2013

Mass Spectrometry-based Proteomics Experiments and Peptide Fragmentation Studies of Lysine and its Homologues

Charlotte B. Platner

College of William and Mary

Follow this and additional works at: https://scholarworks.wm.edu/honorstheses

Recommended Citation
https://scholarworks.wm.edu/honorstheses/628

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.
# Table of Contents

Table of Contents ....................................................................................................................... i  
Acknowledgements ................................................................................................................... ii  
List of Figures .......................................................................................................................... iii  
List of Tables ........................................................................................................................... iv  
Abstract ...................................................................................................................................... v  

## Chapter 1: Introduction ..........................................................................................................1  
1.1 Protein Structure ..................................................................................................................1  
1.2 Proteomics ............................................................................................................................2  
1.3 Peptide Fragmentation .........................................................................................................4  
1.4 Sequence Scrambling Fragmentation ..................................................................................8  
1.5 Lysine and its Homologues ..................................................................................................9  
1.6 Organisms of Interest .........................................................................................................10  
  1.6.1 Mycobacterium smegmatis .........................................................................................10  
  1.6.2 Bacteriophages Larva and CrimD ...............................................................................11  

## Chapter 2: Experimental Procedures ..................................................................................13  
2.1 Proteomics Studies of *Mycobacterium smegmatis* ............................................................13  
  2.1.1 Culture and Lysis of Mycobacterium smegmatis .......................................................13  
  2.1.2 Gel Electrophoresis .....................................................................................................13  
  2.1.3 Tryptic Digest .............................................................................................................15  
  2.1.4 High Performance Liquid Chromatography (HPLC) .................................................16  
2.2 Mass Spectrometry Experiments .......................................................................................18  
  2.2.1 Ionization Techniques .................................................................................................18  
  2.2.2 Electrospray Ionization (ESI) .....................................................................................19  
  2.2.3 Mass Analyzers ..........................................................................................................20  
  2.2.4 Data-Dependent Tandem Mass Spectrometry ............................................................21  
  2.2.5 SEQUEST ...................................................................................................................22  
  2.2.6 Proof of Concept Experiments with Bovine Serum Albumin (BSA) .........................23  
2.3 Peptide Fragmentation of Lysine and its Homologues ......................................................24  
  2.3.1 Solid Phase Peptide Synthesis ....................................................................................24  
  2.3.2 Mass Spectrometry Fragmentation of Tripeptides ......................................................26  

## Chapter 3: Proteomics Studies .............................................................................................27  
3.1 Proof of Concept Experiments with Bovine Serum Albumin (BSA) ...............................27  

## Chapter 4: Tripeptide Fragmentation ..................................................................................40  
4.1 AAX Peptides ....................................................................................................................40  
4.2 AXA Peptides ....................................................................................................................43  
4.3 XAA Peptides ....................................................................................................................46  

## Chapter 5: Conclusions .........................................................................................................50  

References ...............................................................................................................................52
Acknowledgements

I would like to thank Dr. Poutsma for all of his advice, support, and guidance throughout my research these past four years. It has been one of the most rewarding experiences of my college career. I would also like to acknowledge the generous funding I received through the College of William and Mary from Howard Hughes Medical Institute (HHMI) and the Arnold and Mabel Beckman Foundation. The Experimental Genomics and Proteomics lab members also deserve much thanks, without them this project would never have come this far.

I would also like to thank my friends and fellow ionlab members for their friendship, optimism, and for always making me laugh when I desperately needed it. Finally, I would like to thank my family, especially my parents and sister, for their constant support and encouragement. I could not have made it without them.
List of Figures

Figure 1.1 – Basic structure of an amino acid .................................................................1
Figure 1.2 – Steps for mass spectrometry-based proteomics .........................................4
Figure 1.3 – Peptide fragmentation nomenclature ..........................................................5
Figure 1.4 – Mechanism for charge-directed peptide cleavage ......................................7
Figure 1.5 – Sequence scrambling fragmentation ..........................................................8
Figure 1.6 – Lysine and its homologues ......................................................................9
Figure 2.1 – Alkylation of cysteine by iodoacetamide ..................................................16
Figure 2.2 – Diagram of MALDI and ESI ...................................................................19
Figure 2.3 – Schematic of an Electrospray Ionization source ......................................20
Figure 2.4 – Structure of alanine ..............................................................................24
Figure 3.1 – BSA chromatogram from 40 minute MeOH gradient ..................................27
Figure 3.2 – Plot of retention time versus MH+ for BSA 40 minute MeOH gradient ......28
Figure 3.3 – M. smegmatis chromatogram from 130 minute MeOH gradient ...............29
Figure 3.4 – BSA chromatogram from 130 minute MeOH gradient ..............................30
Figure 3.5 – Plot of retention time versus MH+ for BSA 130 minute MeOH gradient ....31
Figure 3.6 – BSA chromatogram from 130 minute ACN gradient ...............................32
Figure 3.7 – Plot of retention time versus MH+ for BSA 130 minute ACN gradient .......33
Figure 3.8 – BSA chromatogram from 65 minute ACN gradient ...................................34
Figure 3.9 – Plot of retention time versus MH+ for BSA 65 minute ACN gradient .........34
Figure 3.10 – BSA chromatogram from 80 minute ACN gradient .................................35
Figure 3.11 – Plot of retention time versus MH+ for BSA 80 minute ACN gradient .......36
Figure 3.12 – BSA chromatogram from 90 minute ACN gradient ..................................37
Figure 3.13 – Plot of retention time versus MH+ for BSA 90 minute ACN gradient ......37
Figure 3.14 – Plot of retention time versus ACN concentration for 90 minute gradient ...38
Figure 3.15 – Blank solvent injection, 90 minute ACN gradient ....................................39
Figure 4.1 – Fragmentation spectrum of AAK ..............................................................40
Figure 4.2 – Fragmentation spectrum of AAO ..............................................................41
Figure 4.3 – Fragmentation spectrum of AA-Daba ......................................................41
Figure 4.4 – Fragmentation spectrum of AA-Dapa ......................................................42
Figure 4.5 – Fragmentation spectrum of AKA ..............................................................44
Figure 4.6 – Fragmentation spectrum of AOA ..............................................................44
Figure 4.7 – Fragmentation spectrum of A-Daba-A .....................................................45
Figure 4.8 – Fragmentation spectrum of A-Dapa-A .....................................................45
Figure 4.9 – Fragmentation spectrum of KAA ..............................................................46
Figure 4.10 – Fragmentation spectrum of OAA ............................................................47
Figure 4.11 – Fragmentation spectrum of Daba-AA ....................................................47
Figure 4.12 – Fragmentation spectrum of Dapa-AA ....................................................48
List of Tables

Table 1.1 – Pathogens related to Mycobacterium smegmatis .................................................. 10
Table 2.1 – Synthesized tripeptides ......................................................................................... 24
Abstract

