29\_16 #1 RT: 0.01 AV: 1 NL: 7.54E5  
T: + p ESI Full ms2 360.00@cid30.00 [95.00-364.00]

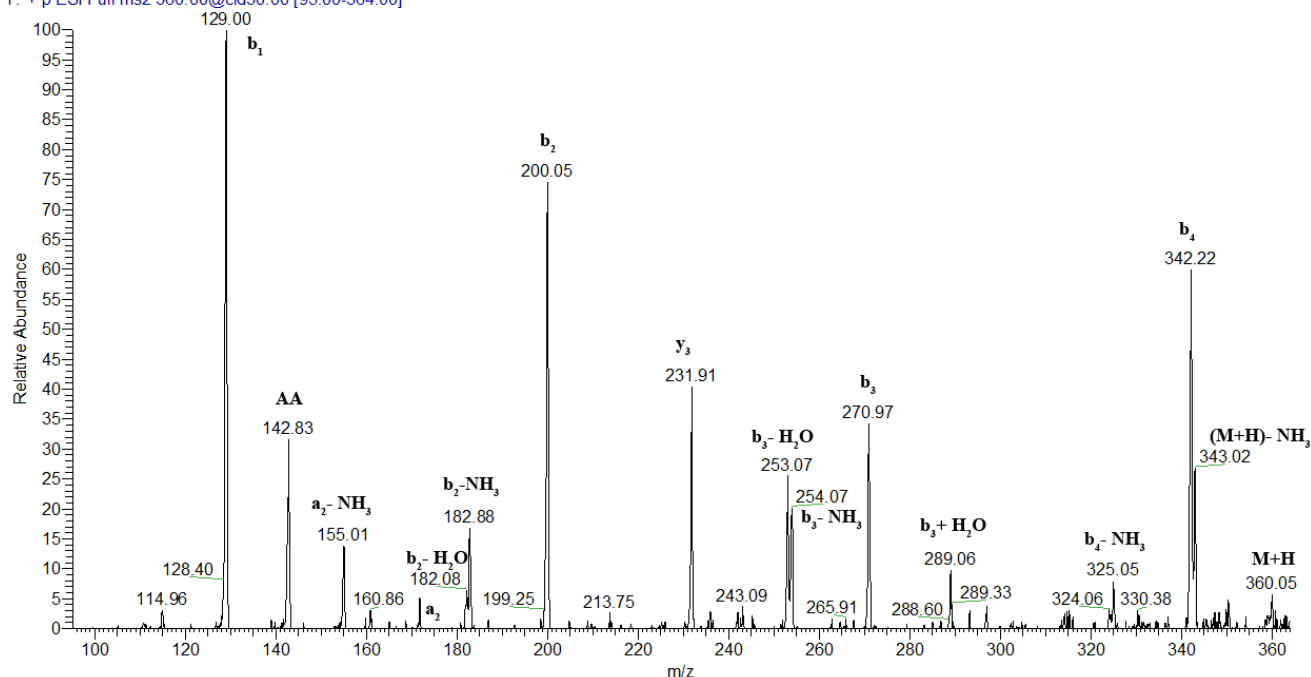


Figure 3.13 KAAA MS/MS Spectrum 30% CID

The fragmentation spectrum for KAAA ( $M+H = 360$ ) shows a base peak at a  $m/z$  of 129. This peak is most likely the formation of a  $b_1$  ion via lysine side chain attack of the adjacent C-terminal carbonyl to form a 7 membered caprolactam derivative b ion (a lysine effect). There is evidence of loss of  $NH_3$  directly from the parent ion to form a peak at a  $m/z$  of 342. The  $b_4$  ion at a  $m/z$  of 343 can lose  $NH_3$  or  $H_2O$  to form  $b_4 - NH_3$  at 325 or  $b_4 - H_2O$  at 324. Similarly, the  $b_3$  and  $b_2$  peaks and lose  $NH_3$  or  $H_2O$ . There is no significant presence of subsequent loss of  $NH_3$  from the  $b_4 - H_2O$  peak (or other b ion peaks). There is significant evidence of the formation of a  $b_3 + H_2O$  ion at a  $m/z$  of 289. This agrees with the results of Hiserodt et al. which found that N-

terminus basic amino acid residues allow for the formation of  $b + H_2O$  ions [56]. There is formation of an  $a_2 - NH_3$  ion, which is mainly observed for the AXAA and AAAX tetrapeptides. Finally, there is a significant intensity of AA dimer internal fragment formation. This dimer formation is noticed in significant quantities when the lysine residue is at the N or C terminus (KAAA or AAAK).

KAAA is likely not undergoing macrocyclization to create scrambled sequence product ion fragments. There are fragments observed that could represent a scrambled ion, but they all have potential explanations within KAAA fragmentation. For example, the low intensity peak at a  $m/z$  of 214 could represent the  $b_3$  of AAAK, but it is more likely to be a  $y_3 - H_2O$  peak. Additionally, the peak at  $m/z$  of 115 theoretically can only be explained as the  $a_2$  of AAKA or AAAK, but it is more likely to be a loss of 28 (CO) from the AA dimer at  $m/z$  of 143.

KAAA fragmentation resulted in the most number of product ion fragments of any of the XAAA tetrapeptides. KAAA required approximately 37% CID to fragment to 50 percent of parent ion count, which is the most %CID for the KAAA peptides (see Table 3.1). Computational studies to better understand the energetics of KAAA could elucidate why certain fragments are formed at higher intensities than others.



### 3.4.2 OAAA

OAAA\_CID30\_2\_29\_16 #1 RT: 0.00 AV: 1 NL: 1.62E6  
T: + p ESI Full ms2 346.00@cid30.00 [95.00-350.00]