This thesis presents mass spectrometry-based proteomics experiments and peptide fragmentation studies of lysine and its homologues. Proof of concept experiments were performed with a tryptic digest of bovine serum albumin (BSA) to optimize the proteomics protocols. Solvent gradients from the high performance liquid chromatography (HPLC) instrument were optimized first using a mixture of methanol and water and subsequently with acetonitrile and water. Data-dependent scans were run to isolate and fragment peptides to gain sequence information. Proteins were identified using the proteomics sequencing software SEQUEST. Peptide fragmentation studies of lysine and its homologues were performed by placing a lysine homologue in each of the three positions on a tripeptide with alanine in the other two positions to give tripeptides in the forms AXA, XAA, and AAX. Our results found a pattern of sequence scrambling in the AAX and AXA peptides to give fragments that appear to be from the XAA peptide. Sequence scrambling was not seen in the XAA peptides. This result suggests that the lysine homologues are most stable at the N-terminus which could be a result of their high gas phase proton affinities. Correlations between side chain length and fragmentation pattern were also seen.
Chapter 1 - Introduction

1.1 Protein Structure

Proteins are fundamental biological macromolecules found in all living systems. They are essential for the survival of living cells and perform an extraordinarily wide variety of functions. These functions include catalysis, transport, metabolic control, and acting as structural molecules. Proteins are linear, chain polymers, or polypeptides, made up of small subunits called amino acids. There are 20 widely accepted protein amino acids coded in the DNA and RNA of living things. All amino acids have the same basic structure but differ in what is known as the side chain, or R group. The basic structure of an amino acid is shown in Figure 1.1 below.

![Figure 1.1](image)

Amino acids are joined together to form polypeptides and proteins by peptide bonds. These bonds form via a condensation reaction between the amine group (N-terminus) of one amino acid and the carboxylic acid (C-terminus) of another. The hydroxyl group from the C-terminus and a hydrogen from the amine leave as water resulting in a bond between the carboxylic acid carbon and the amine nitrogen.

Proteins do not simply exist as long, loose chains of amino acids. They are three-dimensional structures held together by various types of chemical interactions. There are four canonical levels of protein structure. The primary structure (1°) of a protein is its amino acid sequence which is coded in the DNA. The secondary (2°) structure of a protein refers to the
formation of three dimensional folded structures, the most common of which are called α-helices and β-pleated sheets, from extended sequences of amino acids. Tertiary (3°) structure refers to the total three-dimensional folded structure of a protein. The final level, known as the quaternary (4°) structure, occurs only with complexes of multiple proteins held together by non-covalent interactions. In their native state, proteins are folded into specific conformations that give the proteins their functions.

The characteristics of amino acid side chains are largely responsible for how a given protein folds. Hydrophobic amino acid residues tend to interact with one another and be located in the center of the folded protein. Hydrophilic amino acid residues, on the other hand, tend to be located along the exposed surface of the protein because they interact with water molecules that make up much of the protein’s external environment. The three-dimensional structure is the result of various side chain interactions including hydrophobic interactions, hydrogen bonding, and covalent bonding, especially the formation of disulfide bonds. The dependence of protein function on three dimensional structure and therefore on the primary sequence of amino acids has led to an increasing interest in protein sequencing and the field of proteomics.

1.2 Proteomics

Proteomics is the study of all the proteins expressed by an organism under specific physiological conditions. Over the past few decades the field of genomics churned out immense amounts of information about the genetic code of humans and many other species. With the full genome information, and therefore the sequence of every potential encoded protein, of many organisms readily available, focus has turned to study at the protein level. Unlike the genome, the proteome of an organism is highly variable. The expressed proteins within a single cell can differ as a result of external conditions while the protein expression of different cells in a
complex organism depends significantly on the identity of the cell and the tissue with which it is associated.

Prior to mass spectrometry-based studies, protein sequencing was done using a time consuming method called the Edman degradation which involved successive stepwise cleavages of amino acids beginning at the N-terminus. With the advent of two ‘soft’ ionization sources that made mass spectrometry accessible for the study of large biomolecules, the field of mass spectrometry based proteomics has grown exponentially. These ionization methods are electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI).

There are two principle approaches to proteomics using mass spectrometry: Top-down and bottom-up. Top-down proteomics experiments involve tandem mass spectrometry experiments on intact proteins and therefore require high resolution instruments. Bottom-up studies typically involve separation of proteins with either one or two dimensional gel electrophoresis followed by enzymatic digestion of proteins into small pieces, called peptides. These peptides are further separated and then analyzed using liquid chromatography and mass spectrometry. A generalized bottom-up protocol is detailed in figure 1.2.
1.3 Peptide Fragmentation

Bottom-up proteomics studies are characterized by fragmentation studies of tryptic peptides. The fragmentation patterns of these peptides are then used to make an identification of the peptide. This is possible because peptides fragment according to well characterized rules depending on the type of activation used. Typically, the mode of fragmentation used is collision induced dissociation (CID) using an inert collision gas such as helium or argon. CID is a relatively low-energy fragmentation method and, as such, gives reproducible and characteristic fragmentation patterns. Peptides fragmented with collision induced dissociation tend to fragment...
at the amide bond (also called the peptide bond) with little to no side-chain fragmentation. When
the amide bond of a peptide breaks, there are two possible fragment generated, one with the
charge remaining on the N-terminal side and one with the charge remaining on the C-terminal
side of the amide bond. As a mass spectrometer measures the mass to charge (m/z) ratio, only
charged species will be seen by the detector. Singly-protonated peptides will give one of two
fragments depending on which side of the amide bond keeps the proton. If the N-terminal side is
the charged species, that fragment is designated a b-type ion. If the C-terminal side remains
charged, it is designated a y-type ion. For a singly-charged peptide, only one of these two ions
can form from each precursor ion. For multiply charged ions, both are possible. Under more
energetic fragmentation conditions, other bonds along the peptide backbone can break.
Depending on which side of the peptide keeps the proton, this fragmentation can result in x- or a-
type ions and z-or c-type ions. Nomenclature for peptide fragmentation under collision induced
dissociation is summarized in the figure below.

Other common fragments result from the loss of neutral species including, most commonly, loss
of ammonia or water.
The most widely accepted model for peptide fragmentation under low-energy fragmentation conditions, such as collision induced dissociation (CID), is the mobile proton model introduced separately by Wysocki and Gaskell.\textsuperscript{6,7} The mobile proton model explains movement of protons along the peptide backbone to initiate cleavage of specific bonds. This mode of bond cleavage is called charged-directed cleavage.\textsuperscript{6,7} The mobile proton model has been verified using deuterium labeling experiments to show the movement of this proton.\textsuperscript{8} When energy is added to the peptide molecule in the form of gas-phase collisions, protons move to different sites along the peptide backbone to initiate cleavage. The mechanism of charge-directed cleavage of a typical peptide to give a b-type ion is shown in figure 1.4. It involves movement of the proton from a more basic site to a carbonyl oxygen on the peptide backbone. From here, nucleophilic attack of the protonated carbonyl by another carbonyl oxygen results in an intramolecular ring structure which can open to form either a b- or y-type ion. B-type ions can have several structures, but are commonly protonated oxazalone ions. Proton transfer from the oxazalone ion to the y-fragment via a proton-bound dimer intermediate gives a y-type ion.\textsuperscript{8} The mechanism of charge-directed cleavage for a tripeptide is depicted in figure 1.8.\textsuperscript{6,8}
Gas-phase basicity and proton affinity of amino acids play a big role in the energetics of this movement. The more basic an amino acid side chain, the higher its proton affinity. As a result, it requires more energy input to move a proton from a highly basic side chain to a protonation site along the peptide backbone. Of the twenty protein amino acids, in the gas phase, arginine, histidine, and lysine have the highest proton affinities and gas phase basicities.\(^9\)
1.4 Sequence-Scrambling Fragmentation

Sequence-scrambling fragmentation is a mechanism to explain the appearance of fragment peaks that do not belong to the peptide being studied. Bleiholder et al published a study of the fragmentation pattern of the peptide YAGFL (Y=tyrosine, A=alanine, G=glycine, F=phenylalanine, L=leucine) under MS² and MS³ tandem mass spectrometry. Using ¹⁵N labeling at the N-terminus, they were able to study the formation of non-sequence ions. Direct-sequence ions for peptides are the expected a, b, c, x, y, z type ions discussed previously. Non-sequence ions are those whose m/z values cannot be directly predicted from the sequence of the peptide being fragmented. Bleiholder et al. proposed that these ions are the result of intramolecular cyclization followed by ring opening at a different place. Both steps of this process are the result of collisions with the inert collision gas. As a result of the differential ring opening, the resulting ions appear to be fragments of a peptide with a scrambled sequence. This process is summarized with the peptide YAGFL in figure 1.5 below.
1.5 Lysine and its Homologues

As one of the most highly basic amino acids, lysine should influence the fragmentation behavior of peptides according to the mobile proton model under collision induced dissociation. The goal of our peptide fragmentation studies was to determine how the side-chain length of lysine and its homologues influenced the fragmentation behavior of tripeptides. Better understanding of peptide fragmentation behavior is important because fragmentation is the basis of protein identification in bottom-up proteomics studies. The better fragmentation behavior is understood, the better protein identifications will be. Lysine and its homologues ornithine, 2,4 diaminobutanoic acid (Daba), and 2,3 diaminopropanoic acid (Dapa) each differ by one methylene in the length of the side chain. Lysine and its homologues are shown in figure 1.6 below.

![Lysine and its Homologues](image)

Figure 1.6

Lysine’s high gas phase basicity can be explained by the presence of two amino groups. When lysine is protonated, an intramolecular hydrogen bond is formed between the amino group on the N-terminus and the amino group on the side chain. Lysine and its homologues have been studied...
previously in the Poutsma group.11,12 Their gas-phase proton affinities were determined by Poutsma et al. using the Extended Kinetic Method and were confirmed using high level Density Functional Theory (DFT) calculations. Proton affinity is defined as the negative enthalpy of protonation.13 The calculated proton affinities for lysine, ornithine, Daba, and Dapa are 1004.1±8.0 kJ/mol, 1001.1±6.6 kJ/mol, 975.8±6.6 kJ/mol, and 950.2±7.1 kJ/mol respectively.13 From these values, we can see that the proton affinities of Lysine and Ornithine are nearly identical. As a result, differences in their fragmentation patterns will not be enthalpic differences, but likely entropic. Daba and Dapa have significantly different proton affinities so differences in fragmentation patterns here will be both enthalpic and entropic.

1.6 Organisms of Interest

1.6.1 Mycobacterium smegmatis

*Mycobacterium smegmatis* is fast-growing and non-pathogenic soil bacterium that is closely related to other pathogenic mycobacteria that cause disease. As a result, it is often used as a model system for these pathogenic mycobacteria. Some of these bacterial pathogens and the diseases they cause are listed in the table below.

<table>
<thead>
<tr>
<th>Bacterium</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mycobacterium tuberculosis</em></td>
<td>tuberculosis</td>
</tr>
<tr>
<td><em>Mycobacterium avium</em> subsp. <em>Paratuberculosis</em></td>
<td>Johne's disease</td>
</tr>
<tr>
<td><em>Mycobacterium leprae</em></td>
<td>leprosy</td>
</tr>
</tbody>
</table>

Table 1.1

The genomes of mycobacteria have been sequenced but much of the mechanism responsible for growth and proliferation of these species is still unknown. Mycobacteria are gram positive bacteria characterized by a thick, waxy cell wall that makes them highly impermeable. This allows for survival in harsh, antibiotic conditions. The cell envelope of mycobacteria is also
unique. It is made up of a regular lipid inner membrane and a second outer layer of mycolic acids that are tightly connected to the outer layer of peptidoglycan. This feature and the waxy cell wall combine to make mycobacteria highly protected.\textsuperscript{14} \textit{Mycobacterium tuberculosis} (the causative agent of tuberculosis), for example, can exist, upon infection, latent in the human body for the duration of one’s lifetime with no symptoms. Reactivation can be triggered by immunosuppression due to age, infection by another virus or bacteria, malnutrition, and a host of other factors. Of the three diseases mentioned previously, tuberculosis is the focus of the most research. According to the World Health Organization (WHO), there were 8.7 million new cases of tuberculosis and 1.4 million deaths as a result of infection in 2011 alone.\textsuperscript{15} Studying protein expression and growth mechanisms of \textit{M. smegmatis} can provide insights into these same processes in \textit{M. tuberculosis} and could potentially result in new drug and vaccine targets.

1.6.2 Bacteriophages Larva and CrimD

The bacteriophages Larva and CrimD were first isolated and cultured at the College of William and Mary by the freshman Phage Lab funded by the Howard Hughes Medical Institute (HHMI). Bacteriophages are viruses that infect bacteria. They have long been considered a possible mechanism for targeting and killing pathogenic bacteria. Bacteriophages, often abbreviated as ‘phages’, are characterized by their polygonal heads and tubular tails.\textsuperscript{16} Phages are classified according to gross DNA relationships. There are 11 clusters, named A-K. Larva and CrimD belong to Cluster K based on their genomic relationship to the most commonly studied mycobacteriophage, TM4. A 13 base pair sequence involved in the initiation of translation of a specific set of genes is well conserved across the cluster. Cluster K bacteriophages are also called mycobacteriophages because they infect mycobacteria. They infect both fast and slow growing mycobacteria and all, with the exception of TM4, form stable
lysogens in these bacteria. Cluster K is further subdivided based on nucleotide similarities into numbered subclusters. Larva belongs to subcluster K5 while CrimD belongs to subcluster K1.17
Chapter 2 – Experimental Procedures

2.1 Proteomics Studies of Mycobacterium smegmatis

2.1.1 Culture and Lysis of Mycobacterium smegmatis

*M. smegmatis* strain MC\(^2\)155 was cultured for 24 to 48 hours to an optimal optical density in Middlebrook7H9 media at 37°C. This culture was then subdivided into three samples. One remained uninfected while the other two were infected by either the Larva or CrimD bacteriophage. Infected *M. smegmatis* sub-samples were then harvested at specific time points over the course of infection. The time points used in these experiments were 0, 20, 40, 60, 90, and 120 minutes post infection. Cell lysis was performed according to the protocol outlined by Marcotte et al.\(^{18}\) The cells were subjected to both chemical and mechanical lysis to release cellular contents. Mechanical lysis was accomplished using 1mm glass beads in a bead-beating apparatus. This step is key for lysis of mycobacteria because their thick, waxy mycolic acid layer, meant to protect the cell, inhibits chemical digestion by the lysis buffer. Cell lysates were then centrifuged to remove heavier cellular debris such as portions of the membrane. The supernatant was then ready for protein separation by gel electrophoresis.