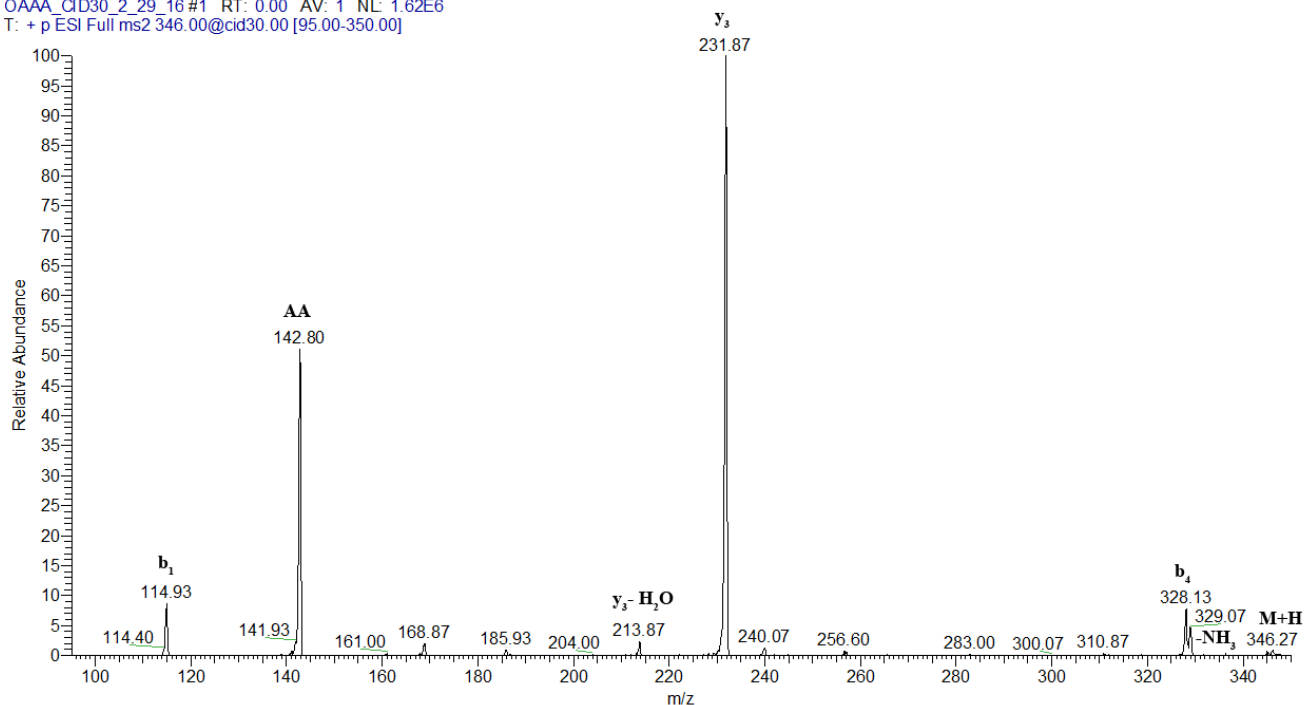


Figure 3.14 OAAA MS/MS Spectrum 30% CID

Fragmentation of OAAA ( $M+H = 346$ ) results in a base peak at a m/z of 232 which is the  $y_3$  ion. This is an interesting result as fragmentation of OAAA would be expected to have a  $b_1$  peak as the most abundant due to the ornithine effect. It seems that C-terminal cleavage after the ornithine residue is occurring in OAAA; however the proton is preferentially transferring to y fragment to form a  $y_3$  ion instead of a  $b_1$  ion. The study by McGee and McLuckey in 2013 which first described the ornithine effect only briefly mentioned that N-terminal ornithine effects mechanistically occur in the same manner as internal ornithine effects, and did not expand upon what occurs in fragmentation of peptides with ornithine at the N-terminus [47]. From this fragmentation spectrum of OAAA it is clear that there is an alternative mechanism or energetic preference at work for the ornithine effect.

There is a small amount of  $b_4$  formation at a  $m/z$  of 328, which can lose  $NH_3$  to form a peak at a  $m/z$  of 311, but in tiny amounts. There is a significant abundance of the AA dimer at a  $m/z$  of 143. This is likely because there is already C-terminal cleavage to the ornithine residue that is energetically favored, and the AA dimer can more easily be formed. There is no significant evidence of macrocyclization in OAAA. OAAA required 31% CID to fragment to 50 percent of parent ion count, which was the least of all XAAA tetrapeptides (see Table 3.1). This may be due to the energetic preference for C-terminal cleavage next to the ornithine residue. To further explore the energetics of OAAA, computational studies will be conducted.

### 3.4.3 BAAA

BAAA\_CID30\_2\_29\_16 #1 RT: 0.01 AV: 1 NL: 4.07E6  
T: + p ESI Full ms2 332.00@cid30.00 [90.00-336.00]

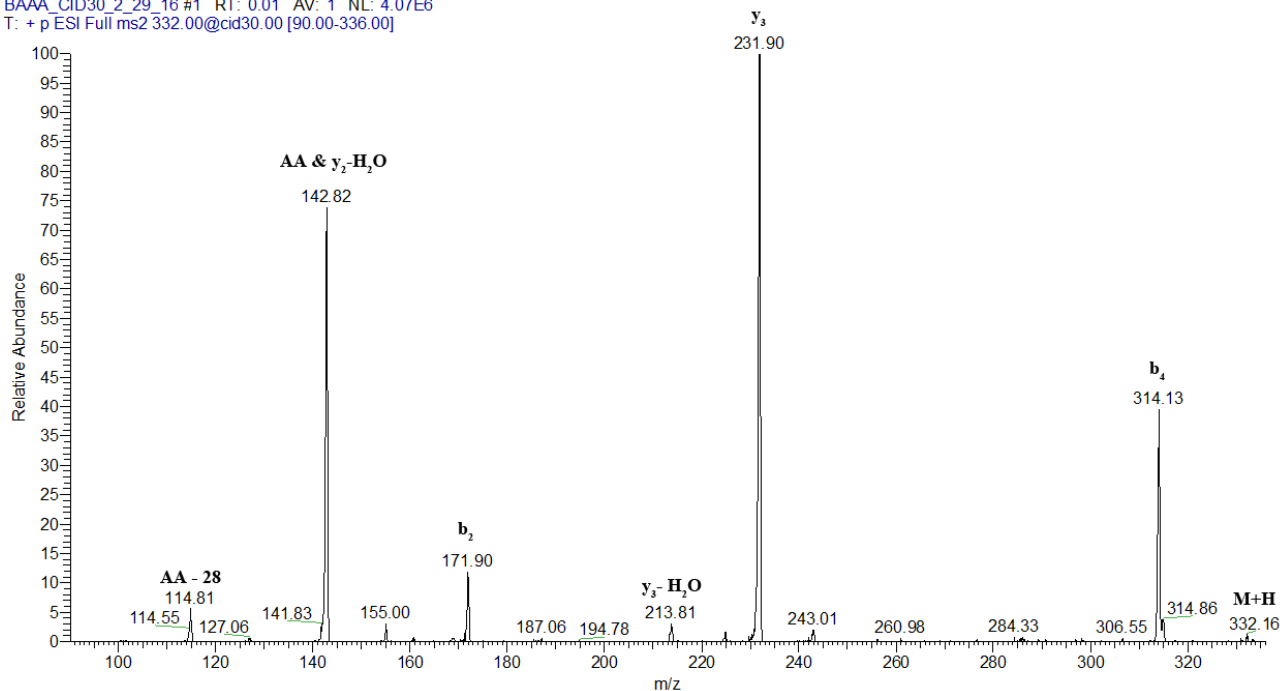


Figure 3.15 BAAA MS/MS Spectrum 30% CID

The fragmentation spectrum for BAAA ( $M+H = 332$ ) results in a base peak at a  $m/z$  of 232 which corresponds to a  $y_3$  ion. There is a significant formation of  $b_4$  ions at an  $m/z$  of 314, but loss of  $NH_3$  or  $H_2O$  from the  $b_4$  ion is not observed. There is no major evidence of  $b + H_2O$  formation for BAAA. There is a strong peak at a  $m/z$  of 143 that could represent an AA dimer internal fragment of a  $y_2 - H_2O$  product ion. Both could lose CO (28) to form the secondary fragment (AA-28) observed at a  $m/z$  of 115. There is no significant evidence of macrocyclization for BAAA. BAAA required 35% CID to fragment to 50 percent of parent ion count, which is greater than that of OAAA and BAAA (see Table 3.1). The energetics of BAAA will be further assessed using computational studies.

### 3.4.4 ZAAA

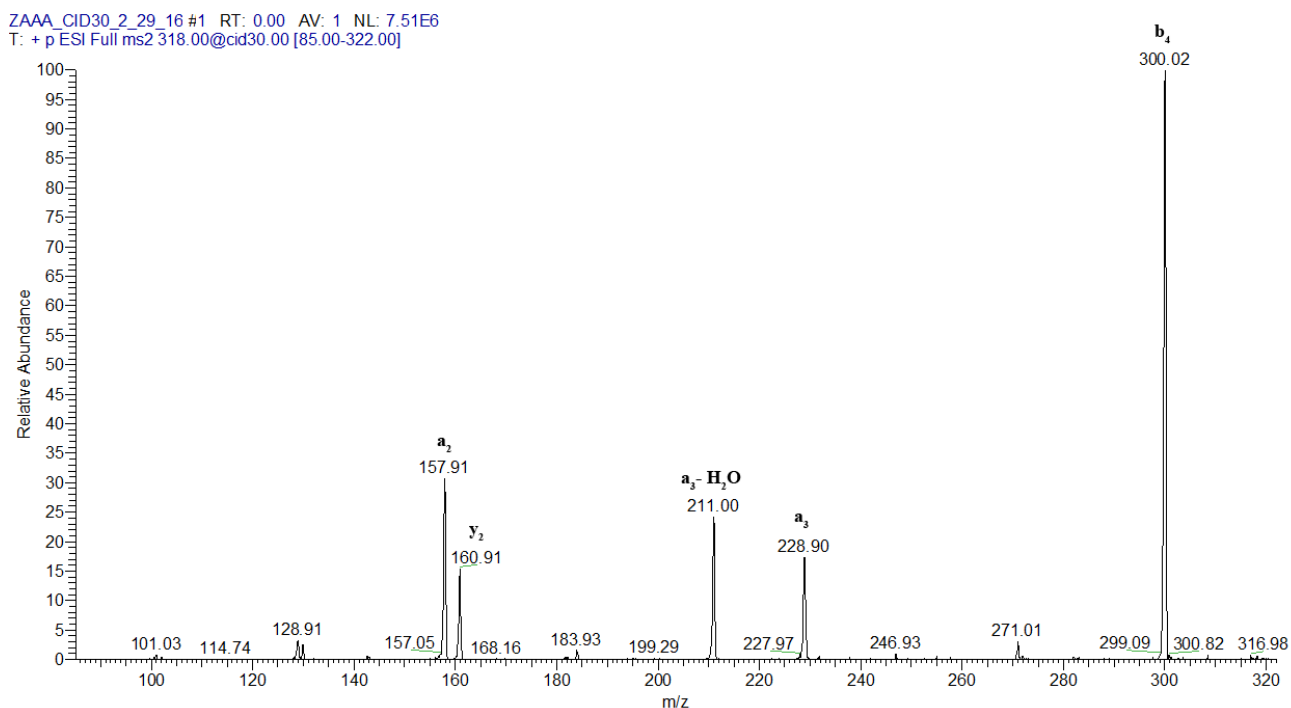


Figure 3.16 ZAAA MS/MS Spectrum 30% CID

The fragmentation spectrum for ZAAA ( $M+H = 318$ ) results in a base peak at a  $m/z$  of 300, which corresponds to the  $b_4$  ion. This  $b_4$  ion does not lose  $H_2O$  or  $NH_3$  in any significant

amount. Interestingly, a ion formation is more abundant than y ion formation for ZAAA. The a<sub>2</sub> ion at a m/z of 229 can lose H<sub>2</sub>O to form a peak at a m/z of 211. This is the only instance of a – H<sub>2</sub>O ion formation in all of the tetrapeptides studied. There is no significant evidence of macrocyclization for ZAAA. ZAAA required 33% CID to fragment to 50 percent of parent ion count, which is greater than that of OAAA, but less than that of KAAA and BAAA (see Table 3.1). ZAAA could be further assessed using computational studies to gain insight into the energetics and mechanisms of preferential a ion formation, and a<sub>2</sub> – H<sub>2</sub>O ion formation.

### **3.5 Tyrosine-Alanine-Glycine-X**

As was discussed in sections 3.1 through 3.4 above, macrocyclization does not seem to occur in any significant amounts for the 16 AAAX, AAXA, AXAA, and XAAA tetrapeptides. It is difficult to tell if macrocyclization is occurring in these 16 peptides because three of the same amino acid residues (alanine) were used. This created isobaric ions that could otherwise have differentiated a normal fragment from a scrambled fragment (especially difficult when ions with loss and addition of H<sub>2</sub>O are formed). The YAGX set of tetrapeptides were synthesized and studied to look for macrocyclization in tetrapeptides because they contain four different amino acids, and resemble the pentapeptides (YAGFL- NH<sub>2</sub>) used by Harrison et al. to discover sequence scrambling after CID [48]. Therefore there are more potential fragment ions at masses diagnostic of sequence scrambling than the previous 16 tetrapeptides. Additionally, spectral analysis of the YAGX tetrapeptides was conducted to look for similar trends to those found for the set of AAAX tetrapeptides.