2.1.2 Gel Electrophoresis

Gel electrophoresis is a method of protein separation used commonly in proteomics experiments to separate complex mixtures of proteins into resolvable bands that can be analyzed individually. Protein separation in polyacrylamide gel can occur in one or two dimensions. One-dimensional gel electrophoresis separates proteins based on their size while two-dimensional gel electrophoresis separates proteins by size in one dimension, and by isoelectric point in the other. Isoelectric point (pI) is the pH at which a peptide or protein has zero net charge. For one-dimensional gel electrophoresis, the anionic detergent sodium dodecyl sulfate (SDS) is used.
SDS binds to proteins in amounts proportional to the molecular weight of the protein. The resulting peptide has a negative charge that is proportional to its molecular weight. When an electric field is applied, the charged proteins migrate through the gel. Small proteins are able to migrate more quickly through the pores in the polyacrylamide gel. The rate of migration is proportional to the logarithm of the molecular weight\textsuperscript{19}.

Two-dimensional gel electrophoresis begins with separation in one dimension based on isoelectric point, a process called isoelectric focusing. Amino acids, the building blocks of proteins, are what are known as amphoteric molecules, meaning that they possess both acid and base properties. As a result, their charge state depends on the pH of their environment. Protein mixtures are applied to a gel strip with an immobilized pH gradient and an electric potential is applied across the gel. Above its isoelectric point, a protein will have a net negative charge and will consequently move towards the anode while a protein below its isoelectric point will have a net positive charge and migrate towards the cathode. When the protein reaches its isoelectric point, it has no net charge and will no longer migrate in response to the applied electric potential. Once the isoelectric focusing step is complete, the resulting gel strip is placed horizontally along the top of an SDS polyacrylamide gel for separation by size according to the same procedure as the one dimensional gel.

The final step for either one or two-dimensional gel electrophoresis is staining to make the protein bands (in a 1-D gel) or spots (in a 2-D gel) visible to the naked eye. The two most common staining techniques are Coomassie blue and Silver stain. Coomassie blue is easier to remove but less sensitive than silver stain. Silver stain is thought to bind preferentially to basic amino acid residues but the mechanism is not yet fully understood. Further chemical washes are necessary to make silver stain visible. Coomassie blue binds directly to basic and aromatic amino
acid residues of a protein and requires no further chemical activation. Coomassie staining is typically considered to be more compatible with mass spectrometry while silver stain requires more extensive removal of the stain prior to mass spectrometry.\textsuperscript{20}

SDS polyacrylamide gels were made according to standard protocols. Samples were each mixed with a running buffer also containing the SDS detergent. To prevent samples from running together, only every other lane was loaded with a sample, with a standard molecular weight ladder located in lane 1. The gels were run for 4 hours at 35mA, or until the dye from the running buffer was approximately 1cm above the bottom of the gel. The gel was then stained for visible bands with Coomassie Blue.

2.1.3 Tryptic Digest

The in gel tryptic digest was performed according to the Arizona Proteomics Consortium Protocol\textsuperscript{21} which is based upon a previously published protocol\textsuperscript{22}. Gel bands were excised and subsequently cut into approximately 1mm\textsuperscript{2} square pieces. They were then washed in a series of steps using acetonitrile (ACN) and ammonium bicarbonate (abbreviated Ambic) to remove the Coomassie blue stain. Next the disulfide bonds formed by cysteine residues in the protein were reduced using dithiotreitol (DTT). To prevent them from reforming, iodoacetamide (IAM) was added to alkylate the cysteine residues.
This modification changes the molecular weight of each cysteine residue by +57Da. Following the reduction and alkylation of the cysteine residues, trypsin is added to begin the digest. Trypsin is a protease that selectively cleaves the amide bond of the peptide backbone to the carboxyl side of lysine and arginine residues. Following overnight incubation, the tryptic digestion is quenched with trifluoroacetic acid and the tryptic peptides are extracted from the gel. The peptide mixtures are subsequently concentrated down to approximately 30µL and ready for analysis.

2.1.4 High Performance Liquid Chromatography (HPLC)

Chromatography is the science of separation. In typical chromatography experiments there is a mobile phase and a stationary phase. The separation of analytes is based on competition between the mobile and stationary phases for interaction with the analyte. Typically separation in liquid chromatography is based on polarity. High Performance Liquid Chromatography (HPLC) is a method of separating analytes dissolved in solution. HPLC is characterized by the use of high-pressure pumps to push a mobile phase containing the analytes of interest through a column containing the stationary phase. The most common method of
HPLC today is reverse-phase chromatography which uses a non polar column and a mixture of a polar solvent (like water) with a less polar (typically organic) solvent for the mobile phase. This is considered to be reverse phase simply because early chromatography experiments were performed with the polarities of these components reversed.

Reverse-phase liquid chromatography was performed on the peptide mixtures using a Hewlett Packard Series 1100 HPLC instrument. The column used was a 30x 2.1mm i.d. packed C18 column from ACE ® HPLC Columns (ACE 3 C18-300). An HPLC flow rate of 0.5mL/min was used. Subsequently, the solvent was run through a 50:1 splitter prior to the column. This resulted in a flow rate of 10µL/min. The split was necessary to interface the HPLC with the mass spectrometer. A precise gradient is best achieved at flow rates at or above 0.5mL/min while the mass spectrometer works best at flow rates on the order of 10µL/min. Various solvent gradients of water and methanol were applied to provide optimal elution of peptides. 1% formic acid was added to each of the solvents to protonate the peptides in preparation for mass spectrometry analysis. In general, the solvent concentrations started at 95% H2O and 5% MeOH and were ramped over the course of various gradient lengths to a final concentration of 5% H2O and 95% MeOH prior to a final flush of 95% H2O and 5% MeOH. The most successful gradient used for proof of concept experiments was a 130 minute step gradient of MeOH in H2O as follows:

- 0-20 minutes – 5% MeOH
- 20-30 minutes – 15% MeOH
- 30-40 minutes – 25% MeOH
- 40-50 minutes – 35% MeOH
- 50-60 minutes – 45% MeOH
- 60-70 minutes – 55% MeOH
- 70-80 minutes – 65% MeOH
- 80-90 minutes – 75% MeOH
- 90-100 minutes – 85% MeOH
- 100-110 minutes – 95% MeOH
- 110-130 minutes – 5% MeOH
Following positive controls and proof-of-concept experiments with predigested bovine serum albumin (BSA) and E. coli cell lysates, it was determined that, seemingly independent of the length of the gradient, most of the peptides were being eluted near the end of the gradient. Literature searches determined that the most common solvent mixtures in bottom-up HPLC-ESI-MS/MS experiments were acetonitrile (ACN) and water. New solvents were prepared as by Hess et al.\textsuperscript{26} This protocol gives a polar solvent consisting of 98% H\textsubscript{2}O, 2% ACN with 0.2% formic acid and a non polar solvent consisting of 98% ACN, 2% H\textsubscript{2}O with 0.2% formic acid.