### 3.5.1 YAGK

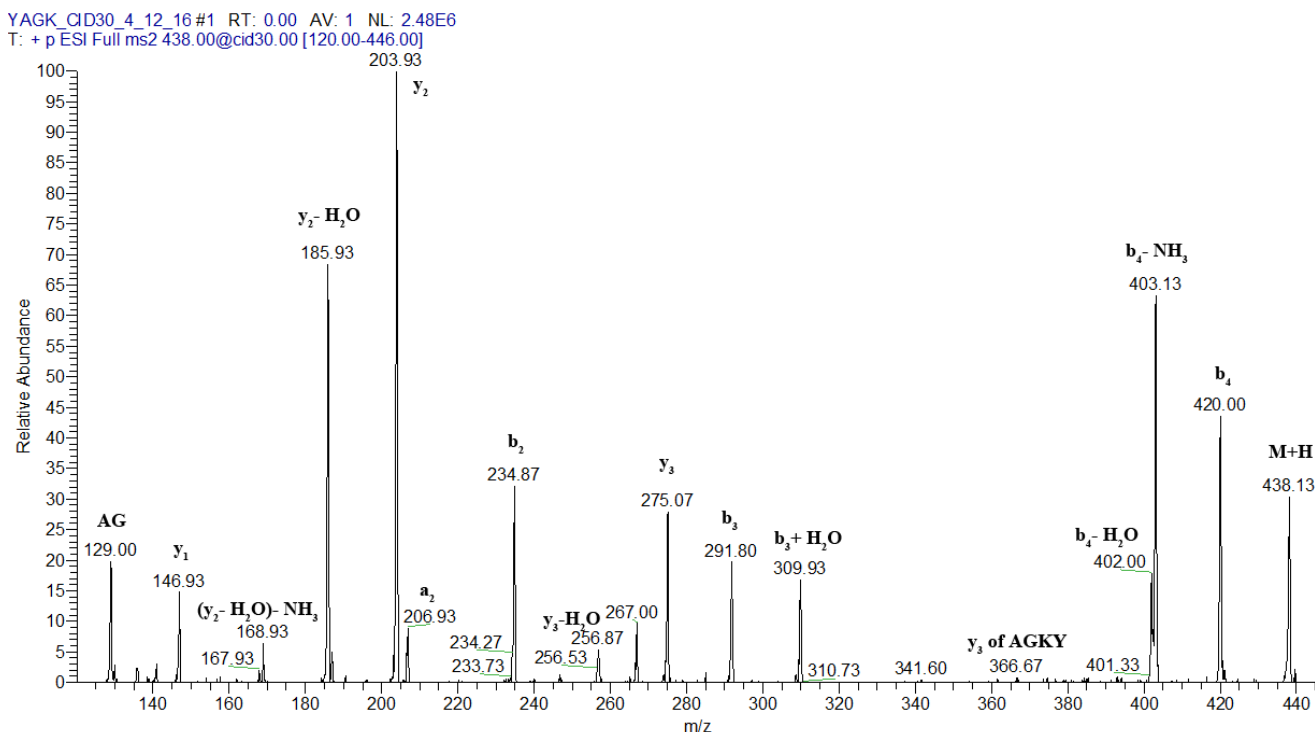


Figure 3.17 YAGK MS/MS Spectrum 30% CID

The fragmentation spectra of YAGK ( $M+H = 438$ ) shows a base peak at a  $m/z$  of 204, which corresponds to the  $y_2$  ion. There is significant formation of a  $b_4$  ion at a  $m/z$  of 420, which can lose  $NH_3$  or  $H_2O$  to form peaks at a  $m/z$  of 403 and 402, respectively. There is a strong abundance of the  $b_3 + H_2O$  peak at a  $m/z$  of 310, which is further evidence that  $b + H_2O$  ions can form with the lysine residue at the C-terminus of the peptide. From the base peak  $y_2$  ion, there is evidence of loss of  $H_2O$  to form a peak at a  $m/z$  of 186, and subsequent loss of  $NH_3$  to form a peak at a  $m/z$  of 169. There is also a significant presence of an AG internal fragment dimer at a  $m/z$  of 129.

Macrocyclization is occurring for YAGK, but in amounts that are not abundant enough to affect peptide sequencing and searching databases. For example, there is a very low intensity peak at a m/z of 367, which is the  $y_3$  of the tetrapeptide AGKY. This indicates macrocyclization is occurring, but in very small amounts. Additionally, the peak at a m/z of 186 could be a  $y_2 - H_2O$  ion, or it could be the sequence scrambled  $b_2$  ion of GKYA.

YAGK required approximately 47% CID to fragment to 50 percent of initial parent ion count (see Table 3.2). The YAGK spectrum shows a significantly larger number of fragments than the YAGO, YAGB and YAGZ spectra. Assessing the energetics of YAGK via computational studies may be beneficial for better characterizing the fragmentation patterns of YAGK.

Tetrapeptide	%CID
YAGK	47
YAGO	33
YAGB	31
YAGZ	33

Table 3.2 Approximate % CID required to fragment each peptide to 50 percent of parent ion count for YAGX tetrapeptides (see Appendix for graphs).

### 3.5.2 YAGO

YAGO\_CID\_30\_4\_12\_16 #1 RT: 0.00 AV: 1 NL: 3.64E6  
T: + p ESI Full ms2 424.00@cid30.00 [115.00-428.00]

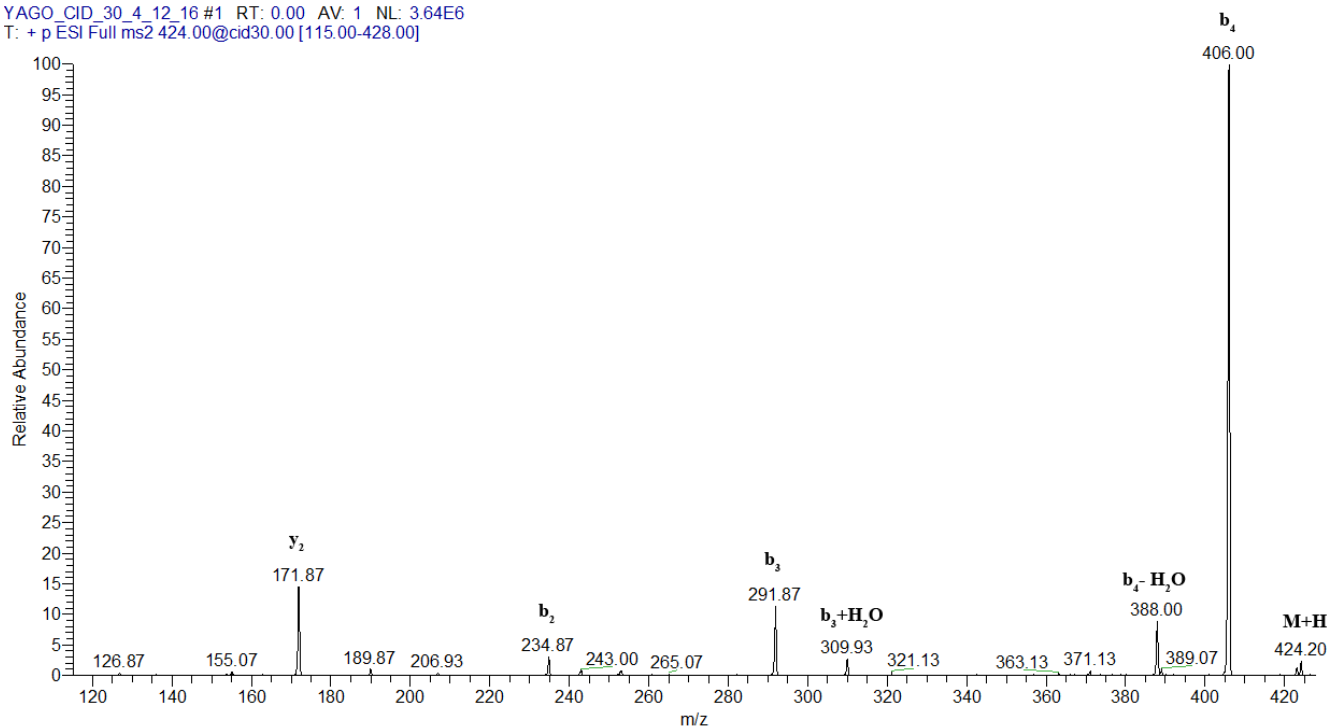


Figure 3.18 YAGO MS/MS Spectrum 30% CID

The fragmentation spectrum of YAGO ( $M+H = 424$ ) show the base peak at a  $m/z$  of 406, which corresponds to the  $b_4$  ion. This  $b_4$  ion is formed as a result of the ornithine effect, and therefore is the base peak as expected. The  $b_4$  peak can lose  $H_2O$  or  $NH_3$  to form peaks at a  $m/z$  of 388 and 389, respectively. The  $b_4 - H_2O$  can then subsequently lose  $NH_3$  to form a peak at a  $m/z$  of 371 (as was seen in AAAO), but this occurs at a very small intensity. There is evidence of a  $b_3 + H_2O$  peak at a  $m/z$  of 310. This is further evidence that ornithine is a basic enough residue to cause the formation of  $b + H_2O$  ions, and that  $b + H_2O$  ion formation can occur with the basic residue at the C-terminus (not just the N terminus as found in [56]). There is no a ion formation in the fragmentation of YAGO.