2.2 Mass Spectrometry Experiments

There are three basic components to a mass spectrometer. The first is an ionization source to produce gas-phase ions of the analyte of interest. The two most common ionization sources in proteomics studies are electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI). Next is the mass analyzer that measures the mass to charge (m/z) ratio of an analyte ion. There are three principle types of mass analyzers common in proteomics experiments. These are quadrupole, time-of-flight (TOF), and ion-trap instruments. These can also be combined to perform tandem mass spectrometry experiments. The final component of a mass spectrometer is the detector which measures the number of ions at a given m/z.

2.2.1 Ionization Techniques

The field of mass spectrometry based proteomics was revolutionized by the advent of two ionization sources that allowed for the intact ionization of large biomolecules: electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI). Both are what are known as “soft” ionization techniques. This means that they form intact ions for analysis in the mass spectrometer and result in little to no side chain fragmentation. Both produce what are known as pseudo-molecular ions (MH\textsuperscript{+}). They differ in that ESI tends to produce multiply
charged ions while MALDI tends to produce singly charged ions. As a result, MALDI is often paired with TOF mass analyzers to measure exact peptide masses while ESI is often paired with ion-trap or triple-quadrupole mass analyzers to measure fragmentation spectra from tandem (MS/MS) mass spectrometry experiments. Basic ESI and MALDI sources are diagrammed in figure 2.2 below. ESI is the method of ionization used in our experiments.

2.2.2 Electrospray Ionization (ESI)

In electrospray ionization, the analyte of interest is ionized in solution and this solution is introduced through a capillary. The charged analyte is sprayed through a charged spray nozzle to produce an aerosol of charged droplets. As the solvent evaporates from these charged droplets, they reach what is called the “Rayleigh limit,” the point at which the coulombic repulsion of the analyte ions, all with the same charge, overcomes the surface tension that holds the droplets together in what is called a “Coulombic Explosion”\textsuperscript{27}. Solvent evaporation is aided by a heated
capillary. This process repeats itself and the droplets get smaller and smaller until the analyte ions are desorbed into the ambient gas of the mass spectrometer as pseudo-molecular ions. A typical ESI source is diagrammed in figure 2.3 below.

![Diagram of ESI source](image)

**Figure 2.3**

For tryptic peptides, ESI tends to produce doubly charged ions (i.e. M+2H^+\textsuperscript{2}) but can produce higher charge states for peptides greater than fifteen amino acid residues in length as well as for peptides that contain especially basic residues like histidine\textsuperscript{2}.

2.2.3 Mass Analyzers

As stated previously, TOF analyzers are typically paired with MALDI ionization sources to produce spectra with very high resolution. Often this allows for measurement of exact protein or peptide masses without needing to do any sort of fragmentation. Quadrupole and ion-trap instruments are typically paired with ESI sources for proteomics experiments because they have lower resolution and mass ranges and therefore need multiply charged ions to see large molecular weight species.
Our studies used an electrospray ionization source paired with an ion-trap mass spectrometer. In this instrument, ions pass through focusing octopoles into an ion trap that is tuned to isolate ions based on their m/z which are then held or ‘trapped’ for a certain specified amount of time before undergoing MS or MS^n analysis. Ions are trapped in a potential well when specific voltages are applied to electrodes located in the ion trap. Ion-traps are limited in their resolution due to the limited number of ions that can be isolated in the center of the trap before charge repulsion distorts the distribution and therefore the mass accuracy.

For many proteomics studies, the mass spectrometer is run in data-dependent mode. Unlike traditional scans, this does not require each peak to be individually isolated and analyzed by hand. A scan can be set up to have the mass spectrometer isolate and fragment the n most intense ions from a given scan, to save these spectra, and move on to the next spectrum. In this manner, an immense amount of data can be gathered from a single experiment. This data is then analyzed by proteomics software such as SEQUEST or Mascot.

2.2.4 Data-Dependent Tandem Mass Spectrometry

Peptides separated by the HPLC were run on-line into an electrospray ionization (ESI) source and from there into a Thermo LCQ-DECA ion trap mass spectrometer. The electrospray ionization source used a spray voltage of 5000 volts, a capillary temperature of 125°C, and a capillary voltage of 5 volts. Nitrogen gas was used as a sheath gas to aid in desolvation of the analyte ions and to push them towards the trap. A data-dependent scan was run for the duration of the chromatography experiment. The selected mass range in the ion trap was 400-1850 m/z. The mass spectrometer was run in centroid mode. The 5 most intense parent ions from each scan were isolated and fragmented at with collision energy of 35% and activation Q of 0.250. Fragmentation was performed using collision induced dissociation (CID) using an inert collision
gas, in this case, Helium. Minimum parent ion signal was 100000 and the minimum MS/MS signal is 5000. To prevent analysis of known noise peaks, an exclusion list consisting of peaks found in ESI calibration solution (195-Caffeine, 524-MRFA, and 1121, 1222, 1322….1822-Ultramark) was used so that these m/z values would be excluded from analysis. A dynamic exclusion list was also used so that the same peptide peaks were not repeatedly isolated and fragmented at the expense of less abundant peptide peaks. Once a parent peak had been seen 3 times within 30 seconds, it was added to the exclusion list which prevents that peak from being analyzed for 10 minutes. The dynamic exclusion list size was to set to the maximum value of 50. This isolation and fragmentation of the 5 most intense ions was repeated every 250 ms and continued throughout the whole scan. Once the scan was complete, the data was ready for analysis by the proteomics algorithm SEQUEST.