Macrocyclization is likely not occurring for YAGO. However, the  $b_3 + H_2O$  peak at a  $m/z$  of 310 could also be the sequence scrambled  $y_3$  ion of OYAG. YAGO required approximately 33% CID to fragment to 50 percent of initial parent ion count (see Table 3.2). Interestingly, YAGO required a similar amount of %CID as that of YAGB and YAGZ. The YAGO spectrum show very few fragments, especially as compared to YAGK. Assessing the energetics of YAGO via computational studies may be beneficial for better characterizing the fragmentation patterns of YAGO.

### 3.5.3 YAGB

YAGB\_CID\_30\_4\_12\_16#1 RT: 0.00 AV: 1 NL: 1.18E7  
T: + p ESI Full ms2 410.00@cid30.00 [110.00-416.00]

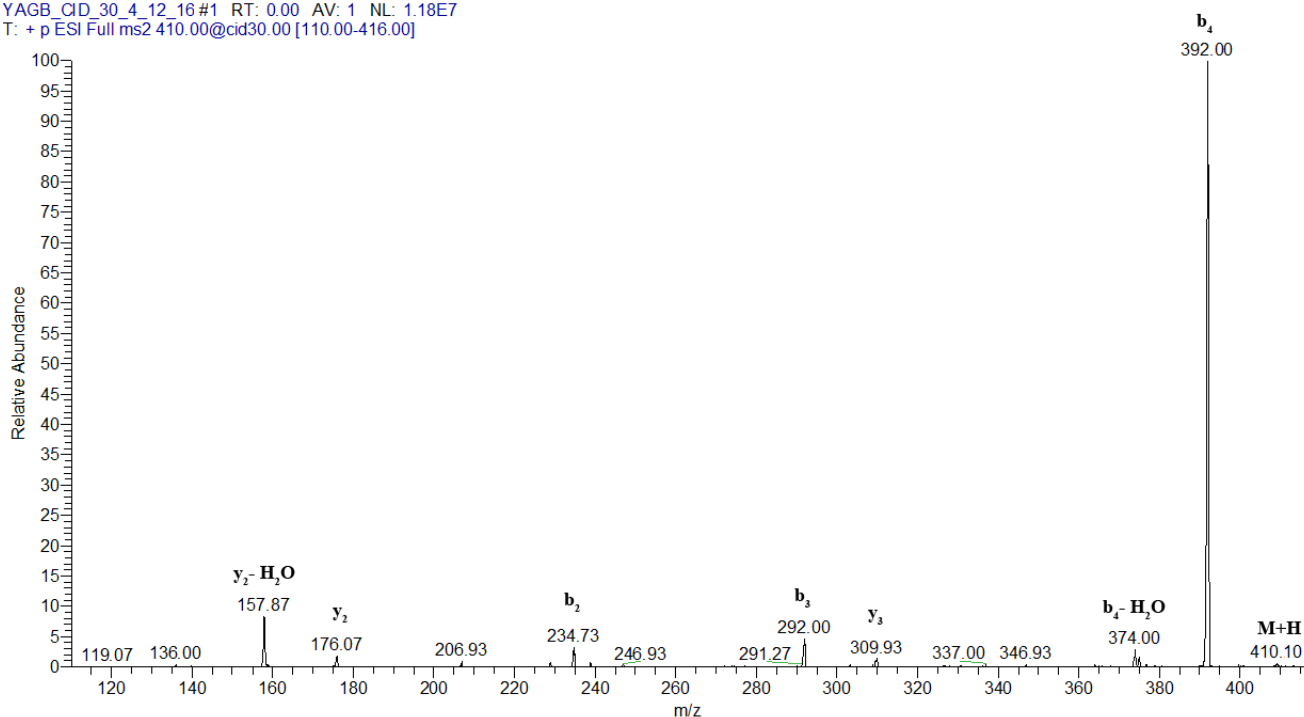


Figure 3.19 YAGB MS/MS Spectrum 30% CID

The fragmentation spectrum for YAGB ( $M+H = 410$ ) shows a  $b_4$  base peak at a  $m/z$  of 392, potentially indicative of a DABA effect. This high relative abundance  $b_4$  ion is also



observed for YAGO and YAGZ (and AAAO, AAAB, AAAZ). There is evidence of  $b_4$  losing  $H_2O$  form a peak at a  $m/z$  of 374. No  $a_n$  ions are observed in this fragmentation spectrum. There is potentially evidence of a  $b_3 + H_2O$  ion at a  $m/z$  of 310, but it is isobaric with the  $y_3$  ion. Macrocyclization is likely not occurring for YAGB. YAGB required approximately 31% CID to fragment to 50 percent of initial parent ion count (see Table 3.2). Again, YAGB required a similar % CID as that if YAGO and YAGZ. The YAGB spectrum show very few fragments, especially as compared to YAGK. Assessing the energetics of YAGB via computational studies may be beneficial for better characterizing the fragmentation patterns of YAGB.

### 3.5.4 YAGZ

YAGZ\_CID\_30\_4\_12\_16#1 RT: 0.01 AV: 1 NL: 2.06E6  
T: + p ESI Full ms2 396.00@cid30.00 [105.00-400.00]