2.2.5 SEQUEST

This software uses search algorithms to search the raw data for peaks. The parent peak m/z data are used to generate theoretical peptide sequences that would produce a parent peak of each specific m/z. These theoretical peptides are then used to produce theoretical fragmentation spectra using well characterized rules of peptide fragmentation. These theoretical spectra are then matched with the experimental spectra to make an identification. Peptide matches are scored based on how well the theoretical fragmentation spectrum matches the experimental spectrum and this is given by a cross correlation value (Xcorr). Identified peptides are then searched against the sequence information of one or many proteins in FASTA files.²⁹

Raw mass spectrometry data was analyzed using the Xcalibur Proteome Discoverer software (Version 1.3). The SEQUEST algorithm allows for the input of various search parameters to make identification more accurate. A mass range of 350-5000 Da was used with
the knowledge that having multiply charged ions would put some of the heavier molecular weight peptides within our m/z range of 400-1850. Activation type was set as collision induced dissociation (CID). The appropriate database was selected based on the sample type. The enzyme used was set to trypsin and a maximum of two missed cleavages was allowed. The precursor (or parent) mass tolerance was set to 1.0 Da and the fragment mass tolerance was set to 0.8 Da. This allows the theoretical parent mass to differ from the experimental by a maximum of 1.0 Da and the theoretical fragment mass to differ from the experimental by a maximum of 0.8 Da. The final parameters that were selected were amino acid modifications that change the molecular weight of the amino acid residues. Dynamic (or variable) modifications allowed for are oxidation of methionine (+16 Da), N-terminal alkylation (+42 Da), and C-terminal oxidation (+16 Da). Carbamidomethyl modification of cysteine (+57 Da) was designated a static modification. This modification accounts for the use of iodoacetamide (IAM) to prevent disulfide bonds from reforming during the tryptic digestion.

2.2.6 Proof of Concept Experiments with Bovine Serum Albumin (BSA)

Predigested BSA was purchased from Michrom Bioresources, Inc out of Auburn, California. It was brought up in a solution of 50:50 MeOH:H₂O to a final concentration of 2pmol/µL. The predigested BSA, as in our procedure, had undergone reduction of cysteine-cysteine disulfide bonds prior to digestion. They used a slightly different alkylation reagent, Iodoacetic acid (IAA), which results in an addition of 58 Da per cysteine residue. BSA was used to adjust the HPLC-MS and SEQUEST portions of our proteomics experiments for optimal protein coverage. The results of these experiments and the subsequent adjustments made to the protocol will be discussed later.
2.3 Peptide Fragmentation of Lysine and its Homologues

Tripeptides AXA, XAA, and AAX, where X is either lysine, ornithine, daba, or dapa and A is alanine were synthesized using standard solid-phase synthetic techniques.\(^{31,30}\) Alanine was chosen because its side chain is simply a methyl group and, consequently, it has a very low gas phase proton affinity compared to lysine. The proton affinity of alanine was determined by Alex G. Harrison to be 214.2 kcal/mol.\(^{9}\) The structure of alanine is shown in figure 2.4 below.

![Figure 2.4](image)

The synthesized peptides are listed in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Lysine</th>
<th>Ornithine</th>
<th>Daba</th>
<th>Dapa</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXA</td>
<td>AKA</td>
<td>AOA</td>
<td>A-Daba-A</td>
<td>A-Dapa-A</td>
</tr>
<tr>
<td>XAA</td>
<td>KAA</td>
<td>OAA</td>
<td>Daba-AA</td>
<td>Dapa-AA</td>
</tr>
<tr>
<td>AAX</td>
<td>AAK</td>
<td>AAO</td>
<td>AA-Daba</td>
<td>AA-Dapa</td>
</tr>
</tbody>
</table>

Table 2.1

2.3.1 Solid Phase Peptide Synthesis

Solid phase peptide synthesis was first introduced by R. B. Merrifield and involves attaching an amino acid to an insoluble particle so that reagents can be washed or filtered through during each step of the synthesis.\(^{30}\) This greatly shortened and simplified peptide synthesis procedures. The synthesis procedure used for these tripeptides involved the use of amino acids bonded to a wang resin at the carboxyl end and the growing peptide chain was synthesized from the carboxyl terminus.\(^{31}\) The amino terminus is bonded to a temporary protecting group to prevent unwanted side reactions. For these studies, Fmoc was used as a
protecting group. Side-chain functional groups required a permanent protecting group that would be unaffected by the synthesis reagents but could be later removed once synthesis is complete.\textsuperscript{31} This was important for our tripeptides because lysine and its homologues have an amino group on the side chain as well as at the N-terminus. In these studies, a tert-butoxycarbonyl group (Boc) was used to protect the amino group on the side chain of these amino acids.

To start, a small amount of wang resin with the C-terminal amino acid attached was loaded into Henke Sass Wolf (HSW) 20 mL plastic syringe with built-in filter. The resin beads are too large to fit through the filter so reagents can be washed through without the growing peptides being lost. Prior to beginning the synthesis procedure, the resin beads were swelled with 50:50 DMF:DCM for 30 minutes. The general procedure for solid phase peptide synthesis involves first deprotecting the amino group of the amino acid bound to the resin. This was accomplished with 20% piperidine in dimethlyformamide (DMF) for a total of 25 minutes with shaking to remove the Fmoc protecting group. The piperidine was subsequently washed off the resin using DMF and dichloromethane (DCM). Coupling of the next amino acid was achieved by using an excess of that amino acid mixed with a coupling reagent, 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylenaminium hexafluorophosphate (HCTU) and N,N-diisopropylethylamine (DIEA) in a solution of DMF. The coupling reaction was run for 45 minutes with shaking. Following coupling, the coupling reagents and excess amino acid are washed off in a series of rinses using DMF followed by DCM. These steps, starting with deprotection, are repeated until the peptide chain reaches the desired length. The next step is cleavage from the wang resin. The cleavage was accomplished with 10 mL 95% trifluoroacetic acid, 2.5% triisopropyl silane, and 2.5% deionized H\textsubscript{2}O for 90 minutes with shaking. The trifluoroacetic acid here also serves to remove the Boc protecting group from the side chain.
lysine and its homologues. Following the cleavage step, peptides are now located in solution as opposed to being bonded to the resin surface. They are eluted through the filter of the syringe into a 100 mL round bottom flask and allowed to crash out in the freezer overnight in 30 mL of anhydrous ether. Typically the ether solution becomes cloudy once sufficient peptide has crashed out of solution. The ether was then placed into test tubes and centrifuged to produce a pellet of peptide at the bottom of the tube. The supernatant is discarded and this process is repeated until all of the ether has been centrifuged and the supernatant discarded. The peptide coating the bottom of the test tube is then dissolved in 50:50 methanol:water for storage in the fridge prior to mass spectrometry analysis.

2.3.2 Mass Spectrometry Fragmentation of Tripeptides

Immediately prior to analysis, 1% acetic acid was added to a dilution of the peptide solution to protonate the peptides. Peptides were ionized by ESI and analyzed by a ThermoFinnigan TSQ triple quadrupole mass spectrometer. Tandem mass spectrometry was used to isolate peptides and collect fragmentation data at various voltages using argon as the collision gas. This voltage was adjusted either up or down to give a good representative fragmentation spectrum with the parent peak still present but with fragments also present with significant intensity. Using the mobile proton model of peptide fragmentation, one can easily predict the expected masses of specific fragment from a given peptide. Expected masses of typical CID fragments (b and y type ions) were predicted and compared to the actual fragmentation spectra.
Chapter 3 – Proteomics Studies

3.1 Proof of Concept Experiments with Bovine Serum Albumin (BSA)

The first experiments with BSA were performed with a Methanol:Water gradient. We started with a 40 minute gradient with the following concentrations of methanol in water (with 1% formic acid).

10 minutes – 5% MeOH
5 min – 5-35% MeOH
5 min – 35-55% MeOH
10 min – 55-95% MeOH
5 min – 95% MeOH
5 min – 5% MeOH

This gradient gave rather poor separation, with most of the peptides eluting at the end of the gradient. The chromatogram is shown below in Figure 3.1.

From this chromatogram and the fragmentation data recorded during the data-dependent scan, SEQUEST identified bovine serum albumin (BSA) with 46.13% coverage. There were a total of 24 unique peptides with statistically significant confidence. The most common
modification was, as expected, carbamidomethylation of cysteine (the result of the alkylation step of the tryptic digest).

To get a better idea of where identified peptides eluted, we plotted protonated peptide mass v. retention time, as shown in Figure 3.2 below.

![Figure 3.2](image)

We used a variation of this gradient to run a sample of msmeq. The gradient used was:

- 10 min – 5% MeOH
- 5 min – 5-35% MeOH
- 5 min – 35-55% MeOH
- 10 min – 55-95% MeOH
- 5 min – 95% MeOH
- 5 min – 5% MeOH

The resulting chromatogram is shown in figure 3.3.
From this data, SEQUEST identified 3 proteins. These were Propionyl-CoA carboxylase betachain (3 peptides, 7.54% coverage), 3-hydroxybutyryl-CoA dehydrase (1 peptide, 3.44% coverage), and transcriptional regulator, GntR family (1 peptide, 6.25% coverage). The coverage and separation of these peptides is poor so we switched back to BSA to optimize the gradient and other protocols. It is evident from the multiple proteins identified that the gel electrophoresis step needs to be optimized better because ideally we should only see one protein per gel band. From here, we decided to use a step gradient where we stepped the concentration of MeOH up 10% every 10 minutes. The gradient used was as follows:

- 20 min – 5% MeOH
- 10 min – 15% MeOH
- 10 min – 25% MeOH
- 10 min – 35% MeOH
- 10 min – 45% MeOH
- 10 min – 55% MeOH
- 10 min – 65% MeOH
- 10 min – 75% MeOH
- 10 min – 85% MeOH
- 10 min – 95% MeOH
- 20 min – 5% MeOH
The chromatogram for the predigested BSA at this gradient is shown below in Figure 3.4.

From this chromatogram and the associated peptide fragmentation data, SEQUEST was able to identify BSA with 80.89% coverage. This included 29 unique peptides with statistically significant confidence. Once again, carbamidomethylation of cysteine was the most common modification. To get a better idea of when identified peptides eluted the protonated mass was again plotted against retention time, as shown in Figure 3.5.
This separation looks better, with retention times better separated but the chromatogram still shows significant intensity near the end of the chromatogram, and a full gradient flush was required to fully clean peptides off the column. From these results and the experimental protocols outlined in the majority of proteomics papers, we decided to switch our less polar solvent from methanol to acetonitrile (ACN). Acetonitrile is significantly less polar than methanol, so it should provide better separation, especially for nonpolar peptides. We switched solvents, flushed the new mixture through the column to condition it, and then ran the same stepped gradient with ACN instead of methanol. The chromatogram is shown in figure 3.6.
This separation looks significantly better and much more like what one would expect from peptide separation. From this data, SEQUEST was able to identify BSA with 73.15% coverage. 31 unique peptides were identified with significant confidence. Carbamidomethylation of cysteine residues was very common in these fragments as well as N-terminal addition of an acetyl group. The protonated masses were again plotted against retention times to get a better idea of where identified peptides actually eluted, shown in figure 3.7.
From this graph, we can see that the vast majority of peptides eluted between approximately 40 minutes and 60 minutes. These retention times correspond to concentrations of 35%-45% ACN. From this separation, we can see that a gradient as long as 130 minutes is not necessary for good separation. From here, we cut this gradient in half as follows for a 65 minute gradient.

10 min – 5% ACN
5 min – 15% ACN
5 min – 25% ACN
5 min – 35% ACN
5 min – 45% ACN
5 min – 55% ACN
5 min – 65% ACN
5 min – 75% ACN
5 min – 85% ACN
5 min – 95% ACN
5 min – 5% ACN
The chromatogram for this run is shown below in Figure 3.8.

Figure 3.8

From this data, SEQUEST identified BSA with 58.81% sequence coverage from 22 unique, statistically significant identified peptides. The protonated mass of identified peptides were once again plotted against their retention times to give the graph below (Figure 3.9).

Figure 3.9
From this graph, it seems that more peptides eluted later in the gradient. The gradient was therefore adjusted to account for this. The next H$_2$O: ACN gradient was as follows:

- 5 min – 5% ACN
- 5 min – 5-25% ACN
- 10 min – 25-45% ACN
- 50 min – 45-85% ACN
- 5 min – 95% ACN
- 5 min – 5% ACN

The chromatogram for predigested BSA is shown below in Figure 3.10.

![Chromatogram for predigested BSA](F:\CBP Proteomics\BSA_predig_80minGrad_032013.RAW)

**Figure 3.10**

From here, we can see peptides eluting towards the middle of the gradient, as one would predict. To get a better idea of elution times for identified peptides we once again graphed protonated mass against retention time, shown in the graph in Figure 3.11.
From this data, we can see that peptides eluted between approximately 15 minutes and 40 minutes. From the gradient, it was determined that 40 minutes corresponds to approximately 60% ACN. The gradient was altered to significantly increase this less polar portion. The new gradient was programmed as follows for a total of 90 minutes.

- 5 min – 5% ACN
- 5 min – 5%-25% ACN
- 10 min – 25%-35% ACN
- 50 min – 35%-65% ACN
- 10 min – 65%-85% ACN
- 5 min – 95% ACN
- 5 min – 5% ACN
The chromatogram from this gradient is shown below in Figure 3.12.

This separation is slightly better than that of the previous gradient, but still not quite ideal. Protonated mass was plotted against retention time for identified peptides giving the graph below (Figure 3.13).
From this, we see that peptides eluted between 15 minutes and approximately 50 minutes. Plotting the gradient in this region as ACN percentage vs time, we can get a better idea of the specific concentrations of these retention times (Figure 3.14).

![Figure 3.14](image)

From this, it was determined that 15 minutes corresponds to 30% ACN and 50 minutes corresponds to 53% ACN. Following this run, a blank injection of solvent was made and the full length gradient was run to determine how well the original gradient washed the peptides out of the column. The chromatogram is shown in Figure 3.15.
From the chromatogram, we can see that the overall signal to noise ratio is much lower, in other words, our analyte intensities are significantly lower in comparison to background noise. SEQUEST identified BSA with 31.8% sequence coverage.
Chapter 4 Tripeptide Fragmentation:

4.1 AAX Peptides

Synthesized tripeptides belong to one of three different classes depending on where the lysine homologue is located. The first class of peptides analyzed were those with the lysine homologue located at the C-terminus of the peptide. Figures 4.1-4.4 show the fragmentation spectra of the four AAX peptides.