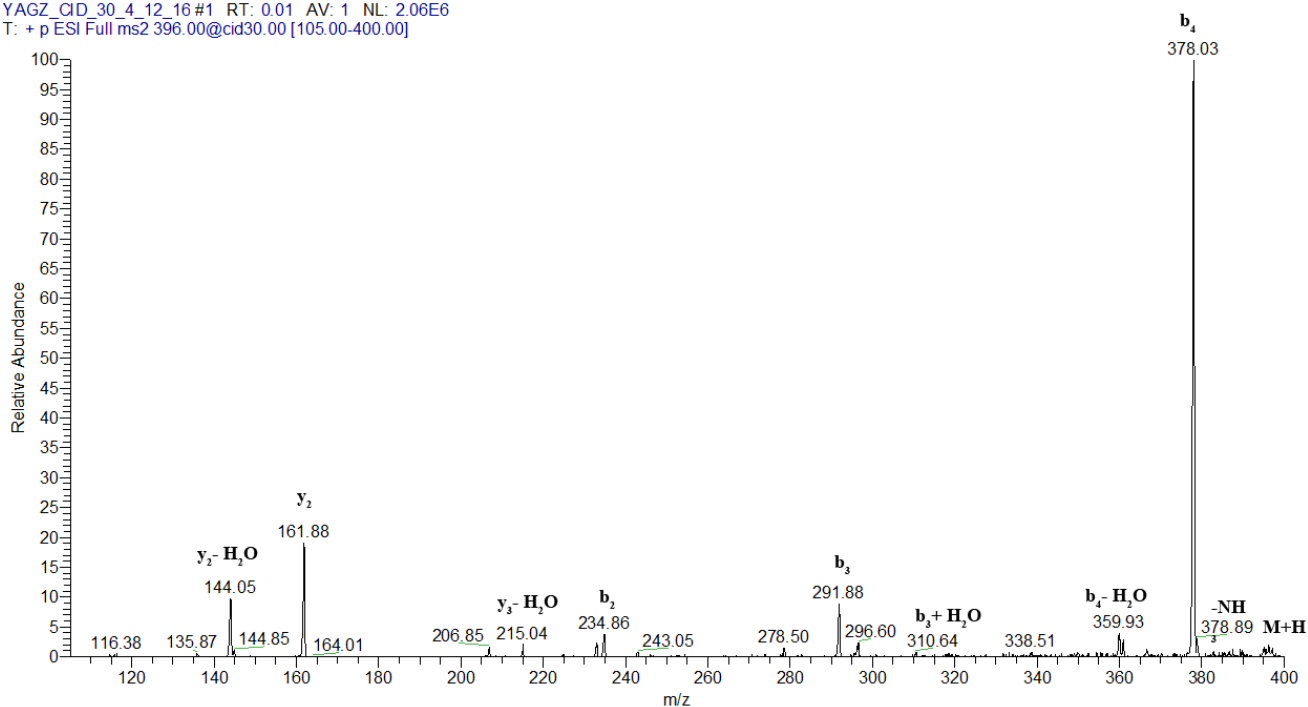


Figure 3.20 YAGZ MS/MS Spectrum 30% CID

The fragmentation spectrum for YAGZ ( $M+H = 396$ ) shows a  $b_4$  ion base peak at a  $m/z$  of 378, which was also observed for YAGO and YAGB. There is evidence of  $b_4$  losing  $H_2O$  or  $NH_3$  to form peaks at a  $m/z$  of 360 and 359, respectively. No  $a_n$  ions are observed in this

fragmentation spectrum. There is evidence of a  $b_3 + H_2O$  ion at a  $m/z$  of 310, but it is formed at a very low relative abundance. Macrocyclization is likely not occurring for YAGZ. YAGZ required approximately 33% CID to fragment to 50 percent of initial parent ion count (see Table 3.2). Again, YAGZ required a similar % CID as that of YAGO and YAGB. The YAGZ spectrum shows very few fragments, especially as compared to YAGK. Assessing the energetics of YAGZ via computational studies may be beneficial for better characterizing the fragmentation patterns of YAGZ.

### 3.6 Trends

In general, little or no macrocyclization leading to sequence scrambling is occurring for the 20 tetrapeptides studied. The AAAX, AAXA, AXAA, and XAAA tetrapeptides were not the most effective choice for studying sequence scrambling because of the large number of isobaric ions possible in the fragmentation spectra, especially when considering the  $y_n + H_2O$ ,  $y_n - H_2O$ ,  $b_n + H_2O$ , and  $b_n - H_2O$  ions formed. Macrocyclization and sequence scrambling was observed in YAGK (via formation of the  $y_3$  ion of AGKY). However, the intensity of the AGKY  $y_3$  peak is too low for it to be a significant detriment to peptide searching and sequencing algorithms.

There were interesting trends based on lysine or lysine analog positional variance within the tetrapeptides, and based on the differing basicities of lysine and its analogs. The AAAO and YAGO tetrapeptides showed a prevalence of  $(b_n - H_2O) - NH_3$  ion formation, whereas the other 18 tetrapeptides did not. DABA and DAPA containing tetrapeptides had minimal  $b_4 - NH_3$  formation as compared to lysine and ornithine containing tetrapeptides. The lower basicities or shorter side chain lengths of DABA and DAPA as compared to lysine or ornithine influences the absence of losses from  $b_n$  ions. Additionally,  $a_n$  ions were the least prevalent for the AAXA and XAAA tetrapeptides. The AXAA and AAAX tetrapeptides all showed  $a_2 - NH_3$  ion formation while only KAAA of the XAAA set and none of the AAXA tetrapeptides showed  $a_2 - NH_3$  formation.

For dipeptides,  $b_n + H_2O$  formation is favored when a basic amino acid residue is on the N-terminus [56]. For the tetrapeptides studied, only KAAA of the XAAA set shows  $b_n + H_2O$  formation. The AAAX AAAO, YAGK, and YAGO tetrapeptides show the formation of  $b_n + H_2O$  formation when fragmented, whereas the others do not. This indicates that  $b_n + H_2O$  ion

formation may also be preferential when the basic residue is at the C-terminus in tetrapeptides. In addition to positional variance, the lack of  $b_n + H_2O$  formation for OAAA, BAAA, ZAAA, AAAB, and AAAZ shows that the basicity of the amino acid residue plays a role (less likely to occur with lower basicity residue). There is intense K and O monomer formation for AAKA, AKAA, AAOA, and AOAA, but minimal or no monomer formation for DABA and DAPA containing tetrapeptides. The mechanism of fragmentation to form these monomers is likely dictated by the length of the amino acid side chain, and/or the basicity of the amino acid residues (DABA and DAPA are shorter and less basic than lysine and ornithine).

Based on number of peaks and CID data (see Table 3.1), it is apparent that positional variance of lysine and its NPAA analogs affects the stabilities of the AAAX, AAXA, AXAA and XAAA tetrapeptides. For example, AAAB required the most % CID to fragment to 50 percent of parent ion count of the AAAX tetrapeptides. This indicates that the intramolecular stabilization via hydrogen bonding between the DABA side chain and oxygen or nitrogen atoms in the peptide backbone is stronger than that of lysine, ornithine, or DAPA (for AAAX). This is likely due to DABA having an optimal side chain length for hydrogen bonding when the X residue is at the C-terminus.

Finally, the presence of the ornithine effect was confirmed in all ornithine containing tetrapeptides. There is some evidence to suggest a lysine effect, which causes preferential C-terminal cleavage, exists (see AAKA, AKAA, and KAAA) but it is not as marked as the ornithine preferential cleavage. Additionally, ABAA fragmentation shows an abundant  $b_2$  product ion peak, potentially indicating that DABA can cause preferential C-terminal cleavage under certain conditions.

## Chapter 4: Conclusions

This project used ESI- ion trap mass spectrometry with CID to fragment the lysine and lysine analog containing tetrapeptides AAAX, AAXA, AXAA, XAAA and YAGX (X = Lys, Orn, DABA, or DAPA). Results show that no macrocyclization leading to sequence scrambling occurs in significant amounts for these tetrapeptides. However, it was difficult to discern if macrocyclization is occurring in the AAAX, AAXA, AXAA, and XAAA tetrapeptides due to isobaric fragment ions between original and scrambled tetrapeptides. The presence of macrocyclic sequence scrambling in YAGK at a very small abundance, but not in YAGO, YAGB, or YAGZ may indicate that sequence scrambling is possible, but not energetically favorable for tetrapeptide fragmentation using CID.

The fragmentation studies did find significant trends in fragmentation patterns of the 20 tetrapeptides based on positional variance and/or basicities of the lysine, ornithine, DABA, and DAPA residues. The formation of  $b_n + H_2O$  ions when basic amino acid residues are at the C-terminus is possible in tetrapeptides (as opposed to findings for dipeptides [56]). Monomer fragment formation was observed for lysine and ornithine containing tetrapeptides at a significantly higher relative abundance than in DABA and DAPA containing tetrapeptides, due to the smaller side chain length and/or lower basicity of DABA and DAPA as compared to lysine and ornithine. The DABA and DAPA containing tetrapeptides had minimal  $b_4-NH_3$  formation as compared to lysine and ornithine containing tetrapeptides. The ornithine effect was confirmed, and a possible lysine effect and DABA effect were observed. Additionally, the CID activation scan data shows differences in the stabilities of the tetrapeptides based on the amino acid residue present.