Figure 4.1 AAK - CID Energy 50%.
Figure 4.2 AAO – CID Energy 35%

Figure 4.3 AA-Daba – CID Energy 35%
For the peptides AAK, AAO, AA-Daba, and AA-Dapa, we saw predominantly b and y type ions, as is expected with low energy fragmentation of peptides. We also saw some peaks that do not correspond to the peptide we were fragmenting. These peaks correspond to sequence-scrambled fragments. Specifically, these peaks correspond to sequence-scrambled fragments from a peptide with the lysine homologue located at the N-terminus of the peptide. These cyclized peptides also gave predominantly b-type fragment ions. This makes intuitive sense because the charge remains on the N-terminus in b-type ions and for these rearranged peptides, the N-terminus would include the most basic amino acid residue on the peptide. Of these sequence scrambled peaks, the relative intensity is greatest in AAO, with the scrambled b₁ peak having an intensity greater than 50% compared to the parent peak at 100%. The b₁ peak corresponds to a fragment with only the amino acid ornithine present. This suggests that scrambled b fragment maybe more stable relative to its parent peptide peak than the other
scrambled fragments are to their respective parents. The length of the ornithine side chain could be responsible for this phenomenon. Ornithine’s side chain could possibly be just the right length to make an intramolecular hydrogen bond which would stabilize the extra proton. It is also important to note that for these four tripeptides, the b₃ peak, if seen, could either come from the original peptide, or the sequence scrambled peptide because the m/z value is the same regardless. AA-Daba and AA-Dapa both showed sequence scrambling to lesser degrees than the other two peptides in the series.

AAK was different from the other peptides in this series in that it required significantly more energy to give a representative fragmentation pattern. This could suggest some sort of intramolecular hydrogen bond formation in the original peptide which would increase the energy required for the mobile proton to move to a location along the backbone thereby inducing fragmentation.

4.2 AXA Peptides

Figures 4.5-4.8 show the fragmentation spectra of the AXA peptides.
Figure 4.5 AKA – CID Energy 35

Figure 4.6 AOA – CID Energy 35%
Figure 4.7 A-Daba-A – CID Energy 35%

Figure 4.8 A-Dapa-A – CID Energy 35%
The peptides AKA, AOA, A-Daba-A, and A-Dapa-A gave the expected CID b and y type ions. In addition, they also gave sequence scrambled fragment ions. Once again, these scrambled ions appeared to be from the sequence scrambled peptide with the lysine homologue located at the N-terminus. In this case, the sequence scrambled b₁ ions from AKA and AOA were the most intense relative to their parent peptides, likely a result of their high proton affinities. Once again, the sequence scrambling was seen in Daba and Dapa containing peptides to a significantly lesser degree, most likely the result of their decreased side chain length and decreased proton affinity. In these peptides we also saw loss of neutral fragments like water (H₂O) and ammonia (NH₃) which correspond to losses of 18 and 17 Da respectively.

4.3 XAA Peptides

Figures 4.9-4.12 show the fragmentation spectra of the XAA peptides.

![Figure 4.9 KAA – CID 35%](image)
Figure 4.10 OAA – CID 35%

Figure 4.11 Daba-AA – CID 50%
For the peptides KAA, OAA, Daba-AA, and Dapa-AA, no non-sequence ions were detected. As before, we see mostly b-and y-type ions (as is typical with CID of peptides) as well as the loss of neutral species like NH₃ and H₂O.

This information, combined with the fragmentation spectra for the other eight tripeptides, suggests that peptides and their fragments are most stable with the most basic amino acid at the N-terminus. As the sequence scrambling mechanism requires multiple collisions and as CID is a relatively low-energy way to induce fragmentation, it makes sense that as these ions undergo a series of low-energy collisions with the inert collision gas, they are able to rearrange and form the most stable species. In the XAA peptides as well as the sequence scrambled peptides, we saw b-type ions with the most relative intensity in the species with higher proton affinity amino acids, specifically lysine and ornithine. Another trend that appears to be highly related to proton affinity, is the complexity of the spectrum. As the proton affinity of the lysine homolog...
decreased, the complexity of the spectrum increased. This makes sense as the higher proton affinity residues are better able to compete with other fragments as well as contaminants for the extra protons and are, therefore, more likely to show up in the mass spectrum with higher intensity.

These results are very important for proteomics applications. Most bottom-up proteomics experiments use trypsin to digest proteins into small peptides to be analyzed by tandem mass spectrometry. Peptides with lysine residues in locations other than the N-terminus could result in sequence scrambled fragments and false positive identifications. As previously discussed, trypsin cleaves C-terminal to lysine and arginine residues. This results in peptides with lysine or arginine at the C-terminus. If these sequence scrambling results can be extended to other small peptides containing lysine, then it is possible that sequence scrambling is leading to misidentification of tryptic peptides. Trypsin is also not always efficient, missed cleavages are common. These result in lysine or arginine residues being located in the middle of a peptide. These could also result in some sort of sequence scrambling and misidentification.
Chapter 5 – Conclusions

Our mass spectrometry-based proteomics experiments yielded optimized protocols for protein identification. The use of a predigested sample of BSA allowed us to focus on optimizing the gradient. We learned that a mixture of acetonitrile (ACN) and water was superior to a mixture of methanol (MeOH) and water. This is likely because of the decreased polarity of ACN which allows more hydrophobic proteins to elute through the nonpolar C$_{18}$ column. Further studies should yield results for *Mycobacterium smegmatis*. Once proof of concept experiments with *M. smegmatis* have been performed, the ultimate goal will be to study the time course of infection by Larva and CrimD as well as any overall proteome differences between wild type *M. smegmatis* and phage infected *M. smegmatis*.

The peptide fragmentation studies provide insight into the relationship between side chain length and fragmentation properties. Our data suggests that certain side chain lengths form stronger intramolecular hydrogen bonds than others which result in higher energies being necessary to generate fragmentation spectra. This data also provides insight into the relationship between amino acid sequence location and fragmentation properties. We saw sequence scrambling fragmentation in the peptides with the lysine homologue in the middle and at the C-terminus. No sequence scrambling was seen in peptides with the lysine homologue at the N-terminus. This suggests that peptides may be more stable with the more basic residue at the N-terminus. Computational studies into the mechanisms of fragmentation and sequence scrambling for these tripeptides are underway.

Our peptide fragmentation studies have implications for proteomics studies. Proteomics studies employ the rules of peptide fragmentation to make peptide identifications and therefore protein identifications. The better these rules are understood, the more accurate identifications
will be. Sequence scrambling could be especially problematic for proteomics studies because it would result in mis-identification of peptides. This could generate false positive identifications or could potentially result in a peptide not being identified at all. With the field of proteomics becoming rapidly more important, studies like these are necessary to make protein identifications as accurate as possible.
29 Xcalibur Proteome Discoverer Version 1.3 User Guide