Future work will involve the synthesis and fragmentation of YAXG, YXAG, and XYAG tetrapeptides to look for macrocyclization and positional variance effects on fragmentation patterns. Additionally, density functional theory computational calculations will be conducted to assess the thermodynamics of the 20 tetrapeptides discussed in this thesis, and the 12 new tetrapeptides. Hydrogen deuterium exchange (HDX) will be performed on these tetrapeptides to explore hydrogen bonding between the amino acid side chains and the peptide backbones, leading to more information on the stabilities of the tetrapeptides. Techniques such as infrared multiphoton dissociation (IRMPD) will be employed to determine how positional variance and/or basicities of lysine, ornithine, DABA, and DAPA can affect the structure of the  $b_n$  ions formed.

Studying the fragmentation patterns of larger lysine and lysine analog containing peptides will provide more insight into the mechanisms of ion formation at play. Analyzing the fragmentation mechanisms of peptides that are 9 or 10 amino acids long which contain lysine or arginine at the C-terminus are particularly relevant to tryptic digests and bottom-up proteomics research. Finally, it will be beneficial to look at doubly protonated lysine species fragmentation studies, and expand research to arginine containing peptides to discern fragmentation mechanisms and ensure a robust proteomics experiment.

## References

- [1] Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., & Smith, R. D. (2012). Mass spectrometry based proteomics: Existing capabilities and future directions. *Chemical Society Reviews*, *41*(10), 3912.
- [2] Raulfs, M. D. M., Brechi, L., Bernier, M., Hamdy, O. M., Janiga, A., Wysocki, V., & Poutsma, J. C. (2014). Investigations of the mechanism of the "proline effect" in tandem mass spectrometry experiments: The "pipercolic acid effect". *J. Am. Soc. Mass Spectrom.*, *25*(10)
- [3] Zhang, Y., Fonslow, B., Shan, B., Baek, M., & Yates, J. I. (2013). Protein analysis by Shotgun/Bottom-up proteomics. *Chemical Reviews*, *113*(4), 2343.
- [4] Roepstroff, P. (2012). Mass spectrometry based proteomics, background, status and future needs. *Protein & Cell*, *3*(9), 641.
- [5] Moradian, A., Kalli, A., Sweredoski, M., & Hess, S. (2014). The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics*, *14*, 489.
- [6] Timmons, M. D., Bradley, M. A., Lovell, M. A., & Lynn, B. C. (2011). Procedure for the isolation of mitochondria, cytosolic and nuclear material from a single piece of neurological tissue for high-throughput mass spectral analysis. *J. Neurosci. Met.*, *197*, 279-282.
- [7] Rani, L., Minz, R., Arora, A., Kannan, M., Sharma, A., Anand, S., . . . Sakhuja, V. (2014). Serum proteomic profiling in granumomatosis with polyangiitis using two-dimensional gel electrophoresis along with matrix assisted laser desorption ionization time of flight mass spectrometry. *Int. J. Rheumatic Disease*, *17*, 910-919.
- [8] Puangpila, C., Mayadunner, E., & Rassi, Z. (2015). Liquid phase based separation systems for depletion, prefractionation, and enrichment of proteins in biological fluids and matrices for in-depth proteomics analysis—An update covering the period 2011–2014. *Electrophoresis*, *36*, 238-252.
- [9] Mitulovic, G., & Mechtler, K. (2006). HPLC techniques for proteomics analysis - a short overview of latest developments. *Briefing in Functional Genomics and Proteomics*, *5*(4), 249-260.
- [10] Monti, M., Orru, S., Pagnozzi, D., & Pucci, P. (2005). Functional proteomics. *Clinica Chimica Acta*, *357*, 140.
- [11] Yan, W., & Chen, S. (2005). Mass spectrometry-based quantitative proteomic profiling. *Briefings in Functional Genomics and Proteomics*, *4*(1), 1.

- [12] Paulo, J., Kadiyala, V., Banks, P., Conwell, D., & Steen, H. (2013). Mass spectrometry-based Quantitative Proteomic profiling of Human Pancreatic and Hepatic stellate cell lines. *Genomics Proteomics Bioinformatics*, *11*, 105-113.
- [13] Ocak, S., Chaurand, P., & Massion, P. (2009). Mass Spectrometry-based proteomic profiling of lung cancer. *Proceedings of the American Thoracic Society*, *6*
- [14] Siuti, N., & Kelleher, N. (2007). Decoding protein modifications using top-down mass spectrometry. *Nature Methods*, *4*(10)
- [15] Whitelegge, J. (2013). Intact protein mass spectrometry and top-down proteomics. *Expert Review of Proteomics*, *10*(2), 127-129.
- [16] Lowenthal, M., Liang, Y., Phinney, K., & Stein, S. (2014). Quantitative bottom-up proteomics depends on digestion conditions. *Anal. Chem.*, *86*, 551-558.
- [17] Kelstrup, C., Frese, C., Heck, A., Olsen, J., & Nielson, M. (2014). Analytical utility of mass spectral binning in proteomic experiments by SPECTral immonium ion detection (SPIID)\*. *Mol. Cell Proteomics*, *13*, 1914-1924.
- [18] Basile, F., & Hauser, N. (2011). Rapid online non-enzymatic protein digestion combining microwave heating acid hydrolysis and electrochemical oxidation. *Analytical Chemistry*, *83*(1), 359-367.
- [19] Turapov, O., Mukamolova, G., Bottrill, A., & Pangburn, M. (2008). Digestion of native proteins for proteomics using a thermocycler. *Analytical Chemistry*, *80*(15), 6093-6099.
- [20] Dong, N., Liang, Y., Xu, Q., Mok, D., Yi, L., Lu, H., . . . Fan, W. (2014). Prediction of peptide fragment ion mass spectra by data mining techniques. *Anal. Chem.*, *86*, 7446-7454.
- [21] Medzihradszky, K., & Chalkley, R. (2015). Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spec. Reviews*, *34*, 43-63.
- [22] Weatherly, D. B., Atwood, J., Minning, T., Cameron, C., Tarleton, R., & Orlando, R. (2005). A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Molecular & Cellular Proteomics*, *4*, 762-772.
- [23] Chamrad, D., Korting, G., Stuhler, K., Meyer, H., Klose, J., & Bluggel, M. (2004). Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data. *Proteomics*, *4*, 619-628.
- [24] Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*, 1367-1372.



- [25] Garwood, K., McLaughlin, T., Garwood, C., Joens, S., Morrison, N., & ... Paton, N. W. (2004). PEDRo: A database for storing, searching and disseminating experimental proteomics data. *BMC Genomics*, 5(68)
- [26] Frewen, B., Merrihew, G., Wu, C., Noble, W., & MacCoss, M. (2006). Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal. Chem.*, 78, 5678-5684.
- [27] Toprak, U., Gillet, L., Maiolica, A., Navarro, P., Leitner, A., & Aebersold, R. (2014). Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Mol. Cell Proteomics*, 13(8), 2056-2071.
- [28] Fenn, J. B., Mann, M., Meng, C., Wong, S., & Whitehouse, C. Electro spray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926), 64-71.
- [29] Ho, C. S., Lam, C., Chan, M., Cheung, R., Law, L., Lit, L., . . . Tai, H. L. (2003). Electro spray ionisation mass spectrometry: Principles and clinical applications. *Clin Biochem Rev*, 14
- [30] Paizs, B., & Suhai, S. (2005). Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews*, 24, 508-548.
- [31] Mosely, J., Smith, M., Prakash, A., Sims, M., & Bristow, A. (2011). Electron-induced dissociation of singly charged organic cations as a tool for structural characterization of pharmaceutical type molecules. *Anal. Chem.*, 83, 4068-4075.
- [32] Zhou, M., & Wysocki, V. (2014). Surface induced dissociation: Dissecting noncovalent protein complexes in the gas phase. *Accounts of Chemical Research*, 47(4), 1010-1018.
- [33] Wysocki, V., Tsaprallis, G., Smith, L., & Brechi, L. (2000). Mobile and localized protons: A framework for understanding peptide dissociation. *Journal of Mass Spectrometry*, 35, 1399-1406.
- [34] Polfer, N., Oomens, J., Suhai, S., & Paizs, B. (2005). Spectroscopic and theoretical evidence for oxazolone ring formation in collision-induced dissociation of peptides. *Journal of the American Chemical Society*, 127, 17154-17155.
- [35] Bythell, B., Somogyi, A., & Paizs, B. (2009). What is the structure of  $b_2$  ions generated from doubly protonated tryptic peptides? *Journal of the American Society for Mass Spectrometry*, 20, 618-624.
- [36] Gucinski, A., Chamot-Rooke, J., Nicol, E., Somogyi, A., & Wysocki, V. (2012). Structural influences on preferential oxazolone versus diketopiperazine  $b_2^+$  ion formation for histidine analogue containing peptides. *Journal of Physical Chemistry*, 116(17), 4296-4304.

- [37] Morrison, L., Chamot-Rooke, J., & Wysocki, V. (2014). IR action spectroscopy shows competitive oxazolone and diketopiperazine formation in peptides depends on peptide length and identity of terminal residue in the departing fragment. *Analyst*, *139*, 2137-2143.
- [38] Perkins, B., Chamot-Rooke, J., Yoon, S., Gucinski, A., Somogyi, A., & Wysocki, V. (2009). Evidence of diketopiperazine and oxazolone structures for HA  $b_2^+$  Ion. *Journal of the American Chemical Society*, *131*
- [39] Nelson, C., Abutokaikah, M., Harrison, A., & Bythell, B. (2016). Proton mobility in  $b_2$  ion formation and fragmentation reactions of histidine-containing peptides. *Journal of the American Society for Mass Spectrometry*, *27*, 487-497.
- [40] Farrugia, J., O'Hair, R., & Reid, G. (2001). Do all  $b_2$  ions have oxazolone structures? Multistage mass spectrometry and ab initio studies on protonated N-acyl amino acid methyl ester model systems. *International Journal of Mass Spectrometry*, *210*(211), 71-87.
- [41] Brechi, L., Tabb, D., Yates, J. I., & Wysocki, V. (2003). Cleavage to N-terminal proline: Analysis of a database of peptide tandem mass spectra. *Anal. Chem.*, *75*(9), 1963-1971.
- [42] Bleiholder, C., Suhai, S., Harrison, A., & 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). *J. Am. Soc. Mass Spectrom.*, *22*, 1032-1039.
- [43] Asakawa, D., Smargiasso, N., Quinton, L., & Pauw, E. (2014). Influences of proline and cysteine residues on fragment yield in matrix-assisted laser Desorption/Ionization in-source decay mass spectrometry. *J. Am. Soc. Mass Spectrom.*, *25*, 1040-1048.
- [44] Tsaprallis, G., Nair, H., Zhong, W., Kuppanan, K., Futrell, J., & Wysocki, V. (2004). A mechanistic investigation of the enhanced cleavage at histidine in the gas-phase dissociation of protonated peptides. *Anal. Chem.*, *76*(7), 2083-2094.
- [45] Huang, Y., Wysocki, V., Tabb, D., & Yates, J. I. (2002). The influence of histidine on cleavage C-terminal to acidic residues in doubly protonated tryptic peptides. *Int. J. Mass Spectrom.*, *219*, 233-244.
- [46] Wang, Y., Vivekananda, S., Men, L., & Zhang, Q. (2004). Fragmentation of protonated ions of peptides containing cysteine, cysteine sulfinic acid, and cysteine sulfonic acid. *Journal of the American Society for Mass Spectrometry*, *15*, 697-702.
- [47] McGee, W., & McLuckey, S. (2013). The ornithine effect in peptide cation dissociation. *Journal of Mass Spectrometry*, *48*, 856-861.
- [48] Harrison, A., Young, A., Bleiholder, C., Suhai, S., & Paizs, B. (2006). Scrambling of sequence information in collision-induced dissociation of peptides. *Journal of the American Chemical Society*, *128*, 10364-10365.

- [49] Riba-Garcia, I., Giles, K., Bateman, R., & Gaskell, S. (2008). Evidence for structural variants of a- and b-type peptide fragment ions using combined ion Mobility/Mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *19*, 609-613.
- [50] Chawner, R., Holman, S., Gaskell, S., & Eyers, C. (2014). Peptide scrambling during collision-induced dissociation is influenced by N-terminal residue basicity. *Journal of the American Society for Mass Spectrometry*, *25*, 1927-1938.
- [51] Schroeder, O., Andriole, E., Carver, K., Colyer, K., & Poutsma, J. C. (2004). Proton affinity of lysine homologues from the extended kinetic method. *Journal of Physical Chemistry*, *108*, 326-332.
- [52] Stawikowski, M., & Fields, G. (2002). Introduction to peptide synthesis. *Current Protocols in Protein Science*, unit 18.1.
- [53] Hood, C., Fuentes, G., Patel, H., Page, K., Menakuru, M., & Park, J. (2008). Fast conventional fmoc solid-phase peptide synthesis with HCTU. *Journal of Peptide Science*, *14*, 97-101.
- [54] Zikos, C., Livaniou, E., Leondiadis, L., Ferderigos, N., Ithakissios, D., & Evangelatos, G. (2003). Comparative evaluation of four trityl-type amidomethyl polystyrene resins in fmoc solid phase peptide synthesis. *Journal of Peptide Science*, *9*, 419-429.
- [55] Hoffmann, E., & Stroobant, V. (2007). *Mass spectrometry principles and applications* (Third Edition ed.). England: John Wiley & Sons Ltd.
- [56] Hiserodt, R., Brown, S., Swijter, D., Harkins, N., & Mussinan, C. (2007). A study of  $b_1 + H_2O$  and  $b_1^-$  ions in the product ion spectra of dipeptides containing N-terminal basic amino acid residues. *Journal of the American Society for Mass Spectrometry*, *18*, 1414-1422.

## Appendix

Below are the graphs of the CID activation parameter scans for AAAX, AAXA, AXAA, XAAA, and YAGX:

