

**Analysis of Protamines and the Investigation of Concentration
Effects of an Immobilized Aspergillopepsin I Reactor**

Robert Anthony D'Ippolito
Toms River, New Jersey

B.S. Chemistry, West Chester University of Pennsylvania, 2014

A Dissertation Presented to the Graduate Faculty of the University of Virginia in
Candidacy for the Degree of Doctor of Philosophy

Department of Chemistry

University of Virginia
May, 2020

Abstract

The following dissertation details three projects utilizing mass spectrometric methods for protein analysis. Each project presented different challenges that required diverse methodology for successful analysis.

The first project identified protamines, which are small arginine rich proteins, from the common liverwort, *Marchantia polymorpha*. Protamines are essential for the condensation of DNA in spermatozoa. The origins of protamines remain elusive. Preliminary evidence detail that protamines evolved from histone H1 through a protamine-like intermediate. However, their evolutionary ancestry has only been loosely related to H1. This work provided definitive evidence linking protamines to their H1 ancestors. Additionally, a novel post translational modification was discovered. The C-termini of protamines were modified with the addition of di-aminopropane.

The second project further developed a novel protein digestion technique utilizing the nonspecific protease Aspergillopepsin I. The protease was immobilized to solid support, enabling digestion on the timescale of seconds to generate hundreds of overlapping peptides. These peptides can be further used to provide unambiguous primary structure analysis of a protein. However, the specific concentration of 0.2 $\mu\text{g}/\mu\text{L}$ was needed to provide the expected results using this system. Multiple protein standards were used to fully characterize the effect of concentration, clearly showing that observed cleavages increased as starting concentration decreased. Additives were used to try to reverse these effects, which led to the discovery of protamines from salmon as an effective competitive inhibitor. Protamine treatment enabled the digestion of proteins at concentrations an order of magnitude lower than previously possible.

The final project incorporated the knowledge gained from the concentration effects of the enzyme reactor and applied them to bispecific antibody analysis. These engineered protein complexes represent a novel avenue for therapeutic molecules and require extensive characterization of their structure. Here, two different classes of bispecific antibodies were digested by the Aspergillopepsin I enzyme reactor to obtain near complete sequence coverage and extensive characterization of specific features of the molecules using novel methodologies.

Acknowledgements

Reaching this point in my career would not be possible without the guidance and support I received from Don and Jeff. I am so grateful for all of the time that you invested in me to improve myself as a scientist. Afternoon coffee discussing the results from my experiments in the few days prior and talking through my thoughts really helped keep my drive focused while keeping you both up to date on the massive number of projects I was juggling. Thank you both for helping shape the scientist that I am today. I am and will always be forever grateful.

Additional gratitude to Dina for always having an open ear for me. I appreciate all of the times when I would come into your office and babble about my day's work. I appreciated that you always welcomed me with a smile on your face, positivity radiating from your office, and a jar full of candy. I also want to thank you for working and coding all of the program ideas that came to mind over my career. Portions of this dissertation were made better directly through your coding.

Graduate school would not be complete without the cohort of the Hunt group. I thank you all for being amazing colleagues. I want to especially thank (in order of appearance in my Hunt Lab journey) Stephanie Miller Lehman for convincing me to join the Hunt Lab, Mark Ross for my initial training of protein mass spectrometry as well as an outlet about baseball, Elizabeth Duselis for training me on whole protein work and the quirks of Velos, Josh Hinkle for training me on the enzyme reactor as well as being my go-to person for all of my problems, Keira Mahoney for being the best cubicle mate to talk through our work on a daily basis, and Maria Panepinto for being an amazing mentee to carry on the work presented here.

I also want to show my appreciation to my defense committee of Dr. Venton, Dr. Hsu, Dr. Grisham, and Dr. Brautigan. Thank you for taking time out of your busy days to read, critique, and participate in my defense.

Last, but certainly not least, I would like to thank the world's most supportive wife, Jenna. After long days in the lab, coming home to you was exactly what I needed, particularly when experiments did not go as planned. I appreciate all of the love and support you have given me over the past five years. I can't thank you enough (good thing you are stuck with me for a while).

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	vii
List of Tables	x
Abbreviations	xi
1 Introduction to the Dissertation	1
1.1 Overview.....	1
1.2 Instrumentation for Biological Mass Spectrometry	2
1.2.1 Reverse Phase High Performance Liquid Chromatography.....	3
1.2.2 Electrospray Ionization	4
1.2.3 Quadrupole Mass Filter	5
1.2.4 Ion Traps.....	7
1.2.5 Orbitraps	9
1.2.6 Collisional Dissociation	12
1.2.7 Electron Transfer Dissociation	15
1.3 Protein Preparation Methods.....	19
1.3.1 Trypsin Digestion.....	19
1.3.2 No Enzymatic Digestion.....	20
1.3.3 Limited Digestion.....	22
1.4 References	24
2 Determination of Protamines from <i>Marchantia Polymorpha</i>	29
2.1 Abstract	29
2.2 Introduction.....	29
2.2.1 Histones and Protamines.....	29
2.2.2 Evolutionary Theory of Protamines	32
2.3 Materials.....	35
2.4 Methods.....	36
2.4.1 Protamine Extraction from Antheridiophores	36
2.4.2 Liquid Chromatography and Mass Spectrometry.....	37
2.4.3 Determination of C-Terminal Di-aminopropanelation by MS ³	37
2.4.4 Data Analysis	38
2.5 Results and Discussion	38
2.5.1 Initial Identification of Protamines.....	38
2.5.2 Discovery of a Novel C-Terminal PTM	40
2.5.3 Confirmatory Evidence of the Origins of Protamines	44
2.6 Conclusions	46

2.7	References	46
3	Concentration Effects of an Immobilized Aspergillopepsin I Enzyme Reactor	50
3.1	Abstract	50
3.2	Introduction.....	50
3.2.1	Immobilized Enzymes.....	50
3.2.2	Aspergillopepsin I.....	52
3.2.3	Immobilized Aspergillopepsin I.....	54
3.3	Materials.....	59
3.4	Methods.....	61
3.4.1	Aspergillopepsin I Bead Conjugation.....	61
3.4.2	Protein Preparation for Digestion	62
3.4.3	Protein Digestion with Immobilized Enzyme Reactor.....	63
3.4.4	Liquid Chromatography and Mass Spectrometry.....	63
3.4.5	Data Analysis	64
3.5	Results and Discussion	64
3.5.1	Reducing Analyte Concentration Results in a Greater Number of Cleavages by Aspergillopepsin I	64
3.5.2	The Search for a Competitive Inhibitor of Aspergillopepsin I.....	68
3.5.3	Protamine as the Ideal Inhibitor	73
3.5.4	Reproducibility of Protamine Inhibition	77
3.6	Conclusions	82
3.7	References	82
4	Analysis of Bispecific Antibodies Using an Immobilized Aspergillopepsin I Enzyme Reactor	86
4.1	Abstract	86
4.2	Introduction.....	86
4.2.1	Antibodies	86
4.2.2	Bispecific Antibodies.....	89
4.2.2.1	IgG-Like Bispecific Antibodies	90
4.2.2.2	Bispecific T-Cell Engager (BiTE)	92
4.2.3	Previous Work with Immobilized Aspergillopepsin I for Antibody Analysis	93
4.2.3.1	Parallel Ion Parking.....	94
4.2.3.2	Disulfide Bond Localization	99
4.3	Materials.....	101
4.4	Methods.....	103
4.4.1	Aspergillopepsin I Bead Conjugation.....	103
4.4.2	Bispecific Antibody Preparation	103
4.4.3	Bispecific Antibody Digestion with Immobilized Enzyme Reactor.....	104
4.4.4	Liquid Chromatography and Mass Spectrometry.....	104

4.4.5	Data Analysis	106
4.5	Results and Discussion	107
4.5.1	Complete Sequence Coverage of an IgG-Like Bispecific Antibody	107
4.5.2	Using Protamine to Generate Peptides with all CDRs for Further Investigation Using Parallel Ion Parking	114
4.5.3	Complete Sequence Coverage of a BiTE with Disulfide Bond Localization in a Single Chromatographic Analysis.....	118
4.6	Conclusions	127
4.7	References	128
5	Conclusions.....	132
	References	135

List of Figures

Chapter 1: Introduction to the Dissertation	
Figure 1.1: Structure of PLRP-s packing material.....	4
Figure 1.2: Mathieu Stability Diagram for Positive a and q Values.....	6
Figure 1.3: Schematic of Linear Ion Trap.....	7
Figure 1.4: Schematic of C-trap and Orbitrap.....	9
Figure 1.5: Orbitrap Transient to Mass Spectrum.....	11
Figure 1.6: Schematics of Instruments.....	12
Figure 1.7: Simplified Representation of b- and y-ions Formed During Collisional Fragmentation.....	13
Figure 1.8: Example CAD MS ²	13
Figure 1.9: Example HCD MS ²	14
Figure 1.10: Example Phosphopeptide Undergoing Collisional Dissociation.....	15
Figure 1.11: ETD Reaction Scheme.....	16
Figure 1.12: Example ETD MS ²	18
Figure 1.13: Charge Density Effects on ETD.....	19
Figure 1.14: Loss of Combinatorial PTM Example.....	20
Figure 1.15: Example Electrospray Distributions for Intact Protein Analysis.....	21
Figure 1.16: Partial Retention of Combinatorial PTM Example.....	23
Chapter 2: Determination of Protamines from <i>Marchantia Polymorpha</i>	
Figure 2.1: Crystal Structure of a Nucleosome.....	30
Figure 2.2: Histone H1 Interaction with a Nucleosome.....	31
Figure 2.3: Differing DNA Compaction by Histones and Protamines.....	32
Figure 2.4: Hypothesized Reason for Shift from Lysine to Arginine via Frameshift Mutation.....	34
Figure 2.5: 2.5 M Urea/5% Acetic Acid PAGE Gel of Protamines from <i>Marchantia polymorpha</i> Antheridiophores.....	35
Figure 2.6: Chromatogram of Non-retained Proteins.....	39
Figure 2.7: MS ¹ Scans from the Two Distinct Regions in the Chromatogram of Non-retained Proteins.....	39
Figure 2.8: Chromatogram of Non-retained Proteins.....	40
Figure 2.9: Protamine Dissociation Example.....	41
Figure 2.10: HCD MS ³ Scan of the Unknown z ¹ -ion.....	43
Figure 2.11: Proposed Structure of the z ¹ -ion Present at 215.1740 m/z.....	43
Figure 2.12: Structures of Polyamide Molecules.....	44
Chapter 3: Concentration Effects of an Immobilized Aspergillopepsin I Enzyme Reactor	
Figure 3.1: Enzyme Conformations when Bound to a Solid Support.....	51
Figure 3.2: Simplified Depiction of Diffusion Layer.....	52
Figure 3.3: Crystal Structure of Aspergillopepsin I.....	53

Figure 3.4: Amino Acid Sequence of the Active Form of Aspergillopepsin I	53
Figure 3.5: Catalytic Mechanism of Aspartic Proteases	54
Figure 3.6: Primary Amine Conjugation Mechanism with Aldehyde Functionalized Beads.....	55
Figure 3.7: Schematic of General Process of Enzyme Reactor Use.....	56
Figure 3.8: Total Ion Current Chromatograms for Different Digestion Times of Apomyoglobin	57
Figure 3.9: Urea Impact on Aspergillopepsin I Digestions.....	58
Figure 3.10: Relative Cleavage Specificity of Immobilized Aspergillopepsin I.....	59
Figure 3.11: ~1 s Digestion Profiles of Apomyoglobin at Various Concentrations.....	65
Figure 3.12: Comparing Digestion Profiles of Apomyoglobin at Various Concentrations to a Standard Digestion.....	67
Figure 3.13: Chromatogram of Tracked Peptides	70
Figure 3.14: Ubiquitin Treated Apomyoglobin Digestion	71
Figure 3.15: β -Lactoglobulin Treated Apomyoglobin Digestion	71
Figure 3.16: BSA Treated Apomyoglobin Digestion.....	72
Figure 3.17: RF + KG Treated Apomyoglobin Digestion	73
Figure 3.18: KKKKK Treated Apomyoglobin Digestion	73
Figure 3.19: Protamine Treated, Diffusion Limited Apomyoglobin Digestions.....	74
Figure 3.20: Protamine Treated, Catalytic Limited Apomyoglobin Digestions.....	75
Figure 3.21: Bar Graph Representation of Selected Peptides from Apomyoglobin Digestions	76
Figure 3.22: Protamine Only Digestion.....	76
Figure 3.23: Effect of Using Organic Rinse.....	77
Figure 3.24: Chromatogram of All Proteins Tested for Protamine Treatment	78
Figure 3.25: Untreated Protein Digestion of Multiple Proteins.....	79
Figure 3.26: Treated Protein Digestion of Multiple Proteins	80
Figure 3.27: Bar Graph Representation of Selected Peptides from Protein Digestions	81
Figure 3.28: Trypsinogen Digestion when Treated with Protamine	82
Chapter 4: Analysis of Bispecific Antibodies Using an Immobilized Aspergillopepsin I Enzyme Reactor	
Figure 4.1: Structure of IgG	87
Figure 4.2: Structures of Common Antibody Glycosylation Trees	88
Figure 4.3: Structures of Bispecific Antibodies	90
Figure 4.4: Depiction of CrossMab	91
Figure 4.5: Structure of a BiTE.....	92
Figure 4.6: Extended ETD Reaction Times Result in the Generation of Internal Fragments.....	95
Figure 4.7: Example pipETD Performed on the Thermo Orbitrap Elite	96

Figure 4.8: IIPT Improves the Ability to Distinguish Fragment Ions	98
Figure 4.9: Disulfide Shuffling	100
Figure 4.10: Sample Preparation of the Common Light Chain bsAb Prior to Digestion	103
Figure 4.11: Data-Dependent Tiered Decision Tree	105
Figure 4.12: Total Ion Current Chromatogram of an ~1s Digestion of Elicizumab	108
Figure 4.13: Total Ion Current Chromatogram of an ~1s Digestion of Elicizumab Treated with Protamine	109
Figure 4.14: Histogram of Peptides Identified in the Digestion of Elicizumab Grouped by Change in Abundance	110
Figure 4.15: Scatter Plot of the Identified Peptides in the Digestions of Elicizumab versus the Molecular Weight.....	111
Figure 4.16: Peptide Map Used for Fragment Ion Analysis from the Digestion of Elicizumab	112
Figure 4.17: Observed Fragment Ions from the Digestion of Elicizumab.....	113
Figure 4.18: Total Ion Current Chromatogram of an ~800ms Digestion of Elicizumab Treated with Protamine	114
Figure 4.19: Peptides Targeted for Further Investigation by pipETD.....	115
Figure 4.20: pipETD of a Large Peptide Containing all 3 CDRs	116
Figure 4.21: Sequence Coverage of the Four Peptides of Interest.....	117
Figure 4.22: Total Ion Current Chromatogram of the ~1.9s Digestion of Nonreduced BiTE.....	119
Figure 4.23: Example Disulfide Peptide.....	121
Figure 4.24: Manually Annotated Peptide Map from the Nonreduced BiTE Digestion	122
Figure 4.25: Collisional Sequence Coverage Obtained from the Analysis of a Nonreduced BiTE.....	123
Figure 4.26: ETD Sequence Coverage Obtained from the Analysis of a Nonreduced BiTE.....	124
Figure 4.27: Composite Sequence Coverage Obtained by Collisional Fragmentation and ETD from the Analysis of a Nonreduced BiTE	125
Figure 4.28: Pictorial Representation of Identified Disulfide Bonds	125
Figure 4.29: Complex Disulfide Bond Network from Bovine Trypsinogen.....	127

List of Tables

Chapter 2: Determination of Protamines from <i>Marchantia Polymorpha</i>	
Table 2.1: Table of Identified Protamines	45
Chapter 3: Concentration Effects of an Immobilized Aspergillopepsin I Enzyme Reactor	
Table 3.1: Table of Tracked Peptides.....	69
Table 3.2: Table of Protamines Observed in Salmine	74
Chapter 4: Analysis of Bispecific Antibodies Using an Immobilized Aspergillopepsin I Enzyme Reactor	
Table 4.1: Sequence Coverage for the pipETD followed by IIPT Data of the Four Peptides of Interest.....	117
Table 4.2: Table of Search Times Required for a Nonspecific Disulfide Analysis ...	126

Abbreviations

μ - micro (1x10⁻⁶)

AC – Alternating Current

AGC - automated gain control

A, Ala - alanine

API atmospheric pressure ionization

C, Cys - cysteine

°C – Degrees Celsius/Centigrade

C₁₈ - octadecylsilane

CID - collision induced dissociation

D, Asp - aspartic acid

Da - Dalton(s)

DC - direct current

E, Glu - glutamic acid

ECD - electron capture dissociation

ESI - electrospray ionization

ETD - electron transfer dissociation

ETnoD - nondissociative electron transfer

f femto - (1x10⁻¹⁵)

F, Phe - phenylalanine

Fab - Fragment antigen binding

Fc - Fragment crystallizable

FT - Fourier transform or high resolution

g - gram(s)

G, Gly - glycine

H, His - histidine

HPLC - high performance liquid chromatography

I, Ile - isoleucine

i.d. - inner diameter

IgG1 - Immunoglobulin G (subclass 1)

IIPT - ion/ion proton transfer

IT - ion trap or low resolution

ITCL - ion trap control language or instrument control software

K, Lys - lysine

L - liter(s)

L, Leu - leucine

LC - liquid chromatography

LIT – Linear Ion Trap

LTQ - linear trap quadrupole (Thermo Scientific QLT)

m - milli (1×10^{-3})

m - meter(s)

M - molar

M, Met - methionine

Min - minute(s)

Mol - mole(s)

ms - millisecond(s)

MS - mass spectrometry

MS¹ - full MS spectrum

MS/MS, MS² - tandem mass spectrometry or tandem mass spectrum

MW - molecular weight

m/z - mass-to-charge ratio

n - nano (1×10^{-9})

N, Asn - asparagine

o.d. outer diameter

p - pico (1×10^{-12})

P, Pro - proline

pipETD - parallel ion parked electron transfer dissociation

PLRP – Polymeric Reverse Phase (polystyrene/divinylbenzene)

ppm - parts per million

PTM - post translational modification

Q, Gln - glutamine

R, Arg - arginine

RF - radio frequency

S, Ser - serine

SF₆ - sulfur hexafluoride

T, Thr - threonine

TIC - total ion current/chromatogram

UV - ultraviolet

V - volt(s)

V, Val – valine

V_H – Heavy Chain Variable Region

V_L – Light Chain Variable Region

W, Trp tryptophan

Y, Tyr tyrosine

Introduction to the Dissertation

1.1 Overview

This dissertation details the analysis and characterization of a unique set of protein samples using mass spectrometry. Following an introduction to the field of protein analysis by mass spectrometry in Chapter 1, Chapter 2 describes the identification of protamines from the plant *Marchantia polymorpha*. Protamines are highly basic proteins in sperm cells consisting of primarily arginine residues that replace histones toward the end of spermatogenesis to compact DNA further. Although studies have implicated that animal protamines evolved from a histone H1 precursor, no direct evolutionary linkage has been identified. This chapter will describe how protamines from *Marchantia polymorpha* were identified using electron transfer dissociation and the discovery of a novel post-translational modification found on the C-terminus.

Chapter 3 details the continuing development of a novel system for digesting proteins utilizing Aspergillopepsin I immobilized onto solid supports. Typical sample preparation for protein analysis involves enzymatic digestion to improve detection and fragmentation of peptides. Common proteases, such as trypsin and chymotrypsin, cleave at specific sites along the protein backbone, but are limited in their ability to obtain comprehensive protein coverage in a single analysis due to the accessibility of the enzyme to the substrate in the native form and the loss of small, hydrophilic peptides. Our lab has demonstrated the ability to quickly and efficiently generate a comprehensive group of peptides using the nonspecific protease Aspergillopepsin I, a unique enzyme that is functional in acidic and denaturing conditions. Previous work has demonstrated that immobilizing Aspergillopepsin I allows for highly reproducible digestions under a well-defined set of working conditions. However, the efficiency of digestion using Aspergillopepsin I is affected by the concentration of protein in the sample; higher protein concentration is required to obtain ideal peptide sizes for analysis. This chapter will explore the conditions that characterize how the analyte concentration affects the digestion by Aspergillopepsin I.

Chapter 4 applies the analyte concentration effects from the immobilized Aspergillopepsin I enzyme reactor, characterized in Chapter 3, to the analysis of two different bispecific antibodies. Bispecific antibodies are currently of significant interest as possible therapeutics since they target two different epitopes. Two bispecific antibody systems are discussed in this chapter. First, an IgG-like bispecific antibody with common light chains but different heavy chains was analyzed for sequence identification. Second, a Bi-specific T-Cell Engager (BiTE) antibody, a class of bispecific antibodies that contain two variable regions expressed as a single amino acid chain, was analyzed for sequence identification as well as disulfide bond localization.

1.2 Instrumentation for Biological Mass Spectrometry

Each individual protein has a specific role within the cell which is dictated by the defined order of amino acids, also known as the primary structure. Specific amino acid residues can be further modified by other proteins (1). These variable or static post-translational modifications (PTMs) further direct protein function and are equally important as the protein's primary structure (1). For example, antibodies are modified by various glycan trees which determines their solubility, stability, and half-life (2). The characterization of all PTMs is necessary to obtain the complete picture of protein function.

Characterizing proteins and their associated PTMs is challenging. A specific protein represents a miniscule fraction of the total protein content within each cell. It is challenging to identify every protein with their respective modifications within a given analysis (3). Instrumentation with high sensitivity is necessary to accurately identify and characterize a specific protein in addition to discriminating the primary structures of other proteins present. Mass spectrometry (MS) has become an invaluable tool for this purpose (1).

Mass spectrometry is the analytical technique of analyzing and dissociating and detecting gaseous ions based on their mass to charge ratios (m/z) (4–6). Generally, the mass spectrometer is broken into three main components. The first is the ion source, a method of generating the gaseous ions necessary for analysis, the second is the mass analyzer, a device that measures the m/z of the gaseous ions, and the third is the detector, a device that measures the signal of the

gaseous ions (6). Importantly, dissociation of the analyte ions is also feasible by the mass spectrometer to determine structural information (7). These highly sensitive components are used to create, measure, and detect species at the attomole level (8). Each individual component will be discussed in greater detail in the proceeding sections as pertained to Thermo instrumentation.

1.2.1 Reverse Phase High Performance Liquid Chromatography

Although a pure, single protein sample can be analyzed via direct infusion to the mass spectrometer, most protein samples consist of a complex mixture. Such complexity requires a separation technique compatible with MS prior to analysis to minimize the number of unique species being analyzed. Reverse phase high performance liquid chromatography (RP-HPLC) is used to resolve individual proteins/peptides in this dissertation. (9, 10).

Reverse phase chromatography separates analytes by their relative hydrophobicity (10–12). A nonpolar stationary phase is immobilized to beads and packed within a column. Analytes are then loaded onto the column in polar, aqueous solution. The hydrophobic regions of the analytes are preferentially attracted to the nonpolar stationary phase. The percentage of polar, organic solvent is increased at a specified rate to form a gradient and remove the analytes from the stationary phase. Analytes partition to the mobile phase when the affinity for the mobile phase is greater than its affinity for the stationary phase. Analytes are released from the stationary phase, concentrate into a band, and elute from the column. However, the analyte band will broaden into a peak from the discrete band due to differences in path length each analyte molecule takes through the column (Eddy diffusion), longitudinal diffusion, and resistance to mass transfer (10). The hydrophobicity of a peptide is determined by its primary structure (13).

A polystyrene, divinylbenzene stationary phase (PLRP-s) is used throughout this dissertation (Figure 1.1) in conjunction with the mobile phases of 0.3% formic acid in water (Solvent A) and 72% ACN, 18% IPA, 10% water, 0.3% formic acid (Solvent B). The stationary phase is packed in silica capillaries with inner diameters of 75 μm with flow rates on the order of ~ 100 nL/min (8, 14).

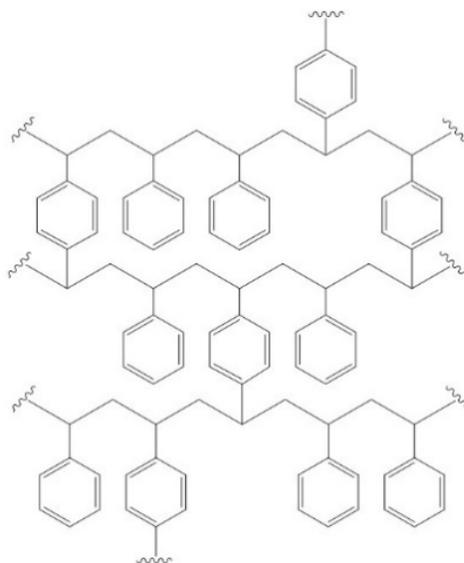


Figure 1.1. Structure of PLRP-s packing material.

1.2.2 Electrospray Ionization

The peptides in the aqueous phase need to be brought into the gas phase for MS analysis. Electrospray ionization (ESI), a soft ionization technique, is used to achieve this goal directly after the peptides are separated to minimize sample handling and analysis time (15–18).

An electrospray tip, which closely resembles a needle, is integrated to the LC column directly following the reversed phase packing material. A high voltage (+2-4 kV) is applied to the LC waste line, creating a positively charged solution while minimizing the possibility of electrochemical reactions occurring with the analyte ions (8). The solvated analytes spray from the tip as positively charged droplets containing multiple analyte ions per droplet and are attracted to the inlet capillary of the mass spectrometer at ground. The charged droplets are then exposed to high temperatures (~250-275°C) within the capillary to facilitate solvent evaporation. This causes the size of the droplets to decrease thus increasing the charge density of the droplets. Once the repulsive Coulombic forces surpass the surface tension of the droplet (the Rayleigh limit), the droplet splits into multiple smaller droplets (18). This process repeats itself until only a single analyte ion is present per droplet. The remaining solvent is evaporated while the positive charges contained by

the droplet through the electrospray process remain on the analyte ion, bringing the analyte to the gas phase.

Peptides will compete for the limited number of charges per droplet. The number of amino acids and the gas phase basicity of the peptide influence the number and distribution of charges observed on a given peptide (19–21). Although these factors split the analyte signal of a given peptide over multiple m/z peaks, their differing charge densities prove useful when fragmenting the peptides for sequence analysis. Coupling ESI to a separation method is incredibly useful to reduce the complexity at a given point in time and increase the dynamic range (22).

1.2.3 Quadrupole Mass Filter

The quadrupole mass filter contains four round parallel rods 90° apart that filter ions based on their m/z . AC and DC potentials are applied out of phase to adjacent rods to produce an electric field that guides ions through the device in which the potential felt by the ions varies quadratically from the center of the device (23). The AC potential, also known as the RF potential due to its operation in the radio frequency domain, is rapidly flipped back and forth from positive to negative voltages, forcing the positively charged ions to attract and repel from the rods.

Ions with stable trajectories in the quadrupolar electric field reach the end of the device while unstable ions collide with the rods. Stable ion trajectories within the quadrupolar electric field can be modeled through the canonical form of the Mathieu equation, a second order differential equation that describe changing but cyclic forces, as shown by:

$$(1.1) \frac{d^2u}{d\xi^2} + (a_u - 2q_u \cos 2\xi)u = 0$$

where u represents the displacement of the ion within the 3D plane, ξ represents a time-dependent value pertaining to the RF frequency of the particular instrument, and a_u and q_u represent the trapping parameters (23). The two unitless parameters, a and q , are related to the stability of an ion and trajectory in the x and y planes while traveling in the z direction. These parameters are defined by the DC potential amplitude and RF potential amplitude applied to the rods, respectively, and are calculated by:

$$(1.2) a_x = \frac{8eU}{\left(\frac{m}{z}\right)r_o^2\Omega^2} \quad (1.3) a_x = -a_y$$

$$(1.4) q_x = \frac{-4eV}{\left(\frac{m}{z}\right)r_o^2\Omega^2} \quad (1.5) q_x = -q_y$$

where e represents the charge of an electron, U represents the DC potential amplitude, V represents the RF potential amplitude, m/z represents the ion's mass-to-charge ratio, r_o represents the radius of the quadrupole, and Ω represents the instrument's specific RF frequency (23). Importantly, the m/z of a given ion is inversely proportional to both a and q . Plotting the values of a and q that satisfies Equation 1.1, which defines ion stability within the quadrupolar electric field, gives rise to the stability diagram shown in Figure 1.2 (23, 24).

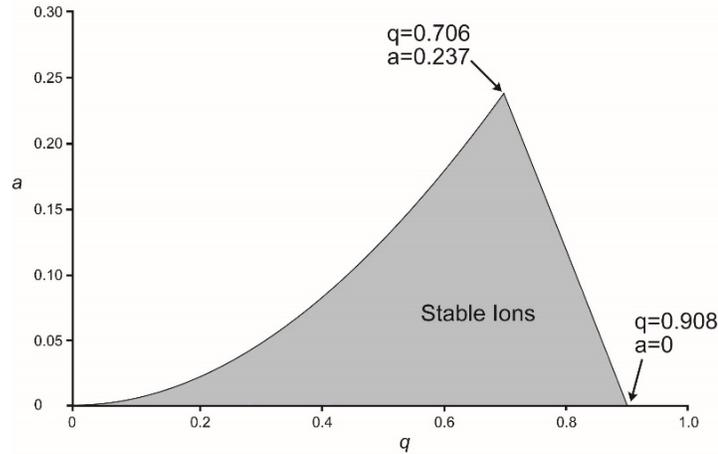


Figure 1.2. Mathieu Stability Diagram for Positive a and q Values. Ions within the shaded region are stable within a quadrupolar electric field.

The simplest operating mode occurs when there is no DC voltage applied, resulting in the a term equaling 0. Therefore, the stability of all ions is reliant on the q value to satisfy Equation 1.1 ($0 < q < 0.908$). The quadrupole acts as a broad window ion guide in this operating mode, allowing a wide range of m/z to maintain stable trajectories (25, 26). The low m/z value for a desired m/z range is given a high, stable q value to set the RF amplitude. Since the value for q decreases as m/z increases with the RF amplitude constant, ions higher in m/z will maintain stability until the point at which the q value reaches 0. Ions that have q values that do not satisfy Equation 1.1 are unstable, collide with the rods, and fail to reach the end of the device.

The quadrupole mass filter isolates a specific m/z range through the application of a static DC voltage to the rods, which introduces the a value for determining stability (25, 26). Ions at certain q values will obtain an unstable trajectory as the a value increases. The ratio of DC/RF amplitudes applied defines the window of specific m/z values that are stable within the device (23). This principle makes a quadrupole the preferred device to isolate a specific m/z range.

1.2.4 Ion Traps

The linear ion trap, developed by Schwartz, Senko, and Syka, is another device that harnesses the ability of a quadrupolar electric field (27). This device uses four parallel rods to trap ions within a quadrupolar electric field. A helium bath gas is used to cool the incoming ions via collisions. This device is split into three sections: front, middle, and back (Figure 1.3). Ions can be transported between sections of the trap through the application of an individual DC offset to each section. The front and back sections of the trap are of equal length at 12 mm while the center section is longer at 37 mm in most models (27). The length of each section directly impacts the maximum number of charges that can be efficiently trapped in each section. The front and back sections can hold $\sim 1\text{-}2 \times 10^5$ charges while the center section can hold $\sim 1\text{-}2 \times 10^6$ charges (27). The newest Thermo instrument, the Orbitrap Eclipse Tribrid, makes use of an extended front section to match the length of the center section. This detail will become important in Section 1.2.7.

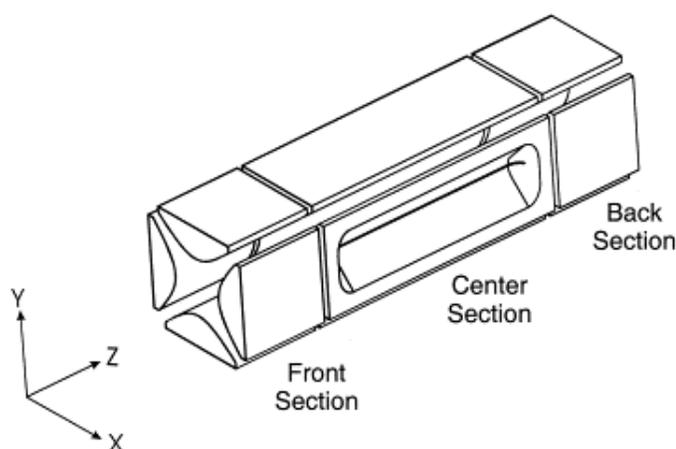


Figure 1.3. Schematic of Linear Ion Trap. The quadrupole ion trap is segmented into three unique sections. Each section has the ability to apply its own distinct DC amplitude. The slit in the center section is used to scan out ions for analysis. Figure was taken from (27).

Ions are trapped within the device in a similar manner to quadrupole mass filters. Stable ions are efficiently trapped with a and q values satisfying Equation 1.1. The DC amplitudes applied to confined ions in a particular section are low enough to be negligible, resulting in an a value (Equation 1.2) of essentially 0. Therefore, the stability of an ion within the ion trap is primarily dependent on the q value of the ion (Equation 1.4).

Additionally, the ions obtain a defined motion in the x and y directions that are independent of one another (23). When combined, the motion, known as the secular frequency of oscillation, can be defined by:

$$(1.6) \omega = \frac{q\Omega}{2\sqrt{2}}$$

where ω represents the secular frequency, and Ω represents the main RF trapping frequency. When substituting Equation 1.4 for q , Equation 1.7 is generated:

$$(1.7) \omega = \frac{-4eV}{2\sqrt{2} \left(\frac{m}{z}\right) r_o^2 \Omega}$$

This equation shows the secular frequency is inversely proportional to the m/z of a specific ion; high m/z ions oscillate in the ion trap at a low rate while low m/z ions oscillate at a higher rate (23).

The ion trap also acts as a low resolution mass analyzer by scanning out the trapped ions. There are slits on the center sections of the x -rods that lead to conversion dynodes and electron multiplier detectors (27). Ions are detected using resonant excitation, which involves applying a supplemental RF voltage to the x -rods that is in resonance with the secular frequency of a small subset of ions (27). This supplemental potential increases the velocity of the ions in resonance in the x -direction, which results in an increase in the oscillation radius of the ion. Once the ions are excited to the point at which the radius of oscillation exceeds the dimensions of the ion trap, the ions are ejected toward the conversion dynode and electron multipliers for mass analysis. The correlation of the electron multiplier response with the supplemental RF potential used to eject the ion determines the ion's secular frequency, and thus its m/z . The supplemental potential is then shifted to excite a different secular frequency for mass analysis. This process is repeated across the mass range to give a mass spectrum. This applied supplemental potential is also used for collisionally activated dissociation (CAD) (to be discussed in Section 1.2.6).

When first implemented, there existed a single linear ion trap to perform both mass analysis and CAD. However, newer Thermo models contain two linear ion traps placed next to one another with a lens separating the traps, referred to as a dual-cell linear ion trap (28). The first trap has a helium pressure of $\sim 7-8 \times 10^{-5}$ torr (the high pressure trap) while the second trap has a helium pressure of $\sim 1-2 \times 10^{-5}$ torr (the low pressure trap). The high pressure trap enables a more efficient collisional cooling mechanism when introducing ions to the trap in addition to reducing the activation time needed for CAD (28). This trap also maintains a more constant quadrupolar field since there are no slits in the x-rods for mass analysis. The low pressure trap, which still contains the slits in the x-rods, benefits from the lower pressure by decreasing the time necessary to acquire a full scan which in turn enables a higher number of scans taken in the same time frame (28).

1.2.5 Orbitraps

Although the linear ion trap is beneficial for multiple low resolution scans in a short time, certain experiments require a more accurate mass analyzer. The high resolution accurate mass analyzer used within Thermo instruments is the Orbitrap, which was invented by Makarov in 2000 (Figure 1.4) (29–33).

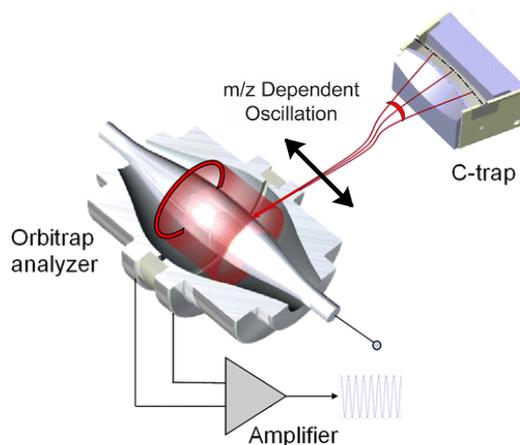


Figure 1.4. Schematic of C-trap and Orbitrap. Ions are cooled and stored in the C-trap prior to off-center injection to the Orbitrap. Figure modified from reference (32).

Ions are trapped in the C-trap prior to analysis by the Orbitrap (Figure 1.4) (34). This trap has a curved structure and is filled with nitrogen gas to collisionally cool the ions after transfer from the

ion trap. It traps up to $\sim 10^6$ ions in a manner similar to that of the linear ion trap: alternating RF potentials are applied to adjacent rods to confine ions (28). Once the ions are cooled to the center of the trap, they are ejected through a slit on the interior curved rod and travel through guiding electrodes to the Orbitrap. Additionally, the guiding electrodes shift the ions to the side to prevent nitrogen gas leaking from the C-trap to the ultra-low pressure Orbitrap.

The Orbitrap is made of two cylindrical outer electrodes and one inner spindle electrode (29). Ions are confined electrostatically while orbiting the center electrode and focused towards the center of the Orbitrap. Ions are introduced off center and perpendicular to the center of the Orbitrap to ensure that the ions move radially along the center electrode. The ions are also radially moving back and forth along the center electrode in the z direction. Each ion has a unique frequency of oscillation that is m/z dependent and is defined by:

$$(1.8) \omega = \sqrt{\frac{z}{m}}k$$

where ω represents the frequency of oscillation and k represents a constant force related to the Orbitrap (29). This is a different frequency of oscillation than in the linear ion trap. The two outer electrodes, separated via a ceramic ring, each measure the image current of ions as they travel back and forth, known as a transient. The transient contains the sum of all sine waves produced by the oscillating ions (Figure 1.5A). The transient is then Fourier Transformed from the time domain to the frequency domain and then to the resulting m/z (Figure 1.5B). The length of acquisition is proportional to the resolution of the resulting mass spectrum (35). Since each individual m/z has a characteristic frequency, the Orbitrap enables ppm accuracy with scan times under one second. Although the current generation of Orbitraps provide maximum resolving powers of $\sim 1,000,000$ in ~ 1 s, experiments performed in this dissertation use resolving powers of 30,000, 60,000, and 120,000 which require under 256 ms to acquire (36, 37).

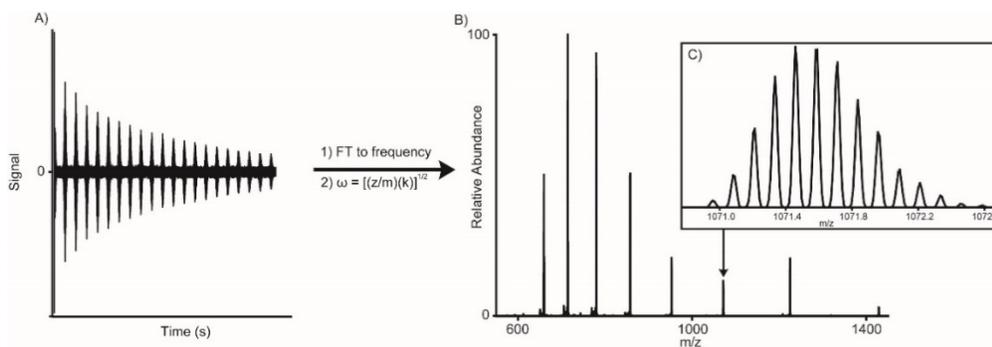


Figure 1.5. Orbitrap Transient to Mass Spectrum. A) Transient acquired for the full scan MS¹ of bovine ubiquitin at 120,000 resolution. The transient contains the sum of all ions' specific sine waves present in the Orbitrap. It is then Fourier transformed to the frequency domain which can then be converted to m/z using Equation 1.8 to give the spectrum shown in B). C) Expanded view of the +8 charge state depicting the isotopic resolution capabilities.

Ideally, ions will oscillate at their defined frequencies for an unlimited amount of time. However, the ions will deviate from their desired frequency due to collisions with gas. The Orbitrap pressure is on the order of $10^{-10} - 10^{-11}$ Torr to minimize these collisions. These low pressures increase the mean free path of the ions to allow for longer transients to be measured prior to collisional cooling (32).

The Exactive series of Thermo instruments primarily use the Orbitrap as the mass analyzer (38, 39). However, Orbitraps are commonly paired with linear ion traps in the hybrid format of instruments (Figure 1.6A) or with a quadrupole mass filter and linear ion trap in the newer tribrid format (Figure 1.6B) (31, 35, 36). Both instrument configurations have the advantage of acquiring data in the Orbitrap and in the linear ion trap at the same time. The linear ion trap can manipulate ions and perform mass analysis for multiple low resolution scans in parallel while the Orbitrap is acquiring a transient over hundreds of milliseconds. The tribrid configuration has the ability to isolate a particular m/z range by the front end quadrupole prior to reaching the linear ion trap, further minimizing the time necessary between low resolution scans (36).

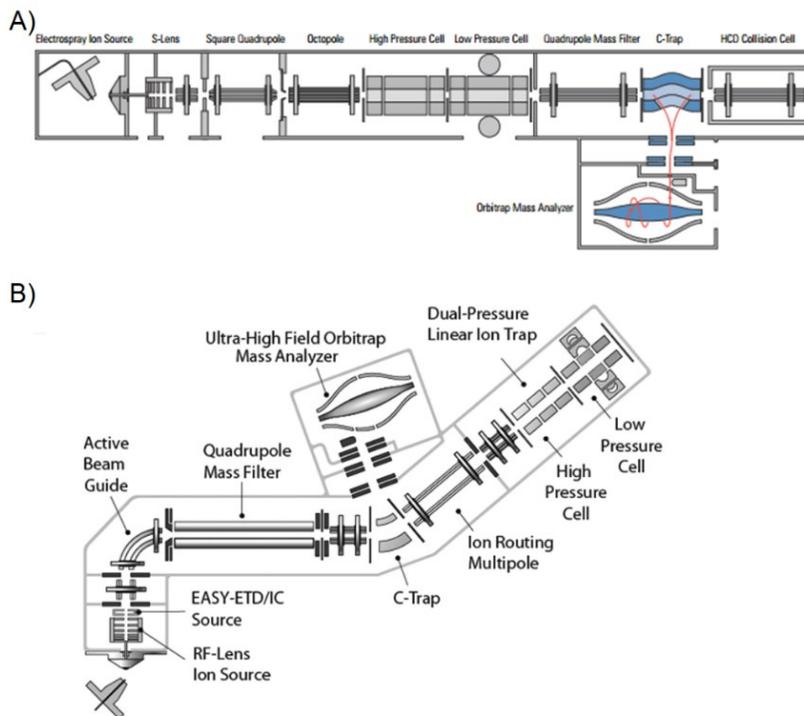


Figure 1.6. Schematics of Instruments. A) Schematic of Thermo Orbitrap Elite hybrid. Image taken from (28). B) Schematic of Thermo Orbitrap Fusion tribrid. Image taken from (36).

1.2.6 Collisional Dissociation

The past few sections described how ions were transported and analyzed within the mass spectrometer. However, dissociation is necessary to determine the primary structure of a peptide or protein (40). The most common dissociation method involves collisional fragmentation (41). There are two types of collisional fragmentation mechanisms: collisionally activated dissociation (CAD) and higher-energy collisional dissociation (HCD). Both mechanisms result in fragment ions with characteristic mass shifts corresponding to amino acid residue masses that determine primary structure information.

CAD occurs in the high-pressure linear ion trap. Ions selected for dissociation, also known as precursor ions, are given a q value equal to 0.25 and are resonantly excited through the application of a supplemental RF potential. This increases the relative velocity, and in turn the kinetic energy, of the precursor ions in addition to increasing their radius of motion. The potential applied is lower than the potential applied for a scan out event to the detector since the ions need to stay within the ion trap for fragmentation. The precursor ions collide with the helium gas present in the ion trap

more frequently and more energetically. Each collision transfers a small amount of kinetic energy into internal vibrational energy. When the vibrational energy overcomes the lowest energy pathway, the precursor dissociates. The lowest pathway is typically an amide bond cleavage along the backbone to give b-type (containing the N-terminus) and y-type (containing the C-terminus) fragment ions (Figure 1.7). Since the resulting fragment ions have different m/z values than the precursor ions, these ions are not resonantly excited and do not fragment further, ensuring that one dissociation event occurs per precursor ion (see Figure 1.8 for example) (7, 42). Additionally, a low mass cut off is introduced to the resulting fragment ions due to the precursor ions having a q value of 0.25. Fragment ions that have a q value greater than the stability limit ($q > 0.908$) of the ion trap are lost. Typically, this cut off value is $\sim 1/3$ the precursor mass.

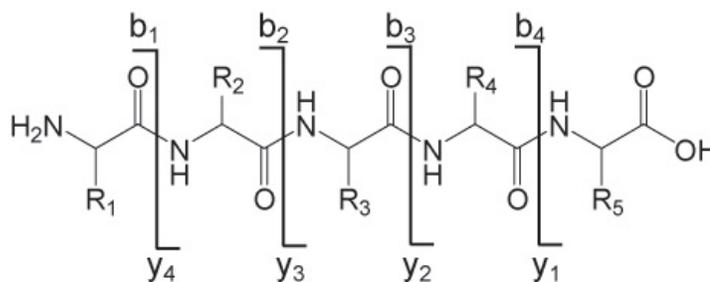


Figure 1.7. Simplified Representation of b- and y-ions Formed During Collisional Fragmentation.

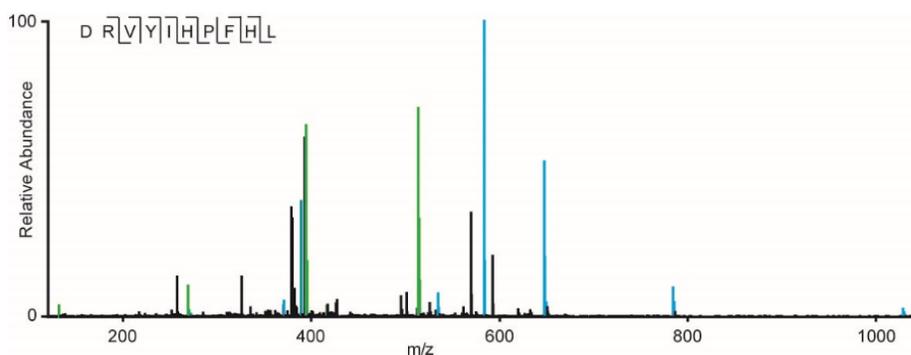


Figure 1.8 Example CAD MS². The triply charged peptide DRVYIHPFHL was fragmented by CAD. b-ions are noted in blue and y-ions are noted in green.

HCD occurs within the HCD cell (in hybrid formats) or in the Ion Routing Multipole (in tribrid formats). These cells contain nitrogen molecules as the bath gas. A voltage gradient is applied to these cells to accelerate precursor ions with much higher kinetic energy than in CAD. The collisions

with the nitrogen bath gas cause the precursor ions to dissociate along the lowest energy pathway. However, this method of dissociation is not m/z dependent. Therefore, first generation fragment ions are still being accelerated across the cell following dissociation, often resulting in secondary fragmentation (example shown in Figure 1.9) (43). Beneficially, this method of collisional fragmentation does not have a low mass cut off, enabling the detection of low mass ions.

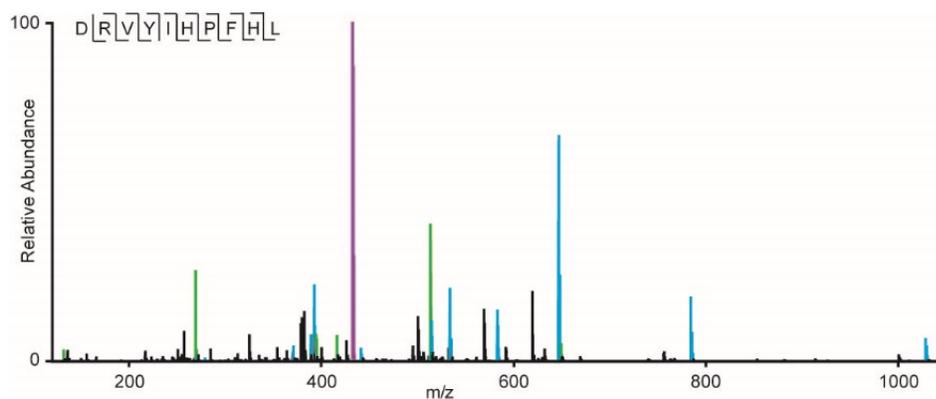


Figure 1.9. Example HCD MS². The triply charged peptide DRVYIHPFHL was fragmented by HCD. b-ions are noted in blue, y-ions are noted in green, and non-dissociated precursor ions are noted in purple.

Although collisional fragmentation is the most frequently used method for dissociation, it suffers from a critical disadvantage due to the loss of labile modifications. Certain modifications, such as phosphorylation or glycosylation, contain a bond that is weaker than an amide bond. Since collisional fragmentation involves the breaking of the lowest energy bond, this results in a fragmentation spectrum dominated by the loss of the modification (Figure 1.10) (44, 45). Although few sequence informative ions are detected, the ability to site localize the modification is greatly inhibited. A different dissociation method is needed to better localize these labile modifications.

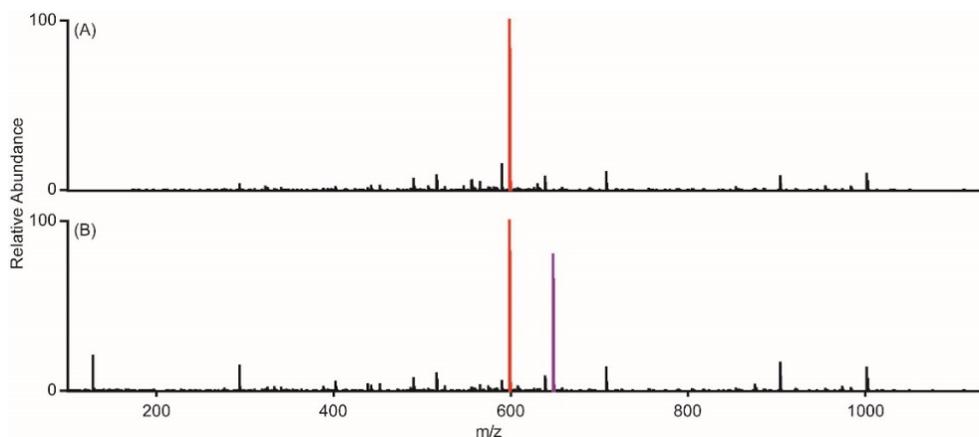


Figure 1.10. Example Phosphopeptide Undergoing Collisional Dissociation. The doubly charged peptide SP RTPVpSPVKF was fragmented by A) CAD and B) HCD. Non-dissociated precursor in B) is noted in purple. The loss of phosphoric acid is noted in red and is the most abundant species present in the scans. Sequence informative fragment ions are present at low abundances. However, fragment ions that contain the modification are prone to lose the modification as well.

1.2.7 Electron Transfer Dissociation

The other major fragmentation method used throughout this dissertation is electron transfer dissociation (ETD). Developed by Hunt et al. in 2004, this radical based fragmentation involves the co-trapping of positive (peptide/protein) and negative (reagent) ions within the ion trap (46). A radical electron is transferred to a carbonyl carbon along the peptide backbone (Figure 1.11). The newly formed negatively charged oxygen atom is stabilized by a nearby basic side chain via hydrogen bonding. Following rearrangement, the N-C α bond is ultimately broken to give c-type ions (containing the N-terminus) and z'-type ions (containing the C-terminus). This reaction reduces the overall net charge by one and is carried out within milliseconds. ETD preserves labile modifications making it the dissociation method of choice for site localization of PTMs.

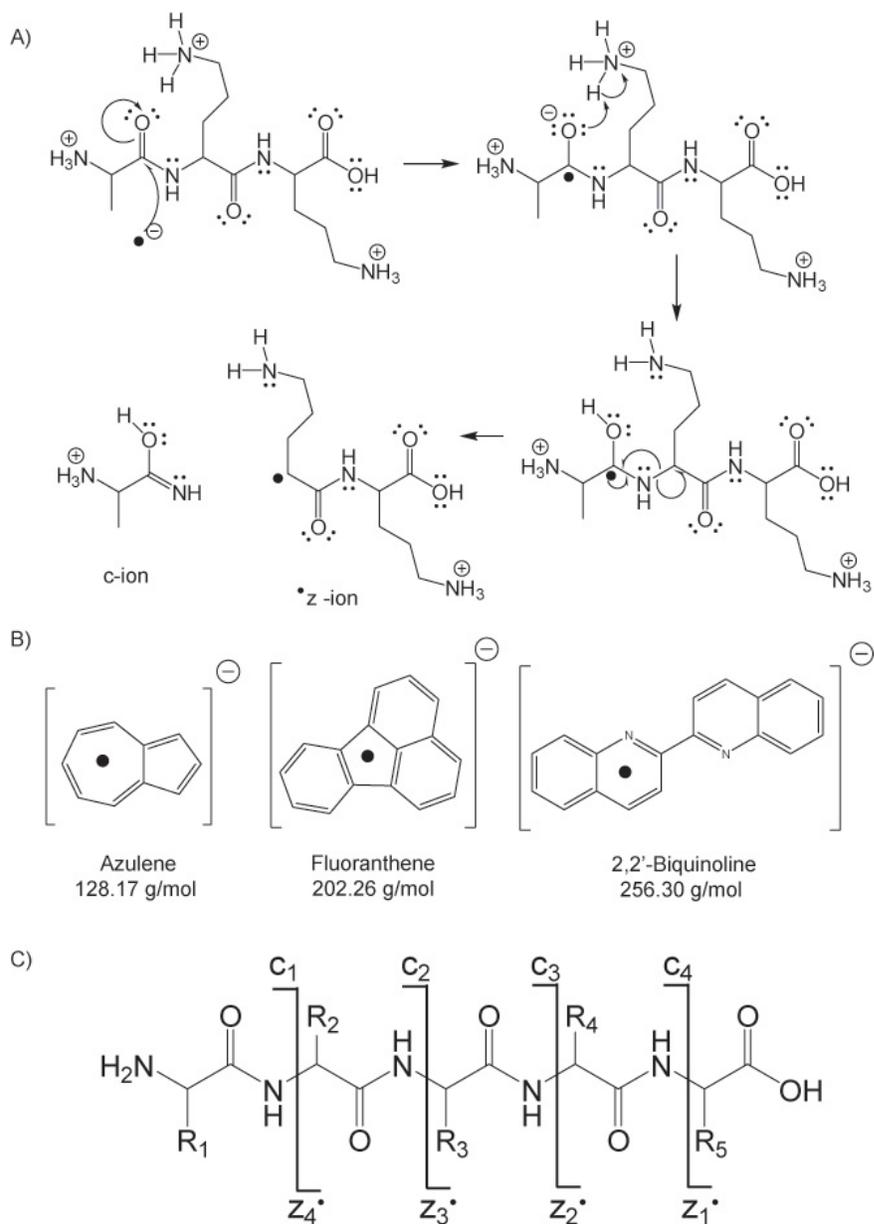


Figure 1.11. ETD Reaction Scheme. A) ETD reaction mechanism. B) Radical reagents used to transfer the radical electron during ETD. C) Simplified representation of c- and z[•]-ions formed during ETD.

This gas phase ion/ion reaction occurs in the linear ion trap. Since the linear ion trap is separated into three sections, a different DC potential can be applied to the individual sections to facilitate co-trapping of positive and negative ions (46). Typically, precursor ions are first stored in the back section of the ion trap while reagent ions are stored in the center section. Once both polarities are trapped, the DC voltages are removed to allow the precursor and reagent ions to mix.

However, this limits the number of precursor charges that can be used for subsequent analysis since the back section can only hold $\sim 1e5$ charges. High-capacity ETD can be employed to increase the number of charges for analysis (47). This involves trapping the precursor ions in the center section, which holds approximately an order of magnitude more charges than the back section, while the reagent is stored in the front section. However, this reduces the number of reagent ions that are capable of performing the reaction and thus slows the overall reaction in comparison to the normal operating mode. This has been rectified in the newest Thermo instrument, the Orbitrap Eclipse Tribrid, where the front section of the ion trap has been extended to efficiently trap more charges. ETD reactions in this system can now be carried out with the optimal number of both precursor and reagent charges to maintain predictable reactions while using the maximum number of precursor charges.

Shown in Figure 1.12A is an example ETD spectrum of the triply charged phosphopeptide SPRTPVsPVKF. Characteristic mass shifts corresponding to amino acid residues enable sequencing of the peptide using both c- and z'- ions. Typical to ETD spectra are charge-reduced species (noted in red in Figure 1.12), also known as electron transfer no dissociation products (ETnoD) (48, 49). Breaking the N-C α bond of a proline residue results in an ETnoD product due to the ring side-chain of proline maintaining connectivity of the peptide. Additionally, products held together via hydrogen bonding or stabilization of the radical electron within the imidazole ring of histidines can contribute to ETnoD products (50, 51). Shown in Figure 1.12B is the doubly charged ETD spectrum of the same peptide. Since ETD neutralizes a charge, a subset of fragment ions do not carry a charge and are not detected by the mass spectrometer. This drastically limits the ability to obtain full sequence coverage of this peptide. However, informative ions are still obtained from these spectra, particularly around one or both termini.

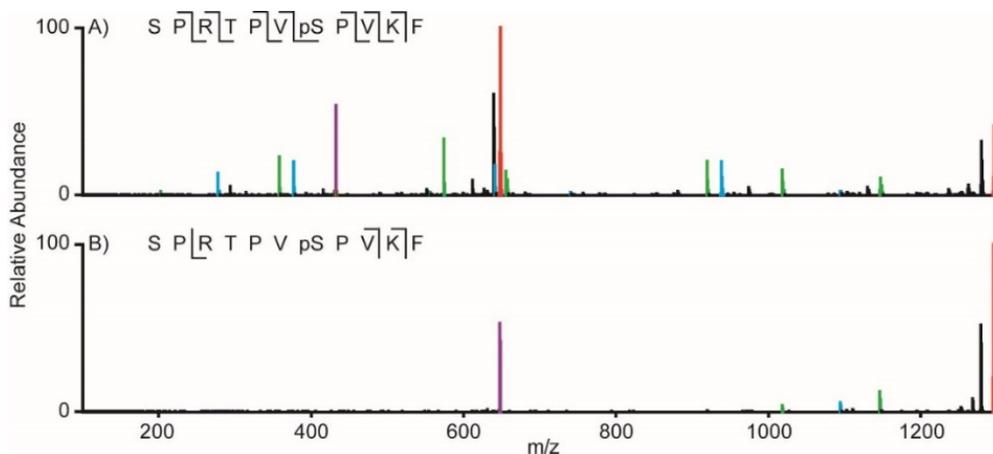


Figure 1.12. Example ETD MS². The phosphopeptide SPRTpSPVKF was dissociated by ETD. A) 45ms of ETD of the triply charged precursor. All sequence informative ions were observed in this spectrum. Fragment ions that are N-terminal to prolines were not observed due to the ring structure maintaining a linear peptide. B) 100ms of ETD of the doubly charged precursor. Significantly fewer sequence ions were observed. All c-ions are noted in blue, z⁺-ions in green, charged reduced precursors in red, and the unreacted precursor in purple.

ETD is much more informative for high charge density species. Precursor ions with m/z greater than ~950 fragment poorly by ETD (52). Figure 1.13 shows the peptide G1-K42 from equine apomyoglobin undergoing 45ms of ETD at the +6 charge state (Figure 1.13A), +5 charge state (Figure 1.13B), and the +4 charge state (Figure 1.13C). Notice as the charge density decreases, as portrayed by the precursor m/z increasing, the ion current of fragment ions decreases, which limits the number of sequence informative fragment ions that can be observed. Although the use of ETD across the mass range is limited, the benefits of random backbone cleavages while maintaining PTM connectivity during dissociation vastly outweigh the limitations.

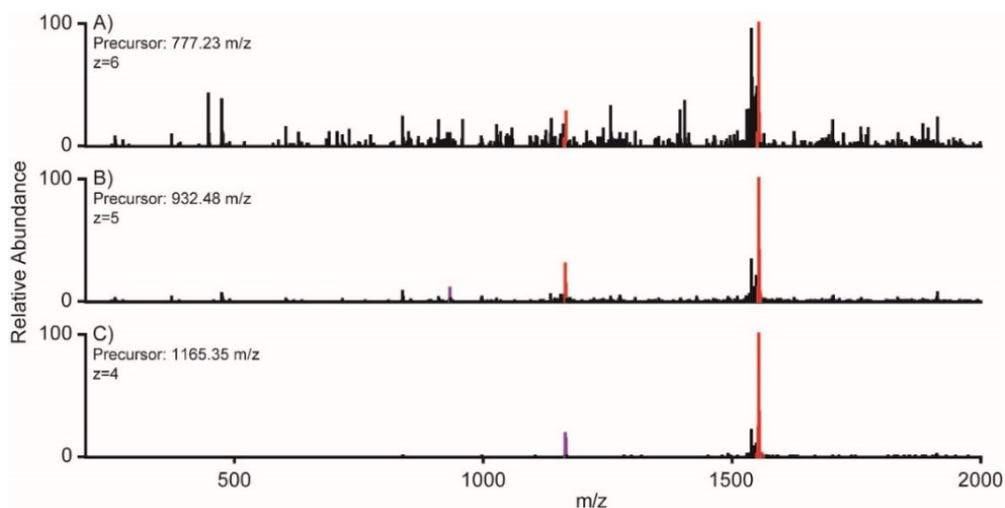


Figure 1.13. Charge Density Effects on ETD. 45ms of ETD performed on the 4.6 kDa peptide G1-K42 from equine apomyoglobin. A) Spectrum from the +6 charge state. B) Spectrum from the +5 charge state. C) Spectrum from the +4 charge state. Precursor ions are noted in purple while the charge reduced species are noted in red for all spectra. As the precursor m/z increases, the ion current from fragment ions decreases in abundance, hindering the ability to accurately sequence this molecule at lower charge states.

1.3 Protein Preparation Methods

Sample preparation of proteins is carried out prior to analysis by LC-MS/MS. The sample must be cleaned up to remove any interfering species such as excess salts, detergents, lipids, etc. Sample analysis can be approached in three different ways: trypsin digestion, no digestion, or limited digestion.

1.3.1 Trypsin Digestion

The most common preparation approach is to digest the sample with trypsin, a specific serine protease (53–55). Trypsin hydrolyzes the peptide bond C-terminal to lysine and arginine residues due to a negatively charged region in the catalytic pocket of trypsin attracting basic residues. This enzymatic digestion has a maximum efficiency at 37°C and is typically carried out overnight to ensure the digestion reaches completion, resulting in smaller peptides for a more manageable analysis (56). The lengths of these peptides are typically 12-15 residues with each termini containing a basic moiety, enabling straightforward separation followed by CAD fragmentation (54, 57). However, since most tryptic peptides are only doubly charged, ETD fragmentation provides

minimal sequence information (52). Although CAD drives the identification of the majority of the peptide, ETD is useful to confirm the residues surrounding the termini.

Although trypsin usually generates peptides that are ideal for analysis, it also generates peptides that are too short to retain on a reverse phase column due to the high abundance of lysine and arginine residues throughout the proteome (58, 59). Since these peptides are not observed during analysis, complete sequence coverage is not typically achieved. Trypsin is useful to determine whether a protein is present in a complex mixture based on a few peptide identifications or as a first pass of sequence analysis. However, trypsin by itself lacks the ability to confirm the entirety of a protein sequence (58, 60, 61). Additionally, digesting proteins with trypsin inhibits the ability to determine multiple PTMs on a single protein (Figure 1.14). The break in continuity of the primary structure prevents the ability to determine if multiple PTMs on the same protein have biological implications (62).

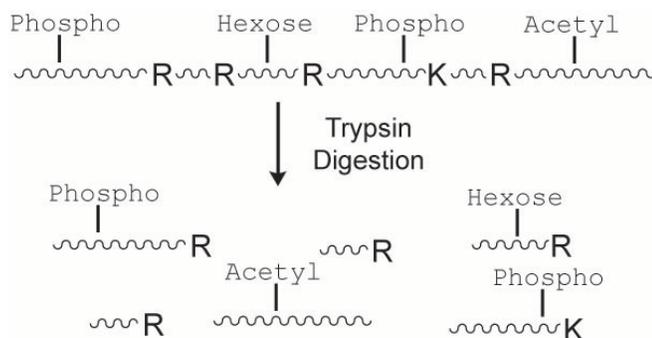


Figure 1.14. Loss of Combinatorial PTM Example. When a multiply modified protein (above) is digested with trypsin, the resulting peptide mixture (below) loses the combinatorial information of the undigested protein.

1.3.2 No Enzymatic Digestion

Recently, the protein mass spectrometry field has been moving toward no enzymatic digestion prior to analysis (63–69). The electrospray ionization process of an intact protein provides a distribution of charge states for a given protein (Figure 1.15) (64). The molecular weight can be directly measured from the distribution. Since the entire protein retains connectivity, all modifications can be observed based on mass shifts in the electrospray distribution (see Figure

1.15B) (70). Additionally, any splice variants or amino acid mutations from the DNA code will be observed from the intact mass measurement.

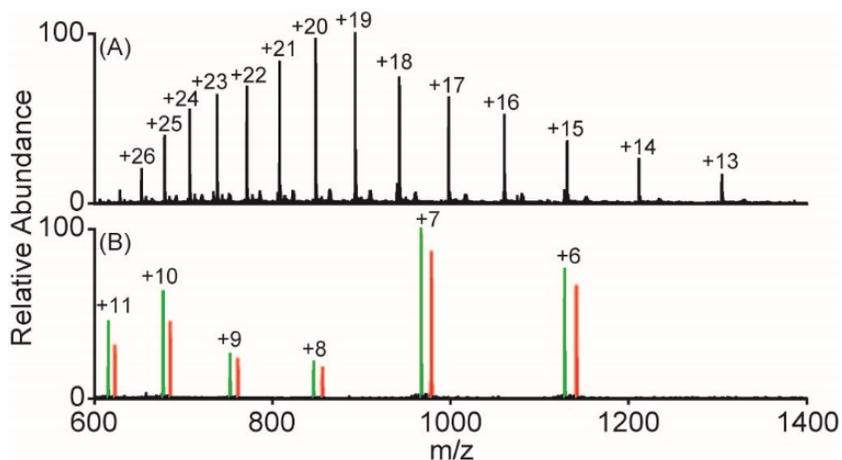


Figure 1.15. Example Electrospray Distributions for Intact Protein Analysis. The charge states for a given peak are located above the peak. (A) Distribution of the ~17 kDa protein equine apomyoglobin. (B) Distribution of the ~6.8 kDa *Marchantia polymorpha* ribosomal L29e protein. The peaks in green represent the unmodified protein. The peaks in red represent a mass shift of +79.96 Da from the green peaks, corresponding to an addition of a single phosphorylation modification.

One of the main drawbacks of analyzing intact proteins is the reduction in dynamic range (69, 71). Increasing molecular weight results in an increase in the channels for available charge states via electrospray (20). Since there are multiple charge states of the protein, the analyte signal is split between multiple peaks. Furthermore, each charge state is divided into multiple isotopic peaks splitting the signal further. Combining all the channels that split the analyte signal results in a lower S/N ratio, representing a challenge in detecting low abundance protein species as well as proteins with large molecular weights (> 30kDa) (69, 71). The S/N drop also affects any fragmentation scans generated. Since the ion trap reaches a target number of *charges* for fragmentation, there are significantly less *molecules* present for fragmentation in comparison to a peptide (27). For example, a target of 100,000 *charges* corresponds to 50,000 *molecules* of a +2 peptide versus 5,000 *molecules* of a +20 protein.

Lack of complete sequence coverage is also an issue when analyzing intact proteins. Extensive fragmentation methods have been developed and tested to increase the extent of sequence coverage (46, 72–75). The sequence coverage of proteins decreases as molecular

weight increases due to the greater number of fragmentation paths in higher molecular weight species (71). Proteins can be sufficiently identified with a small percentage of identified fragmentation ions. However, the lack of complete coverage can prevent the specific localization of any modifications present (69). Eliminating the digestion step prior to analysis provides many benefits for sample analysis, particularly when the sample contains proteins that are less than 30 kDa. This methodology will become more applicable with the advancement in separation strategies, fragmentation strategies, and instrumentation.

1.3.3 Limited Digestion

An alternative to trypsin digestion and no digestion is to digest a protein at less frequent residues with the goal of generating peptides that are larger than the average tryptic peptide but smaller than ~20 kDa (76–82). Typically, specific proteases that cleave at less frequent residues are used to generate these peptides. The commonly used proteases include Lys-C (C-terminal cleavage to K), Glu-C (C-terminal cleavage to E and to a lesser extent D) and Asp-N (N-terminal cleavage to D) (78). Additionally, nonspecific proteases can be used to generate these intermediate sized peptides. Further detail about this class of proteases will be discussed in Chapter 3.

The main advantage for using these alternate enzymes is to create peptides for analysis within the current limitations of instrumentation (71). These digests regularly produce peptides with charge states greater than +2, perfectly suiting ETD (46, 83). A greater portion of a protein can be efficiently sequenced and thus lead to a more complete coverage analysis (82). Although overall connectivity of the protein of interest is still lost due to digestion, peptides generated are still sufficiently large to maintain connectivity of any PTMs in a given region (Figure 1.16).

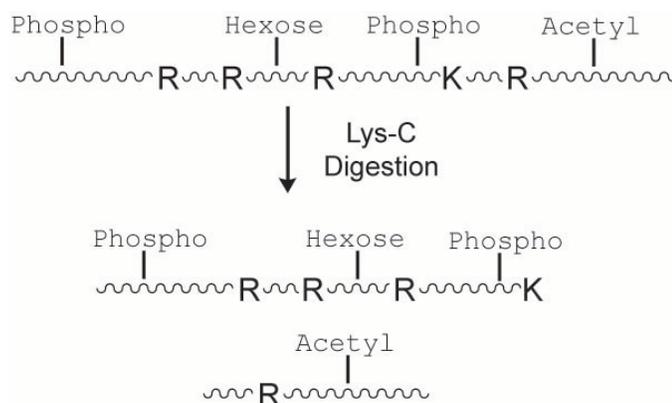


Figure 1.16. Partial Retention of Combinatorial PTM Example. When a multiply modified protein (above) is digested with Lys-C, fewer residues are sites of enzymatic cleavages. The resulting peptide mixture (below) retains a larger portion of the initial primary structure. Although all PTMs are unable to be observed on a single peptide, a majority of PTMs are still connected, giving additional information of the biological relevance of these modifications.

Although a compromise on peptide size distribution is reached with limited digestions, obstacles still obscure its wide-range adaptation. These alternate specific proteases are not as efficient as a trypsin digestion, leading to missed cleavage patterns that further complicate the analysis (84). Additionally, these proteases still have the potential to generate peptides too small to retain on a reversed phase column, which could lead to incomplete sequence coverage and loss of potential PTM localization.

The previous sections described different sample preparation strategies. Although each method has powerful advantages, they each suffer in different ways when used alone. However, analyzing the same sample using a combination of strategies can fill in the gaps produced in one method. For example, an intact analysis of a sample can be done first to obtain the molecular weight of the proteins present as well as any modifications present. Fragmentation of these intact proteins can also provide preliminary sequence information and PTM identification. These intact results can be used to tailor the digestion with a protease to fill in the gaps in sequence coverage as well as localize the observed modifications in the intact analysis. Although the combination of analyses adds time for a single sample, a more in depth analysis of the sample is obtained. This type of methodology combination is also useful for very large proteins, such as antibodies, in which a ~150 kDa sequence analysis is infeasible using a single method.

1.4 References

1. Mann M, Jensen ON (2003) Proteomic analysis of post-translational modifications. *Nat Biotechnol* 21(3):255–261.
2. Mimura Y, et al. (2018) Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein Cell* 9(1):47–62.
3. Nielsen ML, Savitski MM, Zubarev RA (2006) Extent of Modifications in Human Proteome Samples and Their Effect on Dynamic Range of Analysis in Shotgun Proteomics. *Mol & Cell Proteomics* 5(12):2384 LP – 2391.
4. Yates III JR (1998) Mass spectrometry and the age of the proteome. *J Mass Spectrom* 33(1):1–19.
5. Domon B, Aebersold R (2006) Mass Spectrometry and Protein Analysis. *Science (80-)* 312(5771):212 LP – 217.
6. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422(6928):198–207.
7. Rathore D, et al. (2015) Ion Dissociation Methods in Proteomics. *Encycl Anal Chem*:1–26.
8. Martin SE, Shabanowitz J, Hunt DF, Marto JA (2000) Subfemtomole MS and MS/MS Peptide Sequence Analysis Using Nano-HPLC Micro-ESI Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal Chem* 72(18):4266–4274.
9. Aguilar M-I, Hearn MTW (1996) [1] High-resolution reversed-phase high-performance liquid chromatography of peptides and proteins. *Methods Enzymol* 270:3–26.
10. Snyder LR, Kirkland JJ, Dolan JW (2011) *Introduction to modern liquid chromatography* (John Wiley & Sons).
11. Mann M, Hendrickson RC, Pandey A (2001) Analysis of Proteins and Proteomes by Mass Spectrometry. *Annu Rev Biochem* 70(1):437–473.
12. Ardrey RE (2003) *Liquid chromatography-mass spectrometry: an introduction* (John Wiley & Sons).
13. Krokhin O V, et al. (2004) An Improved Model for Prediction of Retention Times of Tryptic Peptides in Ion Pair Reversed-phase HPLC. *Mol Cell Proteomics* 3(9):908 LP – 919.
14. Marginean I, et al. (2008) Analytical Characterization of the Electrospray Ion Source in the Nanoflow Regime. *Anal Chem* 80(17):6573–6579.
15. Yamashita M, Fenn JB (1984) Electrospray ion source. Another variation on the free-jet theme. *J Phys Chem* 88(20):4451–4459.
16. Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB (1985) Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* 57(3):675–679.
17. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science (80-)* 246(4926):64–71.
18. Wilm M (2011) Principles of electrospray ionization. *Mol Cell Proteomics* 10(7):M111.009407-M111.009407.
19. Cech NB, Enke CG (2001) Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* 20(6):362–387.
20. Schnier PD, Gross DS, Williams ER (1995) On the maximum charge state and proton transfer reactivity of peptide and protein ions formed by electrospray ionization. *J Am Soc Mass Spectrom* 6(11):1086–1097.
21. Liu H, et al. (2011) The Prediction of Peptide Charge States for Electrospray Ionization in Mass Spectrometry. *Procedia Environ Sci* 8:483–491.
22. Tang K, Page JS, Smith RD (2004) Charge competition and the linear dynamic range of

- detection in electrospray ionization mass spectrometry. *J Am Soc Mass Spectrom* 15(10):1416–1423.
23. March RE, Todd JF (2005) *Quadrupole ion trap mass spectrometry* (John Wiley & Sons).
 24. Kononkov N V, Sudakov M, Douglas DJ (2002) Matrix methods for the calculation of stability diagrams in quadrupole mass spectrometry. *J Am Soc Mass Spectrom* 13(6):597–613.
 25. Miller PE, Denton MB (1986) The quadrupole mass filter: basic operating concepts. *J Chem Educ* 63(7):617.
 26. Douglas DJ (2009) Linear quadrupoles in mass spectrometry. *Mass Spectrom Rev* 28(6):937–960.
 27. Schwartz JC, Senko MW, Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 13(6):659–669.
 28. Olsen J V, et al. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell proteomics* 8(12):2759–2769.
 29. Makarov A (2000) Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal Chem* 72(6):1156–1162.
 30. Hu Q, et al. (2005) The Orbitrap: a new mass spectrometer. *J Mass Spectrom* 40(4):430–443.
 31. Makarov A, et al. (2006) Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Anal Chem* 78(7):2113–2120.
 32. Zubarev RA, Makarov A (2013) Orbitrap Mass Spectrometry. *Anal Chem* 85(11):5288–5296.
 33. Eliuk S, Makarov A (2015) Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu Rev Anal Chem* 8(1):61–80.
 34. Olsen J V, et al. (2005) Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Mol Cell Proteomics* 4(12):2010 LP – 2021.
 35. Michalski A, et al. (2012) Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* 11(3):O111.013698-O111.013698.
 36. Senko MW, et al. (2013) Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal Chem* 85(24):11710–11714.
 37. Denisov E, Damoc E, Lange O, Makarov A (2012) Orbitrap mass spectrometry with resolving powers above 1,000,000. *Int J Mass Spectrom* 325–327:80–85.
 38. Michalski A, et al. (2011) Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol Cell Proteomics* 10(9):M111.011015.
 39. Scheltema RA, et al. (2014) The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. *Mol Cell Proteomics* 13(12):3698 LP – 3708.
 40. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR (1986) Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci* 83(17):6233 LP – 6237.
 41. McLuckey SA (1992) Principles of collisional activation in analytical mass spectrometry. *J Am Soc Mass Spectrom* 3(6):599–614.
 42. Mitchell Wells J, McLuckey SA (2005) Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol* 402:148–185.

43. Olsen J V, et al. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 4:709.
44. Boersema PJ, Mohammed S, Heck AJR (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* 44(6):861–878.
45. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ (2004) A Neutral Loss Activation Method for Improved Phosphopeptide Sequence Analysis by Quadrupole Ion Trap Mass Spectrometry. *Anal Chem* 76(13):3590–3598.
46. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 101(26):9528.
47. Riley NM, et al. (2016) Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. *J Am Soc Mass Spectrom* 27(3):520–531.
48. Liu J, McLuckey SA (2012) Electron transfer dissociation: effects of cation charge state on product partitioning in ion/ion electron transfer to multiply protonated polypeptides. *Int J Mass Spectrom* 330:174–181.
49. Riley NM, Coon JJ (2018) The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal Chem* 90(1):40–64.
50. Lermyte F, Łacki MK, Valkenborg D, Gambin A, Sobott F (2017) Conformational Space and Stability of ETD Charge Reduction Products of Ubiquitin. *J Am Soc Mass Spectrom* 28(1):69–76.
51. Tureček F, et al. (2010) The Histidine Effect. Electron Transfer and Capture Cause Different Dissociations and Rearrangements of Histidine Peptide Cation-Radicals. *J Am Chem Soc* 132(31):10728–10740.
52. Good DM, Wirtala M, McAlister GC, Coon JJ (2007) Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* 6(11):1942–1951.
53. Rogers JC, Bomgardner RD (2016) Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides. *Modern Proteomics -- Sample Preparation, Analysis and Practical Applications*, eds Mirzaei H, Carrasco M (Springer International Publishing, Cham), pp 43–62.
54. Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates III JR (2013) Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* 113(4):2343–2394.
55. Vandermarliere E, Mueller M, Martens L (2013) Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom Rev* 32(6):453–465.
56. Finehout EJ, Cantor JR, Lee KH (2005) Kinetic characterization of sequencing grade modified trypsin. *Proteomics* 5(9):2319–2321.
57. Burkhart JM, Schumbrutzki C, Wortelkamp S, Sickmann A, Zahedi RP (2012) Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J Proteomics* 75(4):1454–1462.
58. Swaney DL, Wenger CD, Coon JJ (2010) Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* 9(3):1323–1329.
59. Hildonen S, Halvorsen TG, Reubsæet L (2014) Why less is more when generating tryptic peptides in bottom-up proteomics. *Proteomics* 14(17–18):2031–2041.
60. McCormack AL, et al. (1997) Direct Analysis and Identification of Proteins in Mixtures by LC/MS/MS and Database Searching at the Low-Femtomole Level. *Anal Chem* 69(4):767–776.
61. Choudhary G, Wu S-L, Shieh P, Hancock WS (2003) Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary

- LC with ion trap MS/MS. *J Proteome Res* 2(1):59–67.
62. Lanucara F, Eyers CE (2013) Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. *Mass Spectrom Rev* 32(1):27–42.
 63. Cui W, Rohrs HW, Gross ML (2011) Top-down mass spectrometry: recent developments, applications and perspectives. *Analyst* 136(19):3854–3864.
 64. Reid GE, McLuckey SA (2002) 'Top down' protein characterization via tandem mass spectrometry. *J mass Spectrom* 37(7):663–675.
 65. Earley L, et al. (2013) Front-End Electron Transfer Dissociation: A New Ionization Source. *Anal Chem* 85(17):8385–8390.
 66. Siuti N, Kelleher NL (2007) Decoding protein modifications using top-down mass spectrometry. *Nat Methods* 4(10):817.
 67. Garcia BA (2010) What does the future hold for top down mass spectrometry? *J Am Soc Mass Spectrom* 21(2):193–202.
 68. Toby TK, Fornelli L, Kelleher NL (2016) Progress in top-down proteomics and the analysis of proteoforms. *Annu Rev Anal Chem* 9:499–519.
 69. Catherman AD, Skinner OS, Kelleher NL (2014) Top down proteomics: facts and perspectives. *Biochem Biophys Res Commun* 445(4):683–693.
 70. Zhang H, Ge Y (2011) Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circ Cardiovasc Genet* 4(6):711.
 71. Compton PD, Zamdborg L, Thomas PM, Kelleher NL (2011) On the scalability and requirements of whole protein mass spectrometry. *Anal Chem* 83(17):6868–6874.
 72. Fornelli L, et al. (2017) Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J Proteome Res* 16(2):609–618.
 73. Frese CK, et al. (2012) Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal Chem* 84(22):9668–9673.
 74. Ledvina AR, et al. (2010) Activated-Ion Electron Transfer Dissociation Improves the Ability of Electron Transfer Dissociation to Identify Peptides in a Complex Mixture. *Anal Chem* 82(24):10068–10074.
 75. Shaw JB, et al. (2013) Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J Am Chem Soc* 135(34):12646–12651.
 76. Cannon J, et al. (2010) High-Throughput Middle-Down Analysis Using an Orbitrap. *J Proteome Res* 9(8):3886–3890.
 77. Laskay ÜA, Lobas AA, Srzentić K, Gorshkov M V, Tsybin YO (2013) Proteome Digestion Specificity Analysis for Rational Design of Extended Bottom-up and Middle-down Proteomics Experiments. *J Proteome Res* 12(12):5558–5569.
 78. Tsiatsiani L, Heck AJR (2015) Proteomics beyond trypsin. *FEBS J* 282(14):2612–2626.
 79. Cristobal A, et al. (2017) Toward an Optimized Workflow for Middle-Down Proteomics. *Anal Chem* 89(6):3318–3325.
 80. Zhang L, et al. (2016) Analysis of monoclonal antibody sequence and post-translational modifications by time-controlled proteolysis and tandem mass spectrometry. *Mol Cell Proteomics* 15(4):1479–1488.
 81. Sidoli S, Garcia BA (2017) Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert Rev Proteomics* 14(7):617–626.
 82. Pandeswari PB, Sabareesh V (2019) Middle-down approach: a choice to sequence and characterize proteins/proteomes by mass spectrometry. *RSC Adv* 9(1):313–344.
 83. Mikesh LM, et al. (2006) The utility of ETD mass spectrometry in proteomic analysis.

- Biochim Biophys Acta - Proteins Proteomics* 1764(12):1811–1822.
84. Giansanti P, Tsiatsiani L, Low TY, Heck AJR (2016) Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nat Protoc* 11:993.

Determination of Protamines from *Marchantia Polymorpha*

This chapter was a collaborative project with Dr. Juan Ausió, Department of Biochemistry and Microbiology, University of Victoria, Canada.

2.1 Abstract

The genetic material of all organisms is preserved within the DNA sequence and confined within the nucleus. Sexually reproducing species requires the further compaction of DNA within the smaller, more hydrodynamic sperm cells. Protamines are substituted for the histones normally used to compact DNA in sperm cells. These arginine rich proteins interact strongly with the DNA backbone such that the DNA is extensively condensed within spermatozoa. The origins of these proteins were thought to be the result of vertical evolution from histone H1 via a protamine-like intermediate. Previous studies have defined the evolution of protamine-like proteins to protamines, while no direct evidence of the connection of protamine-like proteins to histone H1 has been determined. This chapter details the identification of protamines from the common liverwort *Marchantia polymorpha*, the discovery of a novel post translational modification, and evidence directly linking protamines to histone H1 precursors.

2.2 Introduction

2.2.1 Histones and Protamines

DNA is comprised of two chains consisting of a deoxyribose/phosphate backbone with a nucleobase on each sugar. The four bases form pairs via hydrogen bonding, adenine pairs with thymine and guanine pairs with cytosine, to secure the higher order structure of a double helix (1). DNA interacts with a class of proteins known as histones to compact the millions of base pairs within the nucleus (2, 3). Histones are highly basic proteins and consist of five members: H1, H2A, H2B, H3, and H4. Two H2A/H2B dimers and one tetramer of two H3 and two H4 form the core histone octamer (2, 3). The negatively charged DNA backbone interacts with the positively charged histone octamer such that ~147 base pairs of DNA wrap around the core octamer. This interaction compacts DNA into a nucleosome, the basic unit of chromatin (Figure 2.1) (2, 4).

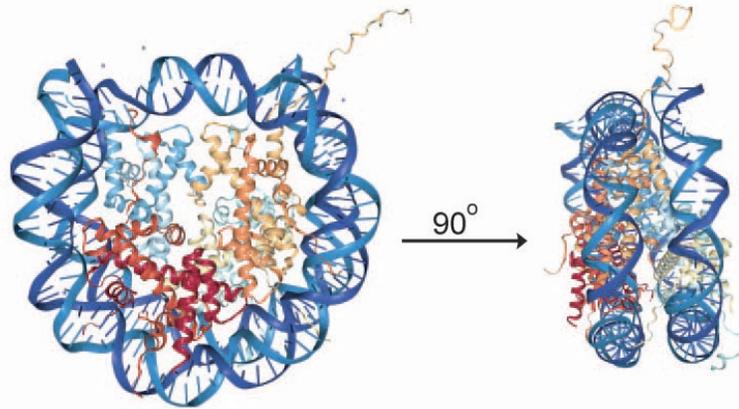


Figure 2.1. Crystal Structure of a Nucleosome. DNA (dark blue) wraps around the core histone octamer. The N-terminal tails of the histones extend outside of the DNA and are subjected to modification. Crystal structure provided by PDB (ID: 1AOI) (10). Solving of structure by reference (11).

Structurally, the core histones can be broken into two sections, a highly basic, disordered N-terminal tail and a globular domain. The N-terminal tails extend outside of the nucleosome and act as substrates for modification by multiple enzymes to enhance or diminish the accessibility of DNA (5–7). Multiple nucleosomes tethered together with regions of unbound DNA (linker DNA) form the first order of compaction for DNA, referred to as “beads-on-a-string” (8). The nucleosomes are compacted further with the help of H1, the linker histone, to generate the 30-nm fiber (8). Although the exact structure of the 30-nm fiber has not been confirmed, the nucleosomes interact with nearby nucleosomes to generate a helical structure (8, 9). The helices interact to further compact the nucleosomes into the final observable chromosome.

H1 interacts with the entry and exit strands of DNA from the nucleosome (~20 base pairs total) to assist in generating higher order structures of nucleosomes (Figure 2.2) (12, 13). Structurally, H1 is divided into three sections. The center section forms a winged helix domain, a feature common to transcription factors to bind DNA (14, 15). It is bordered by a short N-terminal region and a long, intrinsically disordered C-terminal region, both of which are high in lysine abundance. Additionally, H1 is coded by multiple different genes producing additional subtypes of H1 (12, 16, 17). Humans contain eleven genes to code for different H1 subtypes. All of the subtypes maintain the highly conserved globular region but diverge greatly in the length of the C-terminal region (12).

H1 is in a highly dynamic state with linker DNA. Generally, the length of the C-terminal region correlates to the affinity between H1 and DNA, longer C-terminal tails have a stronger affinity (16). Its highly dynamic properties allow H1 to readily dissociate itself from the nucleosome to allow access to the compacted DNA.

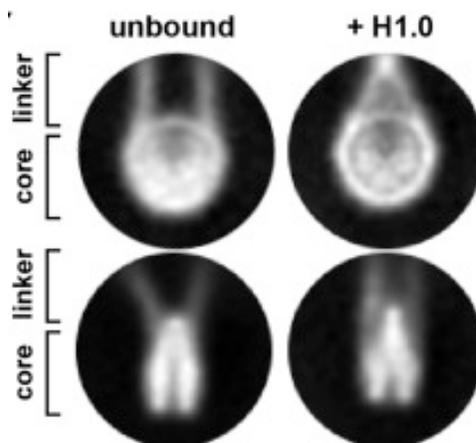


Figure 2.2. Histone H1 Interaction with a Nucleosome. The DNA linking strands freely extend away from the nucleosome without H1 interaction. The DNA linking strands are much more rigid in a defined conformation with H1 interaction. Image modified from reference (4).

Although histones play a major role in DNA compaction in the vast majority of cells, this is not the case for mature spermatozoa. Spermatozoa are continually produced from germ cells in the male sex organ through the process known as spermatogenesis (18, 19). Briefly, this process involves a mitotic division of the undifferentiated germ cells into two copies. One copy will replenish the undifferentiated germ cell population while the second copy will differentiate to the primary spermatocyte. Since the primary spermatocyte contains diploid chromosomes, these cells undergo meiosis into four spermatid with haploid chromosomes (19).

DNA in the differentiating spermatocytes has been compacted using histones (Figure 2.3A). However, the development of the final spermatozoa undergoes one final stage of spermatogenesis, known as spermiogenesis (19). The total process takes the haploid spermatid, which is a spherical cell, to the final mobile spermatozoa, which is elliptical and contains a tail. This process requires further compaction of DNA to prevent the transcription and damage of DNA and to minimize the size of the spermatozoa head to increase its hydrodynamics (20, 21). The majority of histones (~85-

90%) are removed by spermatid nuclear transition proteins and replaced with protamines (22–24). The protamines ultimately give DNA a toroidal conformation, enabling more extensive compaction than histones (Figure 2.3B) (20, 25).

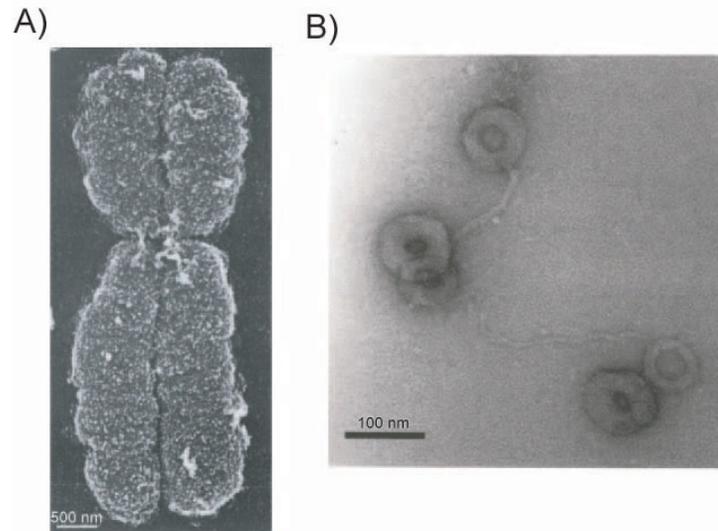


Figure 2.3. Differing DNA Compaction by Histones and Protamines. A) Scanning electron microscope image of a human chromosome found in a majority of cells. Histones are solely used to compact DNA. Image taken from reference (26). B) Transmission electron microscope image of bacterial DNA treated with protamines of salmon. The resulting structure is toroidal and is much further condensed in comparison to histone bound DNA. Image taken from reference (27).

Protamines are members of the sperm nuclear basic proteins (SNBPs). They are small, basic proteins that consist of primarily arginine residues (28). Arginine residues are advantageous over lysine residues to further compact DNA because the guanidine group forms additional hydrogen bonds with DNA (29–31). Some species contain cysteine residues and can form disulfide bonds between protamines to further stabilize their interaction with DNA (24). Protamines are replaced by histones following fertilization in a mechanism that is poorly understood (32, 33). However, it is known that the reduction of disulfide bonds in certain species is necessary (33, 34).

2.2.2 Evolutionary Theory of Protamines

Protamines are not exclusively used by all species to compact DNA in spermatozoa. Other proteins found within different species include histone and protamine-like proteins, both members of the SNBPs (28, 35, 36). These proteins differ in their ratio of arginine to lysine content:

protamines are primarily arginine, histones are primarily lysine with minimal arginine, and protamine-like are an even mix of arginine and lysine (36).

In 1973, the identification of SNBPs in mollusks led to the hypothesis that protamines evolved from a H1 precursor through a protamine-like intermediate via a vertical evolutionary process (36, 37). All three types of SNBPs were observed within this phylum, which contain species of chitons, mussels, clams, snails, and squids. Further studies of *Spisula solidissima* (surf clams) conducted by Ausió et al. in 1987 provided the initial link between protamine-like proteins and H1 (38). The primary protein used to condense DNA in the sperm of clams was much larger in molecular weight than histones or protamines and contained a trypsin-resistant core that shared ~50% sequence similarity to that of bird H5, a terminally differentiated linker histone (36, 38, 39). This evidence in combination with the discovery of similar protamine-like proteins in tunicates and particular fish provided the initial evidence for the transition of protamine-like proteins from H1 (39–42).

Studies of the tunicate *Styela plicata* (sea squirt) presented another piece of the puzzle (39, 40). This organism expresses two basic proteins in its sperm. The first is a large protamine-like protein with a trypsin resistant core, similar to the protamine-like proteins found in mollusks (38–40). This protein has considerable amounts of lysine and arginine residues with the majority of the arginine residues located in the C-terminal region. The second protamine is much shorter in length and has the exact same sequence of the C-terminal region of the large protamine-like protein (39, 40). Therefore, the origin of these short, arginine rich protamines was likely a result of the post-translational cleavage of large protamine-like proteins, solidifying the evolutionary link between protamine-like proteins and protamines (39, 40).

The transition from lysine rich to arginine rich proteins remains uncertain. One plausible hypothesis was found in studies of two closely related tunicates, *Ciona intestinalis* (sea vase) and *Styela montereyensis* (stalked tunicate) (39, 40). *Ciona intestinalis* contains a single protamine-like protein whose C-terminal region is lysine rich while *Styela montereyensis* contains a small, arginine rich protamine and a protamine-like protein whose C-terminal region is arginine rich (40, 42). Analysis of the respective genomes shows that the C-terminal lysine residues in *Ciona intestinalis* were coded in the DNA using AAG codons while the C-terminal arginine residues in *Styela*

montereyensis were coded using AGA codons (39). It is unlikely that each codon in the protein contained two point mutations per lysine residue to drastically shift the composition of lysine to arginine. What was more likely was a frameshift mutation to shift the AAG reads to AGA, effectively mutating lysine into arginine (Figure 2.4) (39).

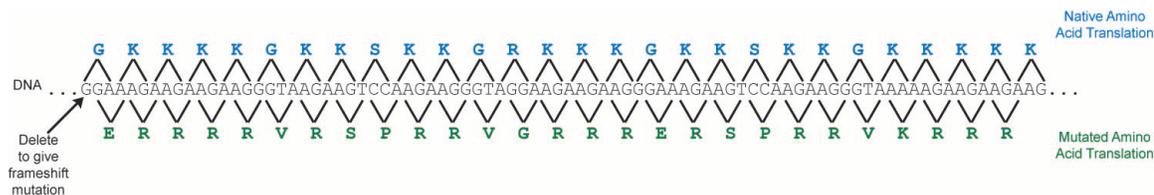


Figure 2.4. Hypothesized Reason for Shift from Lysine to Arginine via Frameshift Mutation. Shown here is the DNA sequence in the C-terminal region of histone H1 from *Ciona intestinalis*. The top line (blue) shows the codon translation from the native sequence. However, if the first G is deleted to ultimately give a frameshift mutation, then the bottom line (green) is the resulting codon translation. The frameshift mutation radically shifted the basic residue content from lysine to arginine. Figure modified from reference (40).

Although the evidence presented supports the hypothesis that protamines evolved from H1 via a protamine-like intermediate, no direct evidence has been shown to link H1 to protamines. This chapter defines the connection through the analysis of protamines from *Marchantia polymorpha* (common liverwort). This plant was one of the major establishers of land plants roughly 400 million years ago and used as a model system to determine evolutionary links between marine and land plants (43). Importantly, it was determined that *Marchantia polymorpha* contained small proteins that travel similar distances on a 2.5M/5% acetic acid PAGE gel to that of arginine rich salmon protamines, implying the presence of arginine rich protamines in this organism (Figure 2.5) (44–46). The amino acid sequences of these protamines are necessary in determining the origin in this species and obtaining additional evidence to support the H1 to protamine evolutionary hypothesis.

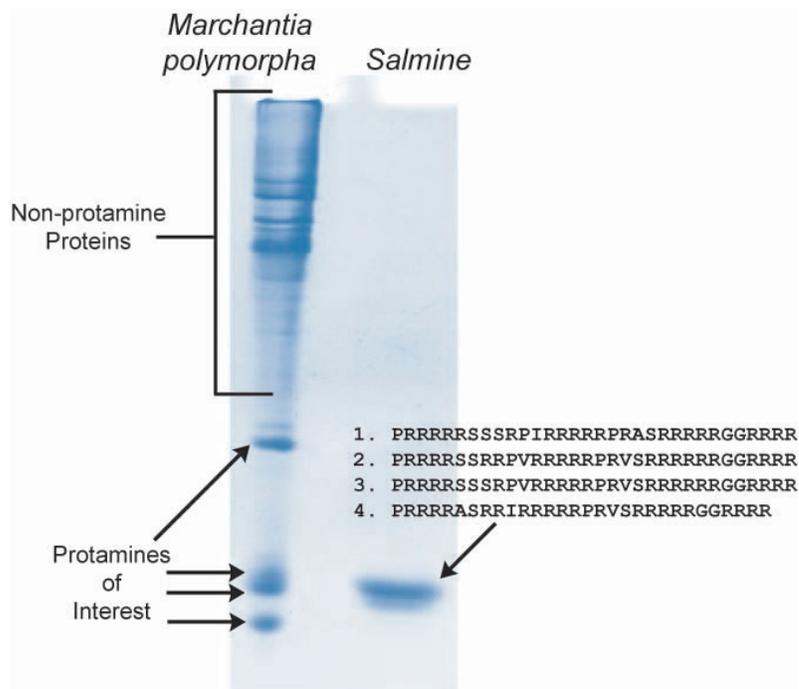


Figure 2.5. 2.5 M Urea/5% Acetic Acid PAGE Gel of Protamines from *Marchantia polymorpha* Antheridiophores. The low pH of the system (~pH 3) causes the majority of the basic side chains to be charged while the acidic residues are not, resulting in a further mobility of basic proteins, such as protamines. The left column shows the presence of four strong bands from *Marchantia polymorpha* while the right column shows the known protamines from salmon (salmine). Note the number of basic residues present in salmine. This would indicate that the bands that travel a similar amount in *Marchantia* are protamines. Image was modified from reference (47).

2.3 Materials

Agilent Technologies (Palo Alto, CA)

1100 Series high performance liquid chromatograph

1100 Series vacuum degasser

PLRP-S 3 μ m diameter reverse phase packing material

Anaspec, Inc. (Fremont, CA)

Vasoactive Intestinal Peptide; Human, porcine, rat; > 95% purity

Eppendorf (Hauppauge, NY)

5414R Benchtop centrifuge

Honeywell (Morristown, NJ)

Burdick and Jackson® Acetonitrile, LC-MS grade

Molex (Lisle, IL)

Polymicro Technologies™ polyimide coated fused silica capillary,
360 μm o.d. x 75 μm i.d.

PQ Corporation (Valley Forge, PA)

Kasil – Potassium silicate solution

Sigma Aldrich (St. Louis, MO)

Angiotensin I acetate salt hydrate, 99% purity (human)

Glacial acetic acid, ≥99.99% trace metal basis

Fluoranthene, >99% purity

Formamide, ≥99.5% (GC)

2-propanol, LC-MS grade

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

Thermo Fisher Scientific (San Jose, CA/Bremen, Germany)

Formic Acid, LC-MS Grade

Orbitrap Fusion™ Tribrid™ mass spectrometer

Pierce® Water, LC-MS Grade

2.4 Methods**2.4.1 Protamine Extraction from Antheridiophores (Performed by Dr. J. Ausió)**

The sperm from *Marchantia polymorpha* were suspended in 100 mM Tris (pH 7.5) and 0.5% Triton X-100 in a 10 μg/mL of tosyl-lysine chloromethyl ketone buffer at a ratio of 20 μL of buffer per antheridiophore. The suspension was centrifuged to pellet the proteins. The pellet was dissolved using 10 μL of 0.4 N HCl per antheridiophore. The resulting extract was precipitated with cold acetone overnight. The precipitate was pelleted by centrifugation, washed with room temperature acetone, and centrifuged again. The pellet was dried via speedvac. The dried proteins were then suspended in 30 μL of water. Figure 2.5 was generated from this suspension depicting four bands of interest in addition to a substantial amount of non-protamine proteins.

2.4.2 Liquid Chromatography and Mass Spectrometry

One μL of the acetone-precipitated protein solution was diluted 10 fold with 0.1% acetic acid in water. ~ 300 nL were pressure-loaded onto a 10 cm of 3 μm diameter, 300 \AA PLRP-s reverse phase column within a 360 μm x 75 μm fused silica capillary integrated with an electrospray tip (48, 49). Additionally, 100 fmol of the internal standards vasoactive intestinal peptide and angiotensin I were also loaded onto the column.

An Agilent Technologies 1100 Series binary HPLC system coupled to a Thermo Orbitrap Fusion tribrid mass spectrometer operating in standard pressure mode was used to characterize the proteins in the sample (50). Proteins were first identified using an isocratic elution at 100% solvent A (0.3% formic acid in water) for 25 minutes at a flow rate of ~ 50 nL/min. This was to account for any highly hydrophilic proteins present that did not retain on column. The column was then washed with solvent A for 25 minutes at ~ 100 nL/min to remove any leftover salts. The proteins retained on column were then eluted using a gradient of 0-10-30-70-100% solvent B (72% ACN, 18% IPA, 10% water, and 0.3% formic acid) in 0-10-20-30-35 minutes at a flow rate of ~ 100 nL/min.

Proteins were first analyzed for molecular weight by a 60,000 resolution Orbitrap MS^1 scan. Proteins with a charge state of 4 or higher were isolated by the quadrupole with a width of 1.5 m/z. Each precursor was fragmented by ETD and CAD. All MS^2 scans were analyzed in the Orbitrap at 60,000 resolution with a precursor target of $2e5$. ETD was performed on precursors with m/z 300-700 using calibrated reaction times (51, 52). CAD was performed on precursors with m/z 400-1500 at 30% normalized collision energy (NCE). Ions were placed on an exclusion list for 30s after 3 scans in 30 seconds. The isocratic identification included an in-source dissociation energy of 40 to remove any acetone or salt adducts from the acetone precipitation.

2.4.3 Determination of C-Terminal Di-aminopropanelation by MS^3

A targeted MS^3 method was used to determine the structure of an unknown z_1^{**} ion (53, 54). Precursor ions at 439.18 m/z were isolated with the quadrupole at a width of 1.5 m/z, a precursor target of $5.0e5$, and a maximum injection time of 300 ms. These ions were then fragmented by ETD using a reaction time of 12 ms with $2e5$ reagent ions. The unknown fragment ion at 215.174 m/z

($z=1$) was then isolated with a width of 1.5 m/z and fragmented by HCD using 30% normalized collisional energy. The resulting fragment ions were analyzed in the Orbitrap at 60,000 resolution.

2.4.4 Data Analysis

The data files were manually inspected for all species present. All major peaks were manually de novo sequenced by averaging all MS^2 scans under the peak using Qual Browser. Sequenced proteins were identified with the NCBI Protein Blast Non-redundant protein sequences database (as of 03/22/2019) against *Marchantia polymorpha* (55, 56).

2.5 Results and Discussion

2.5.1 Initial Identification of Protamines

There was the possibility that the protamines would not retain on the reverse phase column due to their extreme hydrophilicity. Therefore, analysis by the mass spectrometer was carried out immediately upon flowing solvent A. The isocratic elution is shown in Figure 2.6. Since the expected sequence contained primarily arginine residues, the expected charge density of the protamines should be very high, corresponding to peaks between 300 and 600 m/z . Two distinct regions are present in this chromatogram. The first region is at ~3-4 minutes and it contains proteins that did not retain on the column at all. Since there was no interaction with the stationary phase, all the proteins eluted as a defined peak (see Figure 2.7A for the MS^1 at this time point). The second region is from ~5-25 minutes and contains proteins that were rinsed off the column. These proteins had very weak interactions with the stationary phase resulting in a low abundance, broad peak (see Figure 2.7B for the MS^1 at this time point). Similar proteins eluted in both regions of the chromatogram based on the intact molecular weight analysis from these two points in time. However, the green peaks in Figure 2.7A were only observed in the first region of the analysis.

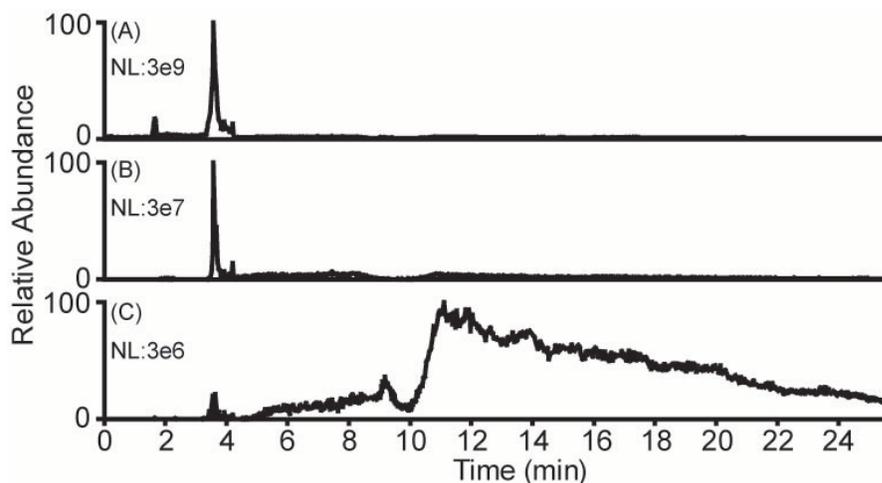


Figure 2.6. Chromatogram of Non-retained Proteins. (A) Total ion current chromatogram. (B) Extracted ion chromatogram of a protamine that did not retain on the column and eluted as a defined peak. (C) Extracted ion chromatogram of a protamine that had very weak interactions with the stationary phase.

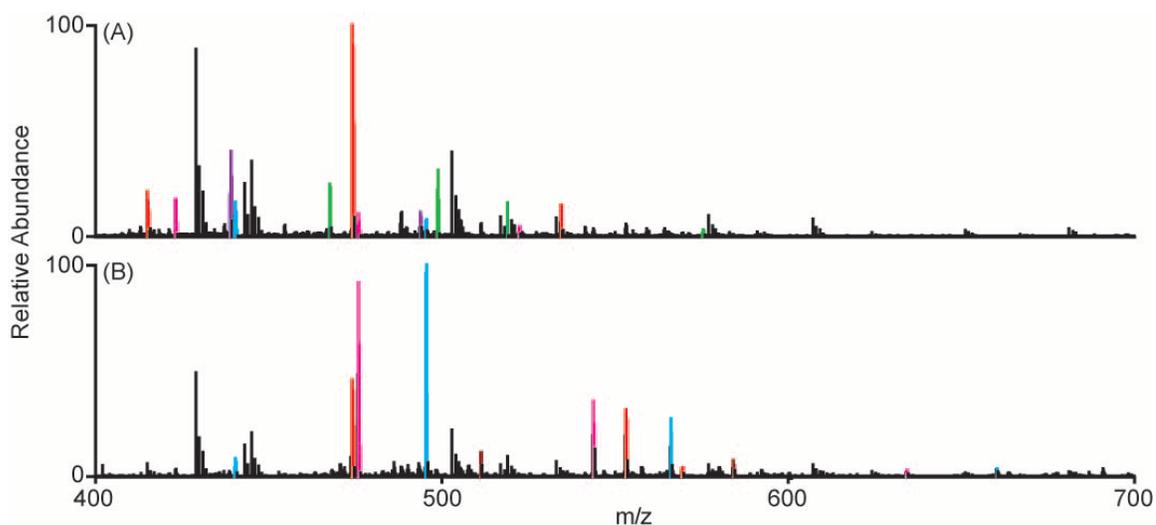


Figure 2.7. MS¹ Scans from the Two Distinct Regions in the Chromatogram of Non-retained Proteins. (A) MS¹ scan taken at ~4 minutes. (B) MS¹ scan taken at ~11 minutes. The different colored peaks correspond to different proteins identified based on their intact molecular weights.

Figure 2.8 shows the gradient elution of the remaining proteins to be identified. The protamines should be some of the first proteins to elute due to their hydrophilicity. Figure 2.8B depicts a protamine peak profile which was broad and eluted over approximately ten minutes. Additionally, there was a high abundance of non-protamine proteins present in the gel analysis of the sample as shown in Figure 2.5. These proteins should have a lower charge density than the

protamines (m/z greater than 700) and have defined peak shapes (as shown in Figure 2.8C). These contaminant proteins were primarily ribosomal proteins.

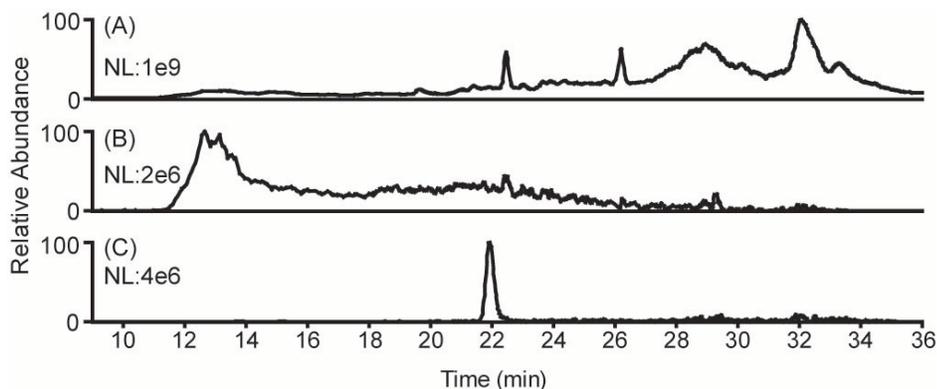


Figure 2.8. Chromatogram of Non-retained Proteins. (A) Total ion current chromatogram. (B) Extracted ion chromatogram of a protamine. These proteins were the first proteins to elute from the column. (C) Extracted ion chromatogram of a contaminant protein. These proteins eluted later in the gradient in relation to the protamines and maintain a defined peak shape.

2.5.2 Discovery of a Novel C-Terminal PTM

ETD is perfectly suited for dissociation of the protamines due to their high charge density. All ETD MS² spectra of each precursor mass were averaged and manually annotated. Figure 2.9 shows an example of the method used for manual analysis. The c-ion series provided sequence coverage for the majority of the identified protamines. When determining the z'-ion series, an unknown mass was observed (noted by the star in Figure 2.9A). This ion at 215.1740 m/z ($z=1$) does not match an expected residue mass for any of the twenty common amino acids. However, characteristic amino acid mass shifts could be observed from this unknown mass that were complementary to the deduced c-ion series (Figure 2.9C). Therefore, this peak must correspond to a modification on the final residue of the protein or on the C-terminus.

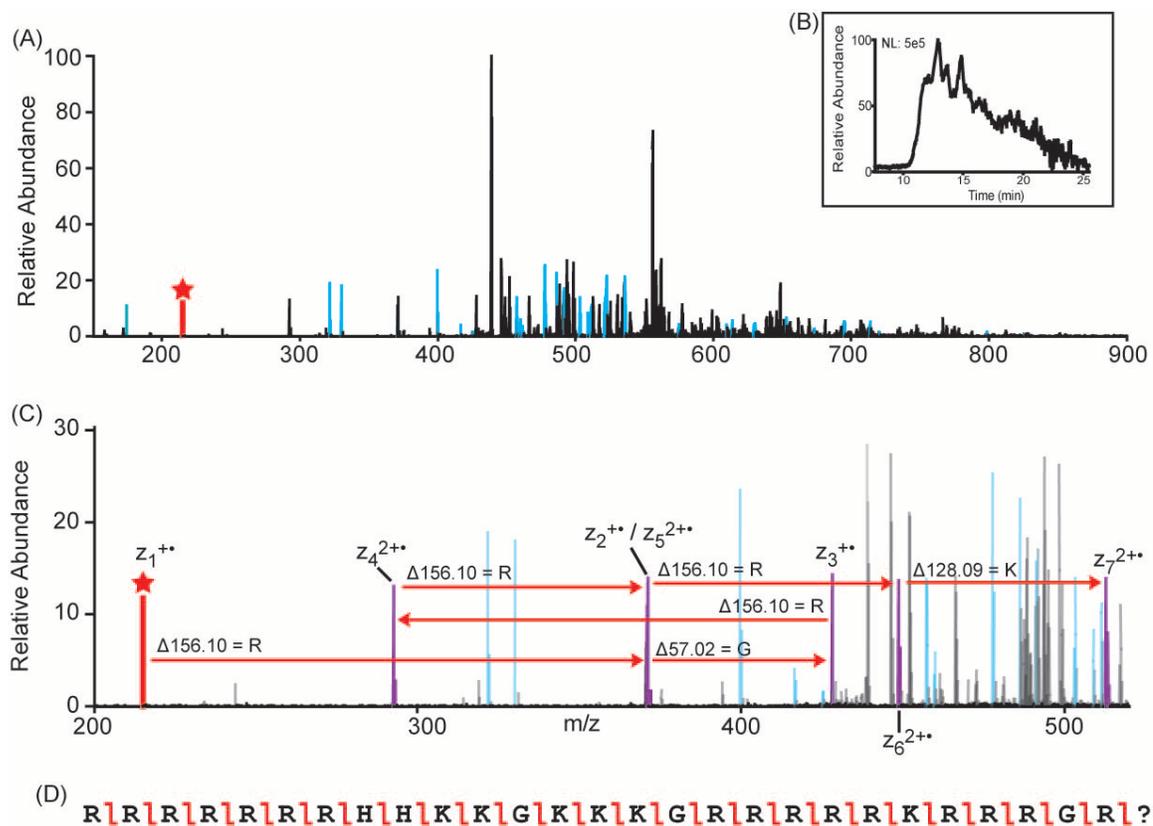


Figure 2.9. Protamine Dissociation Example. (A) ETD MS² of the protamine observed at 438.07 m/z ($z=9$). The red peak and star correspond to the unknown z^+ -ion at 215.1740 m/z. c-ions are colored blue. (B) Extracted ion chromatogram of the peak of interest. Multiple dissociation scans were averaged together due to the broad peak shape to give the MS² shown in (A). (C) Zoom in on the low mass range of the MS² shown in (A). Amino acid mass shifts from the unknown z^+ -ion can be deduced. These ions were complimentary to the observed c-ion series. (D) Fragment ion map depicted the de novo sequenced protamine. Although the z_1^+ residue is unknown, the mass was confirmed to be 215.1740 Da.

The deduced sequence was searched using the NCBI Protein Blast Non-redundant protein sequence database against *Marchantia polymorpha* (last accessed 03/22/2019). This search confirmed the identified sequence and enabled the final unknown residue to be identified as arginine. Therefore, the mass at 215.1740 m/z must be the mass of an arginine plus 56.0738 Da. When de novo sequencing the remaining protamines in the sample, the mass shift of 56.0738 Da was found on additional terminal arginine protamines as well as a terminal lysine protamines (identified at 174.1355 m/z ($z=1$)).

The mass shift does not correspond to any common PTM. A theoretical elemental composition was calculated to determine potential candidates that might explain it. The composition that was the closest based on ppm mass error was $C_9H_{21}N_5O$ (+2.8 ppm error). The modification was determined to be on the C-terminus due to only a single oxygen atom present in the composition. If the modification was on the arginine side chain, then there should be two oxygen atoms in the composition due to the carboxylic acid of the C-terminus. The arginine side chain elemental composition ($C_4H_{11}N_3$) and the protein backbone (C_2HO) were subtracted from the theoretical mass to leave $C_3H_9N_2$ remaining. The exact mass of the remaining elements is 73.0766 Da.

An MS^3 experiment was used to determine the structure of the unknown modification. This type of experiment involves the dissociation of the intact precursor ions, isolation of a particular first generation fragment ion, and an additional dissociation method to fragment the first generation fragment ion. The resulting spectrum, or the MS^3 scan, should contain peaks that correspond to covalent bond breakages. These peaks can then be used to determine the structure of the first generation fragment ion.

The unknown $z^{\prime}1$ -ion at 215.1740 m/z was first formed by an extended ETD reaction time to generate a larger abundance of this ion. The ion trap then isolated these ions and subjected them to HCD to give the MS^3 spectrum shown in Figure 2.10. The single charge was stabilized within the guanidinium group (depicted by the first peak at 59.0479 m/z). The mass differences between peaks can be used to determine the overall structure from the guanidinium group. Every covalent bond was broken during the HCD event as shown by the proposed structures. The data gathered here identified the structure of the unknown modification to be the addition of a di-aminopropane group to the C-terminus (Figure 2.11).

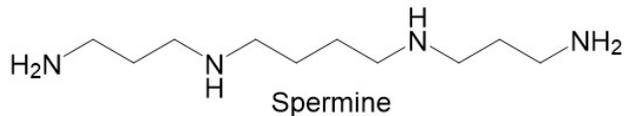
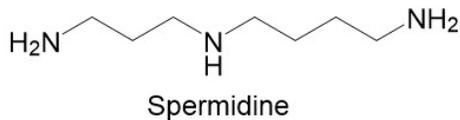
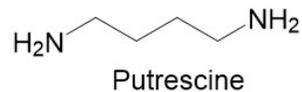
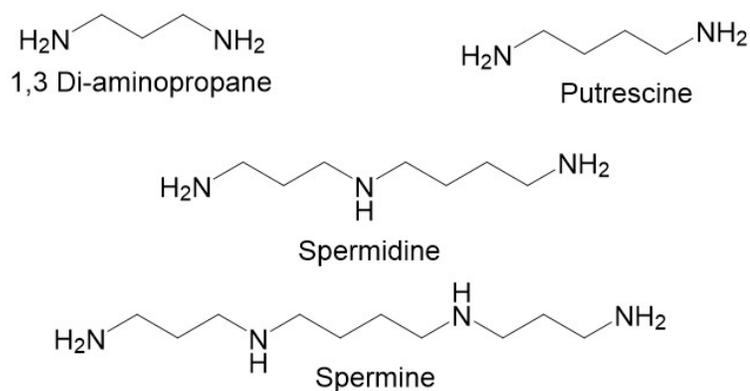


Figure 2.12 Structures of Polyamide Molecules.

2.5.3 Confirmatory Evidence of the Origins of Protamines

Table 2.1 lists all the identified protamines and their respective originator proteins. All of the protamines identified were the result of a post-translational cleavage. More importantly, their originator proteins were all H1 proteins. This is the first time the origin of the observed protamines in a species has been directly linked with H1. This critical piece of evidence solidifies the hypothesis that the evolution of protamines and protamine-like proteins originate from H1.

Table 2.1. Table of Identified Protamines. The lower portion of the table corresponds to low abundance protamines. The amino acids in brackets do not contain confirmatory fragment ions. The residues and sequences within the brackets can be interchanged or potentially contain different residues. The * on the C-terminus of the sequence represents a di-aminopropanelated C-terminus. Table modified from reference (47).

<u>M+H</u>	<u>Sequence</u>	<u>NCBI Accession #</u>	<u>First AA Residue #</u>	<u>Last AA Residue #</u>	<u>Protein Name</u>
3955.50	RRRRRRKSRRRRRRRRRS- KGSRSRRR*	BAU71552	182	209	PL-1
3799.41	RRRRRRKSRRRRRRRRRSK- GSRSPRR*	BAU71552	182	208	PL-1
3942.56	RRRRRRHHKKGKKKGRRR- RRKRRRGRR*	BAU71552	102	129	PL-1
3313.16	RRRRRRRRRRRRRRSRR- RRGRK*	BAU71552	146	167	PL-1
4082.51	RSTSRSRGRRRRRRRHR- RRRGRRRRGRK*	PTQ35141	207	236	PL-2
7466.75	RRRRRRRHGRKGKKPKRHR- RRRRRRGRRKAHRGRKKKR- GGRRKGKRRHRRRR*	PTQ32223	162	216	PL-3
3780.36	RRRRRRKSRRRRRRRR- RR[SKGSH]SPRR*	BAU71552	182	208	PL-1
4139.33	RRRRRRKSRRRR[RRTV- DRRRRRRRRRRR]	BAU71552	182	209	PL-1
4366.65	RRRRRRKSRRRRRRRR- RRSKGSRSRRRRHSSR	BAU71552	182	213	PL-1
7819.01	RRRRRRRHGRKGKKPK- RHRRRRRRRRRGRRKA- HRGRKKKRGRRKGKR- RHRRRRR[GKSH]	PTQ32223	162	220	PL-3

The addition of di-aminopropane most likely formed following the activation of the carboxyl group within the spermatozoa. This would enable the reaction generating this observed PTM. Although the reasoning behind this modification is not definitively understood, it is hypothesized that this modification enables a stronger interaction between protamines and the DNA backbone through the removal of the only negatively charged group, the C-terminus, and replacing it with a positively charged group.

2.6 Conclusions

The investigation of the protamines from *Marchantia polymorpha* provided critical insight into the origins of protamines. Previous studies have only been able to link protamines to protamine-like protein precursors (36, 39, 40, 42). The protamine-like proteins were related to H1 through assumptions based on the folding structures of the proteins, though no direct evidence to H1 was available. The protamines identified here, however, provide the concrete evidence that was missing in the previous studies of protamines in the animal kingdom. These protamines were the result of the post translational cleavage from the direct sequence of H1 precursors. Additionally, the protamines were also found to contain the novel di-aminopropanelated PTM on the C-terminus. The reasoning behind this modification remains elusive and requires further studies. However, replacing the only negative charge on the protein with a positive charge implies that this modification further enhances the binding ability of the protamine to the negatively charged DNA backbone. Together, the results presented here provide the evidence necessary to solidify the claim that protamines indeed evolved from H1.

2.7 References

1. Berg JM, Tymoczko JL, Stryer L, Stryer L (2007) *Biochemistry* (W.H. Freeman, New York).
2. Mariño-Ramírez L, Kann MG, Shoemaker BA, Landsman D (2005) Histone structure and nucleosome stability. *Expert Rev Proteomics* 2(5):719–729.
3. Cutter AR, Hayes JJ (2015) A brief review of nucleosome structure. *FEBS Lett* 589(20 Pt A):2914–2922.
4. Bednar J, et al. (2017) Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol Cell* 66(3):384–397.e8.
5. Garcia BA, Shabanowitz J, Hunt DF (2007) Characterization of histones and their post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* 11(1):66–73.
6. Garcia BA, et al. (2007) Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protoc* 2(4):933–938.
7. Huang H, Lin S, Garcia BA, Zhao Y (2015) Quantitative Proteomic Analysis of Histone Modifications. *Chem Rev* 115(6):2376–2418.
8. Luger K, Dechassa ML, Tremethick DJ (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* 13(7):436–447.
9. Razin S V, Gavrillov AA (2014) Chromatin without the 30-nm fiber: constrained disorder instead of hierarchical folding. *Epigenetics* 9(5):653–657.
10. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242.

11. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389(6648):251–260.
12. Happel N, Doenecke D (2009) Histone H1 and its isoforms: Contribution to chromatin structure and function. *Gene* 431(1):1–12.
13. Fyodorov D V, Zhou B-R, Skoultchi AI, Bai Y (2017) Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol* 19:192.
14. Wolffe AP (1997) Histone H1. *Int J Biochem Cell Biol* 29(12):1463–1466.
15. KASINSKY HE, LEWIS JD, DACKS JB, AUSIÓ J (2001) Origin of H1 linker histones. *FASEB J* 15(1):34–42.
16. The histone H1 family: specific members, specific functions? (2008) *Biol Chem* 389:333.
17. Hergeth SP, Schneider R (2015) The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep* 16(11):1439–1453.
18. Griswold MD (2015) Spermatogenesis: The Commitment to Meiosis. *Physiol Rev* 96(1):1–17.
19. de Kretser DM, Loveland KL, Meinhardt A, Simorangkir D, Wreford N (1998) Spermatogenesis. *Hum Reprod* 13(suppl_1):1–8.
20. Balhorn R, Brewer L, Corzett M (2000) DNA condensation by protamine and arginine-rich peptides: Analysis of toroid stability using single DNA molecules. *Mol Reprod Dev* 56(S2):230–234.
21. Oliva R (2006) Protamines and male infertility. *Hum Reprod Update* 12(4):417–435.
22. Akmal M, et al. (2016) The important role of protamine in spermatogenesis and quality of sperm: A mini review. *Asian Pacific J Reprod* 5(5):357–360.
23. Gardiner-Garden M, Ballesteros M, Gordon M, Tam PP (1998) Histone- and protamine-DNA association: conservation of different patterns within the beta-globin domain in human sperm. *Mol Cell Biol* 18(6):3350–3356.
24. Balhorn R (2007) The protamine family of sperm nuclear proteins. *Genome Biol* 8(9):227.
25. Bao J, Bedford MT (2016) Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. *Reproduction* 151(5):R55–R70.
26. Harrison CJ, Allen TD, Britch M, Harris R (1982) High-resolution scanning electron microscopy of human metaphase chromosomes. *J Cell Sci* 56(1):409 LP – 422.
27. Brewer LR (2011) Deciphering the structure of DNA toroids. *Integr Biol* 3(5):540–547.
28. Lewis JD, Song Y, de Jong ME, Bagha SM, Ausió J (2003) A walk through vertebrate and invertebrate protamines. *Chromosoma* 111(8):473–482.
29. Jeong E, Kim H, Lee S-W, Han K (2003) Discovering the interaction propensities of amino acids and nucleotides from protein-RNA complexes. *Mol Cells* 16(2):161–167.
30. Luscombe NM, Laskowski RA, Thornton JM (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* 29(13):2860–2874.
31. DeRouchey J, Hoover B, Rau DC (2013) A comparison of DNA compaction by arginine and lysine peptides: a physical basis for arginine rich protamines. *Biochemistry* 52(17):3000–3009.
32. Jenkins T, Carrell D (2012) Dynamic alterations in the paternal epigenetic landscape following fertilization. *Front Genet* 3:143.
33. McLay DW, Clarke HJ (2003) Remodelling the paternal chromatin at fertilization in mammals. *Reproduction* 125(5):625–633.
34. Tirmarche S, Kimura S, Dubrulle R, Horard B, Loppin B (2016) Unlocking sperm chromatin at fertilization requires a dedicated egg thioredoxin in *Drosophila*. *Nat Commun*

- 7(1):13539.
35. E Kasinsky H, Maria Eirin-Lopez J, Ausió J (2011) Protamines: structural complexity, evolution and chromatin patterning. *Protein Pept Lett* 18(8):755–771.
 36. Ausió J (1999) Histone H1 and Evolution of Sperm Nuclear Basic Proteins. *J Biol Chem* 274(44):31115–31118.
 37. Subirana JA, Cozcolluela C, Palau J, Unzeta M (1973) Protamines and other basic proteins from spermatozoa of molluscs. *Biochim Biophys Acta - Protein Struct* 317(2):364–379.
 38. Ausio J, et al. (1987) Structural characterization of the trypsin-resistant core in the nuclear sperm-specific protein from *Spisula solidissima*. *Biochemistry* 26(4):975–982.
 39. Saperas N, Ausió J (2013) Sperm Nuclear Basic Proteins of Tunicates and the Origin of Protamines. *Biol Bull* 224(3):127–136.
 40. Lewis JD, et al. (2004) Histone H1 and the origin of protamines. *Proc Natl Acad Sci* 101(12):4148–4152.
 41. Watson CE, Gauthier SY, Davies PL (1999) Structure and expression of the highly repetitive histone H1-related sperm chromatin proteins from winter flounder. *Eur J Biochem* 262(2):258–267.
 42. Eirín-López JM, Frehlick LJ, Ausió J (2006) Protamines, in the footsteps of linker histone evolution. *J Biol Chem* 281(1):1–4.
 43. Bowman JL, Araki T, Kohchi T (2016) *Marchantia*: past, present and future. *Plant Cell Physiol* 57(2):205–209.
 44. Reynolds WF, Wolfe SL (1978) Changes in basic proteins during sperm maturation in a plant, *Marchantia polymorpha*. *Exp Cell Res* 116(2):269–273.
 45. Higo A, et al. (2016) Transcriptional framework of male gametogenesis in the liverwort *Marchantia polymorpha* L. *Plant Cell Physiol* 57(2):325–338.
 46. Bowman JL, et al. (2017) Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell* 171(2):287–304.e15.
 47. D'Ippolito RA, et al. (2019) Protamines from liverwort are produced by post-translational cleavage and C-terminal di-aminopropanelation of several male germ-specific H1 histones. *J Biol Chem* 294(44):16364–16373.
 48. Martin SE, Shabanowitz J, Hunt DF, Marto JA (2000) Subfemtomole MS and MS/MS Peptide Sequence Analysis Using Nano-HPLC Micro-ESI Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal Chem* 72(18):4266–4274.
 49. Udeshi ND, Compton PD, Shabanowitz J, Hunt DF, Rose KL (2008) Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. *Nat Protoc* 3:1709.
 50. Senko MW, et al. (2013) Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal Chem* 85(24):11710–11714.
 51. Compton PD, Strucl J V, Bai DL, Shabanowitz J, Hunt DF (2012) Optimization of electron transfer dissociation via informed selection of reagents and operating parameters. *Anal Chem* 84(3):1781–1785.
 52. Rose CM, et al. (2015) A calibration routine for efficient ETD in large-scale proteomics. *J Am Soc Mass Spectrom* 26(11):1848–1857.
 53. Olsen J V, Mann M (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci U S A* 101(37):13417.

54. Macek B, Waanders LF, Olsen J V, Mann M (2006) Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *Mol Cell Proteomics* 5(5):949–958.
55. Boratyn GM, et al. (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res* 41(W1):W29–W33.
56. Zaretskaya I, et al. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* 36(suppl_2):W5–W9.
57. Fincato P, et al. (2010) Functional diversity inside the Arabidopsis polyamine oxidase gene family. *J Exp Bot* 62(3):1155–1168.
58. Yu C-H, Chou C-C, Lee Y-J, Khoo K-H, Chang G-D (2015) Uncovering protein polyamination by the spermine-specific antiserum and mass spectrometric analysis. *Amino Acids* 47(3):469–481.
59. Pegg AE (2006) Regulation of Ornithine Decarboxylase. *J Biol Chem* 281(21):14529–14532.
60. Pegg AE (2009) Mammalian polyamine metabolism and function. *IUBMB Life* 61(9):880–894.

Concentration Effects of an Immobilized Aspergillopepsin I Enzyme Reactor

3.1 Abstract

Immobilized enzymes are used to increase effective enzyme concentration while minimizing sample preparation time and sample handling. This technique is particularly useful to effectively control the reaction when using enzymes that are nonspecific in nature, such as Aspergillopepsin I. This enzyme has the unique capabilities of functioning in 8M urea and acidic pH for approximately one hour. Digestions using immobilized Aspergillopepsin I are highly reproducible under specific concentrations and digestion times. This chapter details how protein concentration affects digestion profiles when using an immobilized Aspergillopepsin I reactor. When the concentration of the protein decreases, the system becomes externally diffusion limited resulting in an increased number of observed enzymatic cleavages. Several additives were explored to counteract these concentration effects, of which protamine was found to be the ideal additive.

3.2 Introduction

3.2.1 Immobilized Enzymes

Enzymatic digestions are an invaluable step to effectively analyze protein samples by mass spectrometry. Enzymatic reactions are typically carried out overnight in solution with the enzyme at 50-100 times lower in abundance by mass than the substrate. Low protease concentrations are used to minimize protease autodigestion, which consumes valuable enzymes and introduces contaminant peptides to the analysis (1). Immobilization of the enzyme provides an advantageous alternative to circumvent these issues while providing additional benefits to digestion.

Immobilization of an enzyme involves the covalent linkage of the enzyme to a solid support (2). This solid support can include solid or porous beads, magnetic particles, porous membranes, and walls of a capillary (3–6). Ideally, all conjugated enzymes retain their activity. However, a percentage of enzymes are conjugated in such a way that they are rendered inactive (7, 8). This can occur if the catalytic pocket is inaccessible to the solution or if the catalytic pocket is altered by

the conjugation process (Figure 3.1). Since the enzymes are unable to freely move, the rate of autolysis precipitously drops (9). The enzyme can be easily removed from the substrate through removal of the support, enabling simple quenching of the enzymatic reaction as well as the ability to reuse the enzyme.

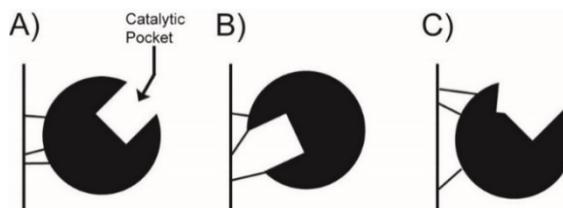


Figure 3.1. Enzyme Conformations when Bound to a Solid Support. A) Enzyme is bound in a favorable conformation in which the catalytic pocket is undisturbed and solution exposed. B) Enzyme is bound in an unfavorable conformation in which the catalytic pocket is not solution exposed. C) Enzyme is bound in an unfavorable conformation in which the catalytic pocket is distorted by the conjugation process.

Immobilized enzymes introduce diffusional effects that impact the digestion. Since the enzyme is bound to the support, the high concentration of the enzyme imparts solvent molecules acquiring an ordered structure forming a diffusion layer (Figure 3.2) (7, 10, 11). The thickness of this layer can be altered by varying solution viscosity (increase viscosity, increase thickness) and flow rate of the bulk solution (increase flow rate, decrease thickness) (8). The relative concentration of the substrate in this layer is much lower than the concentration of the substrate in the bulk solution. This creates a concentration gradient in which the substrate from the bulk solution diffuses from high concentration to the low concentration surface of the support to undergo a catalytic reaction (7, 8). The products then diffuse away from the support through the diffusion layer and back into the bulk solution.

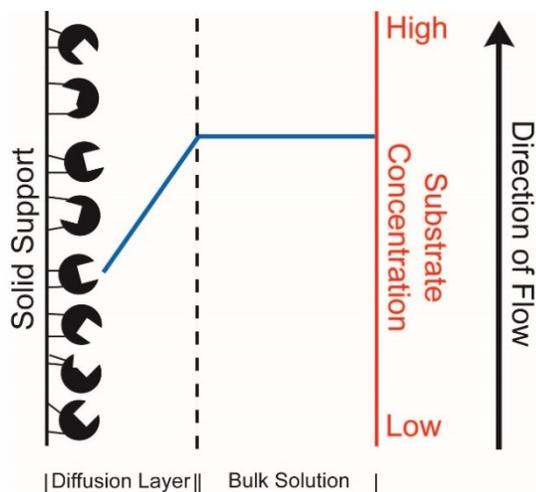


Figure 3.2 Simplified Depiction of Diffusion Layer. The blue line depicts the approximate concentration of the substrate. Since molecules traverse from high concentration to low concentration, substrate molecules are pulled by the gradient towards the surface of the support.

Figure was redrawn and modified from references (8, 11).

The diffusion layer causes fewer substrate molecules to be presented to the enzymes. Since the surface of the support contains a high concentration of enzyme, the ratio of enzyme to substrate increases substantially and thus increases the overall rate of substrate digestion (12). The time needed to fully digest a protein is drastically reduced from overnight to minutes (1). Since such short digestion times are needed, immobilized enzymes are commonly coupled directly in-line with LC-MS/MS analyses (13–15).

3.2.2 Aspergillopepsin I

It is important to generate large peptides to better enhance ETD fragmentation and sequence coverage when using the limited digestion approach (as stated in 1.4.3). Specific proteases rely on a specific residue to be evenly spaced throughout the target protein to effectively generate mid-ranged peptides (16). Conversely, nonspecific proteases, such as pepsin and proteinase K, can be used to provide multiple sites for enzymatic cleavages and generate ideal sized peptides (16, 17). This work utilizes the nonspecific protease *Aspergillopepsin I*, an aspartic protease.

Aspergillopepsin I originates from the genus *Aspergillus*, a fungus. *Aspergillus* is commonly found in everyday life. For example, *Aspergillus niger* is responsible for almost all of the citric acid

produced commercially (18, 19). Additionally, *Aspergillus oryzae* is used to ferment rice (sake), soybeans (soy sauce), and other starchy ingredients (20–22). Humans are frequently exposed to inhalation of these spores without issue; yet those with a weakened immune system or those allergic to certain proteins will exhibit symptoms of the disease aspergillosis (23, 24).

Aspergillopepsin I is a 325 residue protein with a molecular weight of ~34.2 kDa and a pI of ~4 (see Figure 3.3 for structure and Figure 3.4 for sequence) (25–27). It contains 3 aspartic acid residues to hydrolyze peptide bonds: D32, D76, and D214 (26). D32 and D214 are essential to carry out the reaction with a water molecule (see Figure 3.5). D76, while not directly involved in the reaction, is needed for accepting lysine residues in the catalytic pocket (28). The mutation of D76 to a serine drastically decreases the activity of Aspergillopepsin I towards basic residues (28, 29).

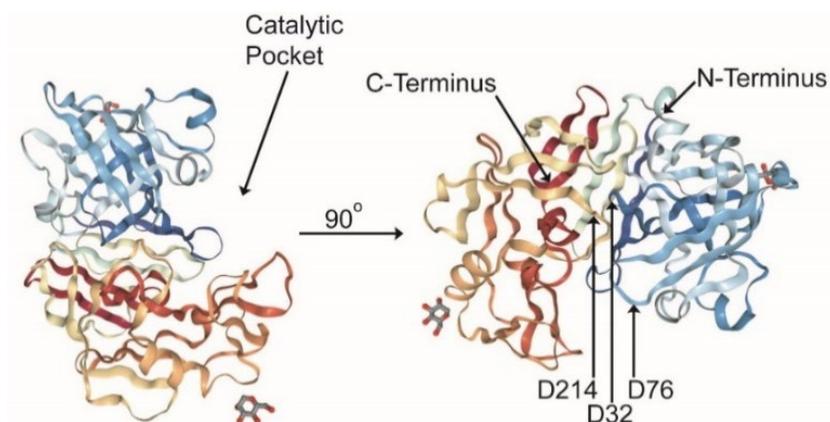


Figure 3.3. Crystal Structure of Aspergillopepsin I. D32, D76, and D214 of the active site are noted. Crystal structure provided by PDB (ID: 1IBQ) (30). Solving of structure by reference (31).

```

1  S K G S A V T T P Q N N D E E Y L T P V T V G K S 25
26 T L H L D F D T G S A D L W V F S D E L P S S E Q 50
51 T G H D L Y T P S S S A T K L S G Y S W D I S Y G 75
76 D G S S A S G D V Y R D T V T V G G V T T N K Q A 100
101 V E A A S K I S S E F V Q D T A N D G L L G L A F 125
126 S S I N T V Q P K A Q T T F F D T V K S Q L D S P 150
151 L F A V Q L K H D A P G V Y D F G Y I D D S K Y T 175
176 G S I T Y T D A D S S Q G Y W G F S T D G Y S I G 200
201 D G S S S S S G F S A I A D T G T T L I L L D D E 225
226 I V S A Y Y E Q V S G A Q E S Y E A G G Y V F S C 250
251 S T D L P D F T V V I G D Y K A V V P G K Y I N Y 275
276 A P V S T G S S T C Y G G I Q S N S G L G L S I L 300
301 G D V F L K S Q Y V V F N S E G P K L G F A A Q A 325

```

Figure 3.4. Amino Acid Sequence of the Active Form of Aspergillopepsin I. D32, D76, and D214 of the active site are highlighted in red.

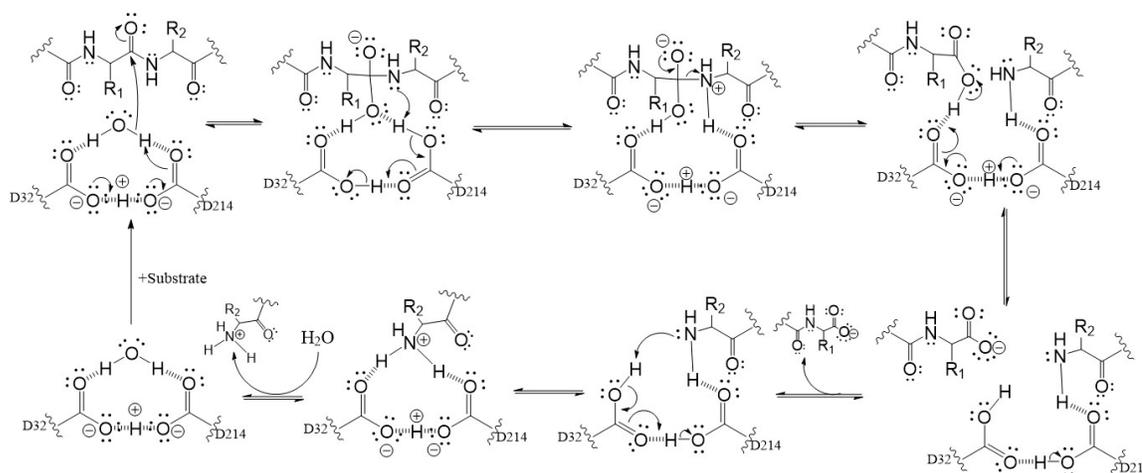


Figure 3.5 Catalytic Mechanism of Aspartic Proteases. D32 and D214 correspond to the essential aspartic acid residues of Aspergillopepsin I. Figure modified from reference (32).

Aspergillopepsin I has additional beneficial attributes to its nonspecificity. Its active pH range is from ~2.5-6 with the optimal pH of ~3 (26, 27). The activity of the enzyme in acidic conditions prevents sample handling artifacts, such as disulfide bond scrambling and deamination (33, 34). Aspergillopepsin I is also functional in 8M urea (3). Urea acts as a denaturing agent to disrupt the tertiary structure of proteins for increased solvent accessibility (35, 36). Denatured and mostly unfolded sample proteins provide Aspergillopepsin I with increased accessibility, allowing it to digest the protein anywhere along the backbone. However, if some tertiary structure remains, then this could bias Aspergillopepsin I towards a more accessible residue.

3.2.3 Immobilized Aspergillopepsin I

If Aspergillopepsin I is unconstrained from digesting the target protein, then the end result will be small peptides (3, 37). Since the goal of using Aspergillopepsin I is to generate large peptide products, the digestion must be stopped after minimal cleavages. Most specific proteases are inactivated by acidifying the reaction, rendering the protease inactive. However, another method must be employed specifically for Aspergillopepsin I since it functions in acidic conditions. Through immobilization and construction of an enzyme reactor, the digestion can be stopped quickly and efficiently by removing the substrate from enzyme bound solid support.

Aspergillopepsin I is conjugated to aldehyde functionalized beads (Figure 3.6) (38). Conjugated beads are pressure loaded into a fritted fused silica capillary to create an enzyme reactor (Figure 3.7). The protein sample is flowed over the packed bed of immobilized enzymes using a pressure vessel to facilitate digestion. The digestion stops as soon as the bulk solution is expelled from the capillary. The first generation of this enzyme reactor by Zhang et al. utilized 20 μm porous beads to demonstrate the high reproducibility of Aspergillopepsin I (3). Successful digestion and full sequence reconstruction of a monoclonal antibody was obtained in 3 LC-MS/MS analyses. The second generation by Hinkle et al. utilized 1 μm solid sphere beads (39). This drop in bead size and elimination of pores reduced the Eddy diffusion of the analyte over the packed bed to give a more uniform path through the column, and thus a more uniform digestion (40). Additionally, the introduction of the Thermo Orbitrap Fusion enabled the full sequence reconstruction of a monoclonal antibody in a single LC-MS/MS analysis (41).

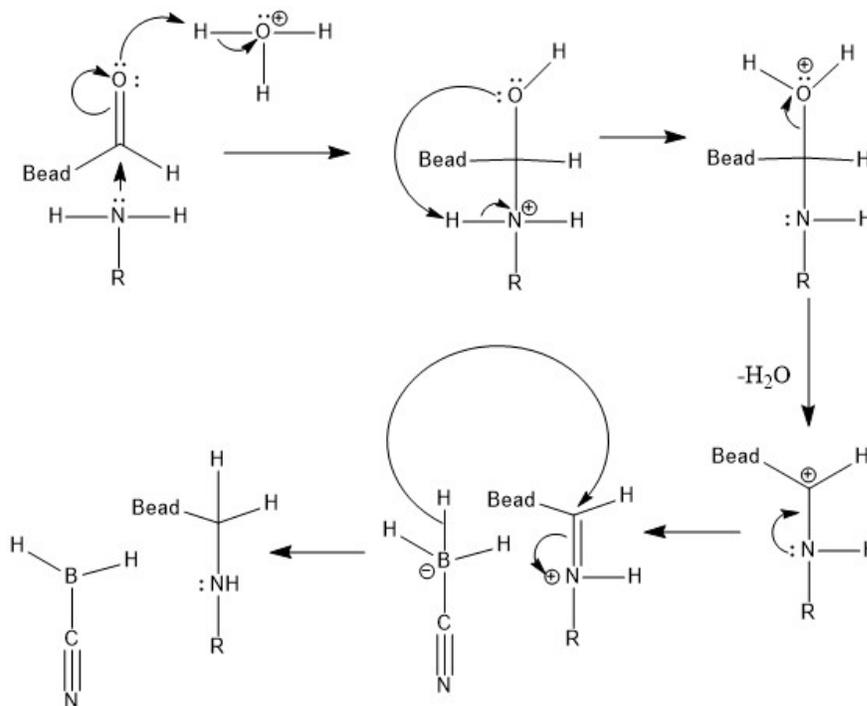


Figure 3.6 Primary Amine Conjugation Mechanism with Aldehyde Functionalized Beads. When conjugating an enzyme to beads, the primary amine represents the N-terminus or a lysine side chain.

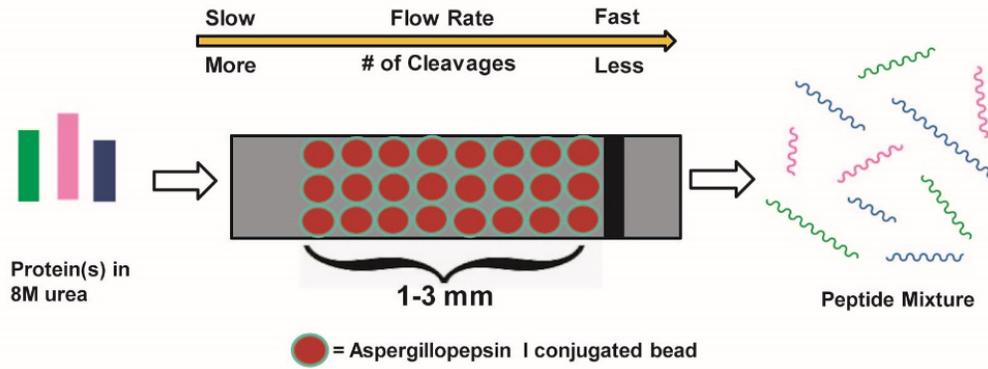


Figure 3.7 Schematic of General Process of Enzyme Reactor Use (39).

The reaction occurs only while flowing substrate over the packed bed. Thus, varying the flow rate of the analyte and/or the length of the packed bed will dictate the digestion time, which can be calculated using the following equation:

$$(3.1) \text{ digestion time (s)} = \pi * \left(\frac{\text{inner diameter [cm]}}{2} \right)^2 * \text{bead porosity} * 60000 * \frac{\text{bed length (cm)}}{\text{flow rate} \left(\frac{\mu\text{L}}{\text{min}} \right)}$$

The inner diameter corresponds to the fused silica used, which was 150 μm in this dissertation. The bead porosity signifies the percentage of packed bed length that consists of liquid. The porosity of 1 μm beads was determined to be $\sim 28.5\%$ (39). 60,000 is a unit conversion from $\text{mL} * \text{min}$ to $\mu\text{L} * \text{s}$. Equation 3.1 can be simplified using these constant values:

$$(3.2) \text{ digestion time (s)} = \pi * \left(\frac{0.015 \text{ [cm]}}{2} \right)^2 * 0.285 * 60000 * \frac{\text{bed length (cm)}}{\text{flow rate} \left(\frac{\mu\text{L}}{\text{min}} \right)}$$

$$\text{digestion time (s)} = 3.02 * \frac{\text{bed length (cm)}}{\text{flow rate} \left(\frac{\mu\text{L}}{\text{min}} \right)}$$

This equation depicts the digestion time is proportional to the bed length and inversely proportional to the flow rate. The longer the digestion time, the longer the protein is in contact with the enzyme, which will result in more enzymatic cleavages. Figure 3.8 shows how digestion time affects the digestion products of apomyoglobin.

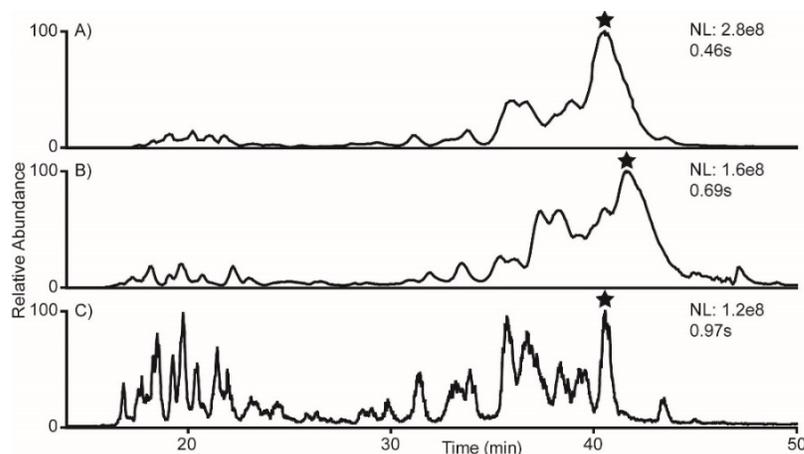


Figure 3.8. Total Ion Current Chromatograms for Different Digestion Times of Apomyoglobin. The peak signified by a star represents undigested apomyoglobin A) Digestion time = ~ 500 ms. B) Digestion time = ~ 700 ms. C) Digestion time = ~ 1 s.

The ability to choose a digestion time allows digests of Aspergillopepsin I to be tuned to a distribution of peptide molecular weight ranges that will satisfy experimental needs. Importantly, Aspergillopepsin I completes these digestions on the time scale of hundreds of milliseconds (3). This drastically reduces the amount of time needed to digest a protein and thus decrease the amount of time needed for sample preparation.

Aspergillopepsin I is functional in 8 M urea for up to one hour (3). It is important to have a denaturing agent when using this enzyme reactor on this timescale. Shown in Figure 3.9 is the digestion of apomyoglobin for ~ 1 s in 8 M urea (Figure 3.9A) and without urea (Figure 3.9B). The removal of urea from the digestion buffer prevents apomyoglobin from denaturing. The retention of higher order structures limits Aspergillopepsin I from accessing the entire backbone, which results in minimal observed digestion products.

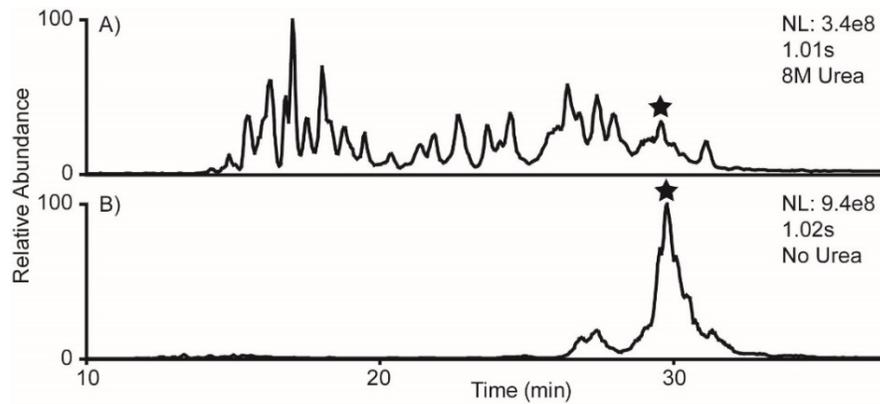


Figure 3.9 Urea Impact on Aspergillopepsin I Digestions. The peak signified by a star represents undigested apomyoglobin. A) Apomyoglobin digested in 8M urea. B) Apomyoglobin digested without urea.

Although Aspergillopepsin I is overall nonspecific, it does have some preferred residues. Aspergillopepsin I typically favors the cleavage of hydrophobic residues and C-terminal to lysine (26, 37). Figure 3.10 shows the frequency of amino acids at the P1 position (cleaves C-terminally to) and P1' position (cleaves N-terminally to) based on internal data sets generated by immobilized Aspergillopepsin I. Although it is true that hydrophobic residues are preferred based on this data set, Aspergillopepsin I also shows preference to charged residues minus histidine at the P1 position.



Figure 3.10 Relative Cleavage Specificity of Immobilized Aspergillopepsin I. The larger the amino acid, the more frequently it is observed at that position.

Previous work by Zhang et al. and Hinkle et al. demonstrated the ability to successfully digest monoclonal antibodies with high reproducibility (3, 39). However, digests performed were at an analyte concentration of 0.2 $\mu\text{g}/\mu\text{L}$. This value is relatively concentrated and not all biological samples can be concentrated to this value. Additionally, if multiple proteins are present within a sample to be digested at varying concentrations, how will the concentration, both overall and of each individual protein, affect how the protein is digested? The following chapter details the digestion of standard proteins at multiple concentrations to characterize how concentration affects digestion by immobilized Aspergillopepsin I.

3.3 Materials

Agilent Technologies (Palo Alto, CA)

1100 Series high performance liquid chromatograph

1100 Series vacuum degasser

PLRP-S 3 μm diameter reverse phase packing material

Anaspec, Inc. (Fremont, CA)

Vasoactive Intestinal Peptide; Human, porcine, rat; > 95% purity

Bio-Rad (Hercules, CA)

Tris base (Tris[hydroxymethyl]aminomethane), ≥99.8% pure, powder

Eppendorf (Hauppauge, NY)

5414R Benchtop centrifuge

Honeywell (Morristown, NJ)

Burdick and Jackson® Acetonitrile, LC-MS grade

Labconco Corporation (Kansas City, MO)

Centrivap centrifugal vacuum concentrator

Molex (Lisle, IL)

Polymicro Technologies™ polyimide coated fused silica capillary:

360 μm o.d. x 75 μm i.d., 360 μm o.d. x 150 μm i.d.

PQ Corporation (Valley Forge, PA)

Kasil – Potassium silicate solution

Protein Metrics (Cupertino, CA)

Byonic™

Sigma Aldrich (St. Louis, MO)

2-propanol, LC-MS grade

Angiotensin I acetate salt hydrate, 99% purity (human)

Apomyoglobin from equine skeletal muscle, protein sequencing standard, lyophilized powder

Arg-Phe acetate salt, ≥98%

β-Lactoglobulin from bovine milk, ≥85% (PAGE), lyophilized powder

Bovine Serum Albumin, heat shock fraction, protease free, ≥98%

Lysozyme chloride from chicken egg white, ≥80%

Formamide, ≥99.5% (GC)

Fluoranthene, >99% purity

Glacial acetic acid, ≥99.99% trace metal basis

Lys-Gly hydrochloride, ≥97%

Lys-Lys-Lys-Lys-Lys (penta-lysine), ≥55% peptide basis

N-(2-Aminoethyl)maleimide trifluoroacetate salt, ≥95% (HPLC)

Protamine sulfate salt from salmon, Grade X, amorphous powder (acquired by Dr. Juan Ausió)

Protease from *Aspergillus saitoi* (Aspergillopepsin I), Type XIII

Trypsin inhibitor from *Glycine max* (soybean), powder

Trypsinogen from bovine pancreas, essentially salt-free, lyophilized powder

Ubiquitin from bovine erythrocytes, ≥98% (SDS-PAGE), essentially salt-free, lyophilized powder

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

Thermo Fisher Scientific (San Jose, CA/Bremen, Germany)

Aldehyde/Sulfate Latex Beads, 4% w/v, 1.0 μm

Ammonium Acetate (Crystalline/Certified ACS) [via Fisher Chemical]

Formic Acid, LC-MS Grade

LTQ-FT-ICR hybrid mass spectrometer, custom modified with front-end ETD

Orbitrap Elite™ hybrid mass spectrometer, custom modified with front-end ETD

Pierce® Water, LC-MS Grade

Tris Hydrochloride, powder

Urea (>99.0%)

3.4 Methods

3.4.1 Aspergillopepsin I Bead Conjugation

~250 μL of 1.0 μm aldehyde/sulfate latex beads, corresponding to ~10 mg of beads, were measured into an Eppendorf tube. The beads were centrifuged at 10,500 rcf for 1 min to pellet the beads and the supernatant was discarded. The beads were then washed three times with 500 μL of water using the same procedure.

A 10 mg/mL Aspergillopepsin I solution was created by dissolving the enzyme in a saturated solution of Na₂SO₄ (~0.28 g/mL). 300 μL of this solution was added to the washed beads.

5 μL of 80 mg/mL NaCNBH_3 in saturated Na_2SO_4 was carefully added to the suspended beads. The beads were shaken at room temperature for 20-24 hours.

The beads were pelleted and the Aspergillopepsin I solution was removed. The beads were washed three times with 500 μL of water. 300 μL of 0.2 M Tris (pH \sim 6.5) was added to the washed beads to block any unconjugated aldehyde sites. 5 μL of 80 mg/mL NaCNBH_3 in water was carefully added to the suspended beads. The beads were shaken at room temperature for 3 hours.

The beads were pelleted and the Tris solution was removed. The beads were washed three times with 500 μL of water. The beads were stored dry at 4°C.

3.4.2 Protein Preparation for Digestion

One nmol (\sim 17 μg) stocks of dried apomyoglobin were reconstituted with digestion buffer (8 M urea in 50 mM ammonium acetate at pH \sim 3.9) to the desired concentration of either 0.2 $\mu\text{g}/\mu\text{L}$ (\sim 11.8 pmol/ μL), 0.1 $\mu\text{g}/\mu\text{L}$ (\sim 5.9 pmol/ μL), 0.05 $\mu\text{g}/\mu\text{L}$ (\sim 2.9 pmol/ μL), or 0.02 $\mu\text{g}/\mu\text{L}$ (\sim 1.2 pmol/ μL).

All other proteins required disulfide bonds to be reduced and alkylated. Protein stock solutions from powders were first prepared by creating 5 $\mu\text{g}/\mu\text{L}$ solutions of protein in 0.1% acetic acid. Soybean trypsin inhibitor was created with water since the pI caused the protein to precipitate in acidic conditions. 6 μL of 5 $\mu\text{g}/\mu\text{L}$ of protein, corresponding to 30 μg , were evaporated to dryness by a centrivap. The dried protein was reconstituted in 10 μL of 15 mM TCEP in 8 M urea and 0.5% acetic acid to reduce the disulfide bonds. Reduction was carried out for 10 minutes at 50°C. The reduced protein was brought to pH \sim 7 using 0.2 M NH_4OH . 10 μL of 20 mM NAEM in 8 M urea and 0.5 M ammonium acetate was added to the reduced sample to alkylate free cysteines. Alkylation was carried out for 10 minutes in the dark. The alkylated protein was brought to pH \sim 4 with 1 μL of 25% formic acid. Digestion buffer was then used to bring the reduced and alkylated protein to the desired concentrations of 0.2 $\mu\text{g}/\mu\text{L}$, 0.05 $\mu\text{g}/\mu\text{L}$, or 0.02 $\mu\text{g}/\mu\text{L}$.

Digestions of apomyoglobin at 0.05 $\mu\text{g}/\mu\text{L}$ were used to determine if an additive would slow the rate of digestion of apomyoglobin. The additives tested were ubiquitin, β -lactoglobulin (after

reduction and alkylation), BSA (after reduction and alkylation), a mixture of the dipeptides RF and KG, penta-lysine (KKKKK), and a mixture protamines from salmon (salmine). 0.2 $\mu\text{g}/\mu\text{L}$ of apomyoglobin was diluted with digestion buffer and an additive to give apomyoglobin a final concentration of 0.05 $\mu\text{g}/\mu\text{L}$ and the additive a concentration of 0.15 $\mu\text{g}/\mu\text{L}$ for ubiquitin, β -lactoglobulin, and BSA and 0.2 $\mu\text{g}/\mu\text{L}$ for all other additives.

Replicate digestions of chicken lysozyme, β -lactoglobulin, soybean trypsin inhibitor, and bovine trypsinogen were analyzed at 0.05 $\mu\text{g}/\mu\text{L}$ and 0.02 $\mu\text{g}/\mu\text{L}$ with protamine as an additive at 0.2 $\mu\text{g}/\mu\text{L}$. These samples for digestion were created by diluting 0.2 $\mu\text{g}/\mu\text{L}$ of (reduced and alkylated) protein with digestion buffer and protamine to give the desired final concentration of protein and protamine at 0.2 $\mu\text{g}/\mu\text{L}$.

3.4.3 Protein Digestion with Immobilized Enzyme Reactor

Dried Aspergillopepsin I conjugated beads were suspended in water and pressure-loaded to create a ~ 2 mm bed in a 360 o.d. x 150 i.d. fritted piece of fused silica. The beads were washed with water for 30 minutes at 500 psi to compact the beads. The sample to be digested was flowed over the bed at the rate necessary to achieve the desired digestion time as given by Equation 3.2. Samples were collected for a calculated amount of time to give ~ 9 μL of digestion products. Digestion products were diluted with digestion buffer to 1 pmol/ μL if necessary.

3.4.4 Liquid Chromatography and Mass Spectrometry

An Agilent Technologies 1100 series binary HPLC system coupled to an in-house modified Thermo LTQ-FT-ICR hybrid was used to analyze each enzyme digestion. ~ 400 fmol of the respective digests were pressure loaded onto a 10 cm PLRP-s analytical reversed phase column. 100 fmol each of vasoactive intestinal peptide and angiotensin I were also loaded as internal standards. The column was rinsed at ~ 100 nL/min with Solvent A (0.3% formic acid in water) for 1 hour to remove all salts from the sample. Peptides were eluted using a gradient of 0-25-55-100% Solvent B (72% ACN, 18% IPA, 10% water, 0.3% formic acid) in 0-5-35-40 minutes at a flow rate of ~ 100 nL/min. Peptides were selected for fragmentation in a data-dependent manner using a

100,000 resolution FT MS¹ scan. The top 3 species were selected for fragmentation by CAD and 45 ms ETD. All MS² scans were analyzed in the linear ion trap at normal speed with a precursor target of 1e4. Ions were placed on an exclusion list for 10 seconds after 3 scans in 10 seconds.

An Agilent Technologies 1100 series binary HPLC system coupled to a Thermo Orbitrap Elite hybrid was used to screen a mixture of all reduced and alkylated proteins to be investigated prior to digestion. ~400 fmol of each protein were pressure loaded onto a 10 cm PLRP-s analytical reversed phase column in addition to the internal standards. The column was rinsed and peptides were eluted similarly to the digestion analysis. Proteins were observed using a 60,000 resolution Orbitrap MS¹ scan.

3.4.5 Data Analysis

Raw files from the enzyme digestions were split into two MGF (Mascot Generic Format) files, one for all CAD scans and the other for all ETD scans, using an in-house program. These files were individually searched using Byonic (version 3.3.11) against the protein sequence that was digested. A nonspecific search was performed with a precursor mass tolerance of 15 ppm and fragment mass tolerance of 0.35 Da. Modifications searched were methionine oxidation, serine/threonine/tyrosine glycosylation and GlcNAcylation, N-terminal pyro-glutamate from leading glutamine/glutamic acid peptide residues, and fixed cysteine NAEM alkylation.

A selection of peptides identified from each digestion were manually validated to confirm the sequence. Mass areas of these peptides were collected by filtering the original raw files for each observed charge state and taking the area under the peak.

3.5 Results and Discussion

3.5.1 Reducing Analyte Concentration Results in a Greater Number of Cleavages by Aspergillopepsin I

Initial experiments to characterize how concentration affects digestion were carried out with the standard protein equine apomyoglobin. Previous digestions of apomyoglobin were well characterized at 0.2 µg/µL (~11.8 pmol/µL) across various digestion times. Apomyoglobin was

digested at concentrations of 0.1 $\mu\text{g}/\mu\text{L}$ (~ 5.9 pmol/ μL), 0.05 $\mu\text{g}/\mu\text{L}$ (~ 2.9 pmol/ μL), and 0.02 $\mu\text{g}/\mu\text{L}$ (~ 1.2 pmol/ μL) to determine the dependence of concentration. All initial digestion times for the lower concentration samples were ~ 1 s to serve as a constant for comparison.

Shown in Figure 3.11 are the resulting chromatograms of ~ 1 s apomyoglobin digestions at decreasing concentrations. The total ion current was evenly distributed throughout the gradient elution at higher concentrations. Lower concentrations had a greater abundance of signal in the beginning of the gradient (~ 12 -20 minutes). This early region contains primarily smaller peptides (~ 0.9 -3 kDa). Additionally, the large peptides found in Figure 3.11A were eliminated once the concentration reached 0.05 $\mu\text{g}/\mu\text{L}$. Since these large peptides were being eliminated while smaller peptides were being generated, this implies that Aspergillopepsin I continually digested the large peptide products. Therefore, the decrease in apomyoglobin concentration resulted in an increase in the number of observed enzymatic cleavages by Aspergillopepsin I. This represents an immobilized enzyme reactor working in an externally diffusion limited manner.

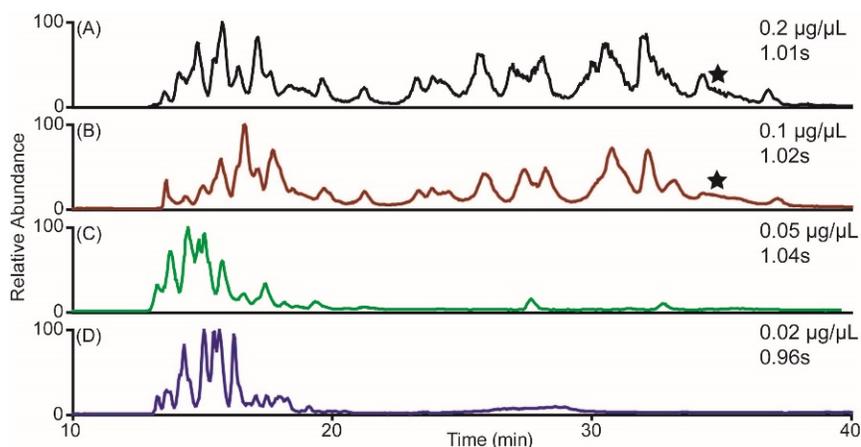
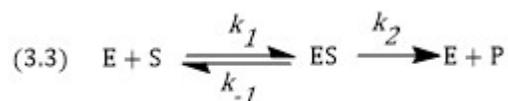


Figure 3.11 ~ 1 s Digestion Profiles of Apomyoglobin at Various Concentrations. Total ion current chromatograms for the ~ 1 s digestions of (A) 0.2 $\mu\text{g}/\mu\text{L}$, (B) 0.1 $\mu\text{g}/\mu\text{L}$, (C) 0.05 $\mu\text{g}/\mu\text{L}$, and (D) 0.02 $\mu\text{g}/\mu\text{L}$ apomyoglobin. The star in (A) and (B) represents undigested apomyoglobin.

Enzymes in solution follow the general reaction equation proposed by Michaelis and Menten:



where E is the enzyme, S is the substrate, ES is the enzyme-substrate complex, and P is product (32). Each step has its own rate constant, k . The formation of the enzyme-substrate complex is reversible while product generation is irreversible. Immobilizing an enzyme creates an additional step to this overall reaction (42). The presence of a diffusion layer as shown in Figure 3.2 requires the substrate to first traverse the diffusion layer prior to interaction with the enzyme. When the rate limiting step is the traversal through the diffusion layer, the reaction is in an externally diffusion limited system (10, 11, 42–45). This occurs with low substrate concentration, low substrate diffusivity, and high enzyme concentration on the surface of the solid support (8). When the concentration of substrate increases, the system shifts towards a catalytic limited system. The rate limiting step for a catalytic limited system is the generation of product from the enzyme-substrate complex (11, 42).

The nonspecific nature of Aspergillopepsin I is unique because the products of an enzymatic reaction are also the substrates for additional reactions. The system was closer to a catalytic limited system at concentrations between 0.1 $\mu\text{g}/\mu\text{L}$ and 0.2 $\mu\text{g}/\mu\text{L}$. Since the majority of the enzymatic sites were occupied, the products of an enzymatic reaction have sufficient time to diffuse through the diffusion layer into the bulk solution. The system shifted to an externally diffusion limited system as the concentration dropped below 0.1 $\mu\text{g}/\mu\text{L}$. The undigested apomyoglobin must have been consumed early in the traversal through enzyme reactor bed at these concentrations. The resulting products, already being low in concentration and within the diffusion layer, were then free to be digested by available enzymes further along in the bed. This process repeated itself until the final products eluted from the enzyme reactor. This would account for the loss of large peptides in the low concentration digestions.

The samples of apomyoglobin at 0.1 $\mu\text{g}/\mu\text{L}$ and 0.05 $\mu\text{g}/\mu\text{L}$ were digested at a faster flow rate (resulting in a lower digestion time) to obtain a digestion profile similar to the ~ 1 s digestion of apomyoglobin at 0.2 $\mu\text{g}/\mu\text{L}$. Increasing the flow rate decreases the time the analyte is in contact with the enzyme, which in turn decreases the amount of observed enzymatic cleavages. Shown in Figure 3.12 are the digestion times needed to obtain the similar profile. A ~ 700 ms digestion of 0.1 $\mu\text{g}/\mu\text{L}$ apomyoglobin has a strong resemblance to that of the ~ 1 s digestion at 0.2 $\mu\text{g}/\mu\text{L}$ while a

significantly shorter time of ~400 ms at 0.05 $\mu\text{g}/\mu\text{L}$ was necessary to obtain a similar profile. A digestion time for a 0.02 $\mu\text{g}/\mu\text{L}$ to match the profile of a ~1 s 0.2 $\mu\text{g}/\mu\text{L}$ digestion would be in the range of 150-200 ms based on this trend.

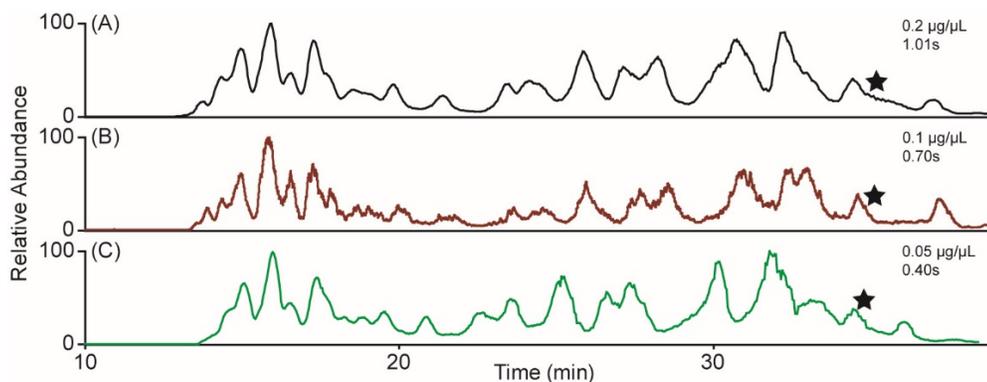


Figure 3.12. Comparing Digestion Profiles of Apomyoglobin at Various Concentrations to a Standard Digestion. Total ion current chromatograms for the apomyoglobin digestions of (A) 0.2 $\mu\text{g}/\mu\text{L}$ for ~1 s, (B) 0.1 $\mu\text{g}/\mu\text{L}$ for ~700 ms, and (C) 0.05 $\mu\text{g}/\mu\text{L}$ for ~400 ms. The star indicates undigested apomyoglobin.

The digestion times necessary for a 0.02 $\mu\text{g}/\mu\text{L}$ comparison digestion are nearly impossible to obtain using the current materials for reactor creation (~2 mm packed bed within a 150 μm I.D. fused silica capillary). Bed lengths for this digestion time would need to be less than 1 mm in length, introducing a high degree of systematic error in measurements. Conversely, flow rates needed to obtain these digestion times (~4 $\mu\text{L}/\text{min}$) are impractical. Since the flow rate was dictated by the applied pressure of the pressure vessel, pressures greater than ~2,000 psi are needed. However, pressures > 750 psi are incompatible with the current reactor construction (for example, the Kasil frit can only withstand pressures up to ~750 psi). An alternative is necessary to successfully inhibit the rate of digestion for low concentration samples to confidently reproduce the digestion and obtain large peptide fragments.

Digesting a protein in an externally diffusion limited system resulted in the generation of primarily small peptides. The lack of large peptides inhibits the potential for site localizing multiple modifications, results in the loss of quality ETD fragmentation, and hinders a more confident sequence continuity across the protein. Since not all biological samples can be ideally concentrated to 0.2 $\mu\text{g}/\mu\text{L}$, alternatives are needed to inhibit the enzyme to effectively bring the externally

diffusion limited system to a catalytic limited system. Although a lower digestion time is a viable option in certain instances, another method is needed for low concentration samples. The primary method of experimentation will be to introduce an additive that competitively inhibits the enzyme.

3.5.2 The Search for a Competitive Inhibitor of Aspergillopepsin I

Initial experiments to reduce the rate of enzymatic digestion of apomyoglobin were performed by treating a 0.05 $\mu\text{g}/\mu\text{L}$ apomyoglobin solution with a protein at three times the mass concentration (0.15 $\mu\text{g}/\mu\text{L}$). Since there are a greater number of enzymatic sites available for Aspergillopepsin I present within the competitive inhibitor, the number of cleavages to apomyoglobin specifically should decrease. Three proteins were used as additives to test this hypothesis: 1) a protein lower in molecular weight than apomyoglobin (bovine ubiquitin at ~ 8.5 kDa), 2) a protein similar in molecular weight to apomyoglobin (bovine β -lactoglobulin at ~ 18.4 kDa), and 3) a protein larger in molecular weight (bovine serum albumin [BSA] at ~ 66.4 kDa). Twelve generated apomyoglobin peptides at varying molecular weights were tracked to determine if the additive slowed the enzymatic digestion (see Table 3.1 for list and Figure 3.13 for chromatographic positions in a 0.2 $\mu\text{g}/\mu\text{L}$ and 0.05 $\mu\text{g}/\mu\text{L}$ apomyoglobin digestion). The larger peptides are eliminated in an externally diffusional limited system. Therefore, if all the peptides were observed in a treated digestion, then the additive was successful at inhibiting the rate of digestion of apomyoglobin.

Table 3.1. Table of Tracked Peptides. The peptide number corresponds to labeled peaks in Figure 3.13. The superscript number prior to the sequence corresponds to position of the peptide in the protein

<u>Pep.</u> #	<u>M+H</u> (Da)	<u>Sequence</u>
1	1292.70	⁴³ FDKFKHLKTE
2	1271.66	³² LFTGHPETLEK
3	4382.32	³² LFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVL
4	3242.66	¹⁰⁷ ISDAIIHVLHSHKHPGDFGADAQQGAMTKALEL
5	3774.90	¹⁰⁷ ISDAIIHVLHSHKHPGDFGADAQQGAMTKALELFRND
6	4051.01	¹⁰⁵ EFISDAIIHVLHSHKHPGDFGADAQQGAMTKALELFRND
7	5080.61	¹⁰⁷ ISDAIIHVLHSHKHPGDFGADAQQGAMTKALELFRNDIAAKYKELGFQG
8	5356.72	¹⁰⁵ EFISDAIIHVLHSHKHPGDFGADAQQGAMTKALELFRNDIAAKYKELGFQG
9	9193.95	⁷⁰ TALGGILLKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSHKHPGDFAGADAQQGAMTKALELFRNDIAAKYKELGFQG
10	3403.74	¹ GLSDGEWQQVLNVWGKVEADIAGHGQEVLR
11	7767.04	¹ GLSDGEWQQVLNVWGKVEADIAGHGQEVLRIFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVL
12	4656.38	¹ GLSDGEWQQVLNVWGKVEADIAGHGQEVLRIFTGHPETLEK

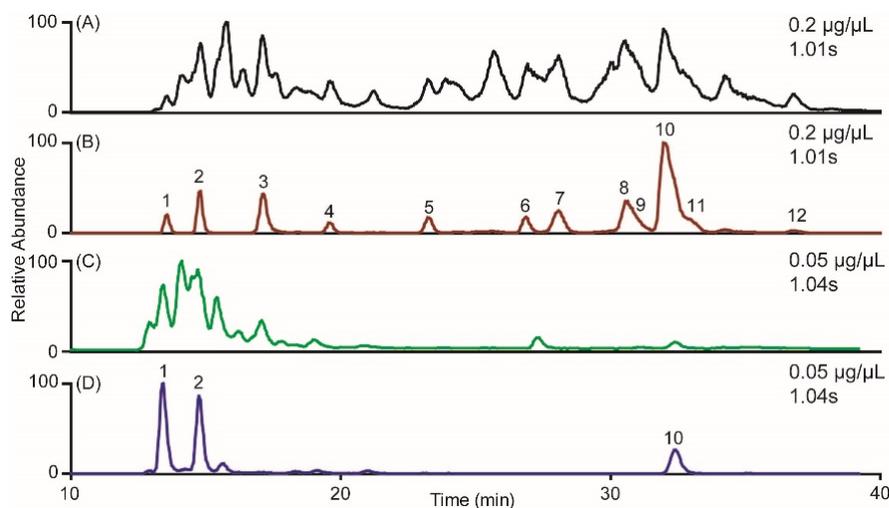


Figure 3.13. Chromatogram of Tracked Peptides. A) The total ion current of a ~ 1 s digestion of $0.2 \mu\text{g}/\mu\text{L}$ apomyoglobin. B) Extracted ion chromatogram of 12 peptides. The numbers represent the Peptide # given in Table 3.1. Since this digestion is closer to a catalytic limited system, all 12 peptides should be present when apomyoglobin is digested under these conditions. C) The total ion current of a ~ 1 s digestion of $0.05 \mu\text{g}/\mu\text{L}$ apomyoglobin. D) Extracted ion chromatogram of the same 12 peptides as B). Notice that only 3 of the 12 peptides are present in this run.

Figures 3.14 and 3.15 show the results of ~ 1 s digestion of $0.15 \mu\text{g}/\mu\text{L}$ ubiquitin + $0.05 \mu\text{g}/\mu\text{L}$ apomyoglobin (Figure 3.14) and of $0.15 \mu\text{g}/\mu\text{L}$ β -lactoglobulin + $0.05 \mu\text{g}/\mu\text{L}$ apomyoglobin (Figure 3.15). These proteins did not effectively reduce the rate of digestion for apomyoglobin. However, these experiments did reveal an important fact: proteins of similar molecular weight digest regardless of the total protein concentration. Therefore, the individual protein concentration dictates if Aspergillopepsin I will digest that specific protein at an externally diffusional or catalytic limited rate. This observation becomes relevant when digesting bispecific antibodies (to be discussed further in Chapter 4).

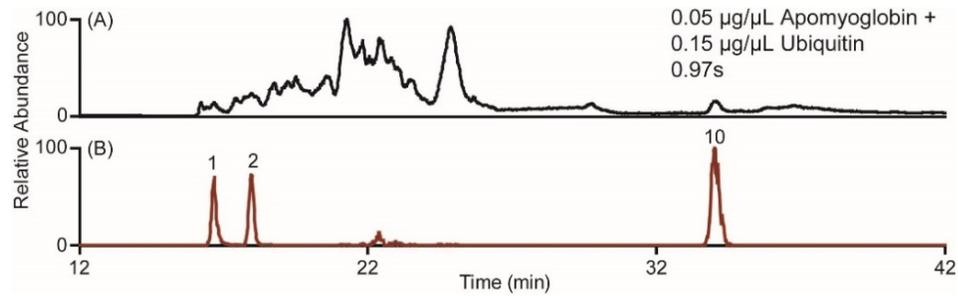


Figure 3.14. Ubiquitin Treated Apomyoglobin Digestion. A) The total ion current chromatogram of the ~1 s digestion of 0.15 µg/µL ubiquitin + 0.05 µg/µL apomyoglobin. B) Extracted ion chromatogram of the 12 tracked peptides. Only 3 of the 12 possible peptides were observed which implies that apomyoglobin is still in an externally diffusion limited system.

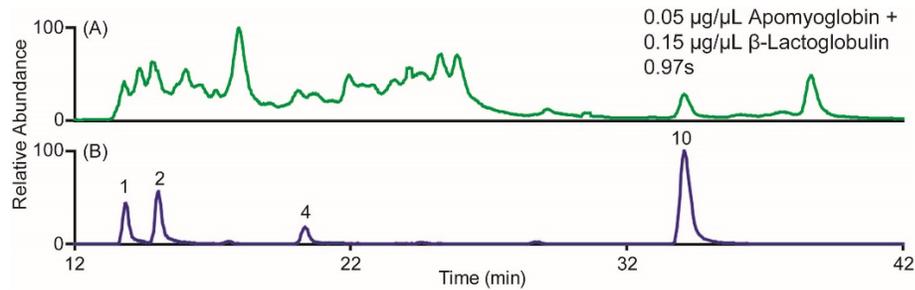


Figure 3.15. β-Lactoglobulin Treated Apomyoglobin Digestion. A) The total ion current chromatogram of the ~1s digestion of 0.15 µg/µL β-Lactoglobulin + 0.05 µg/µL apomyoglobin. B) Extracted ion chromatogram of the 12 tracked peptides. Only 4 of the 12 possible peptides were observed which implies that apomyoglobin is still in an externally diffusion limited system.

Figure 3.16 shows the results of ~1 s digestion of 0.15 µg/µL BSA + 0.05 µg/µL apomyoglobin. BSA was successful in slowing the rate of digestion of apomyoglobin as seen by the production of the large molecular weight species. This was most likely due to the much larger molecular weight of BSA than apomyoglobin which in turn affects the relative diffusion rate in the system. Generally, the rate of diffusion is inversely proportional to the molecular weight of the species (i.e. large proteins diffuse slowly) (46–48). The large molecular weight prevents BSA and its respective digestion products to diffuse quickly enough from the surface of the beads to the bulk solution, thus blocking apomyoglobin from reaching the active sites of Aspergillopepsin I.

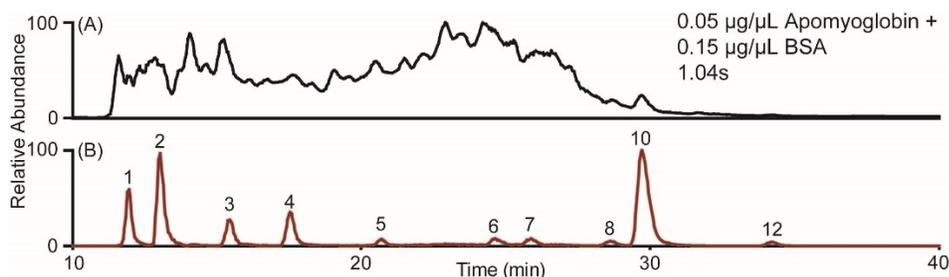


Figure 3.16. BSA Treated Apomyoglobin Digestion. A) The total ion current chromatogram of the ~1 s digestion of 0.15 µg/µL BSA + 0.05 µg/µL apomyoglobin. B) Extracted ion chromatogram of the 12 tracked peptides. 10 of the 12 possible peptides were observed which implies that apomyoglobin was successfully inhibited. However, many of the peaks were suppressed by peptides from BSA.

When protein additives were used, the resulting chromatograms were dominated by digestion products of the more abundant protein. This greatly inhibited the dynamic range and thus the ability to detect low concentration species. Three low molecular weight additives at 0.2 µg/µL were tested as additives: a mixture of the dipeptides RF (~620.8 pmol/µL) and KG (~979.7 pmol/µL), the pentapeptide KKKKK (~303.26 pmol/µL), and a mixture of protamines from salmon (~47.4 pmol/µL). These peptides all have a favored site in the P1 position based on the relative enzyme specificity (Figure 3.10). Additionally, the dipeptides and KKKKK are small and hydrophilic which prevent their retention on a reverse phase column, thereby eliminating the potential for interference with the digestion products of interest. However, RF/KG and KKKKK were unsuccessful in reducing the rate of digestion of apomyoglobin as shown in Figures 3.17 and 3.18. Since these peptides and any potential digestion products did not retain on the reverse phase column, it is impossible to determine using the methods at hand if these peptides interacted with Aspergillopepsin I in any way.

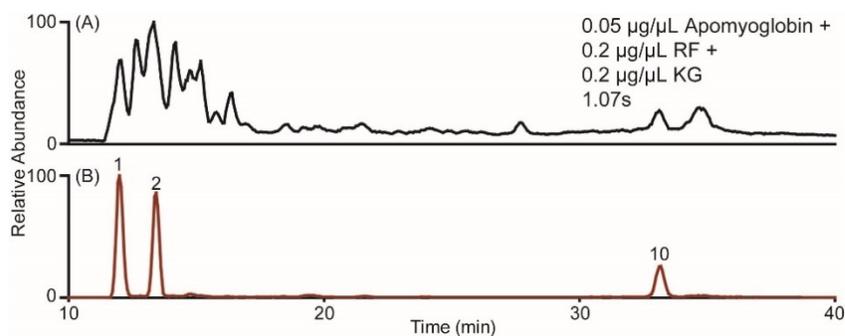


Figure 3.17. RF + KG Treated Apomyoglobin Digestion. A) The total ion current chromatogram of the ~1s digestion of 0.2 µg/µL RF + 0.2 µg/µL KG + 0.05 µg/µL apomyoglobin. B) Extracted ion chromatogram of the 12 tracked peptides. Only 3 of the 12 possible peptides were observed which implies that apomyoglobin is still in an externally diffusion limited system.

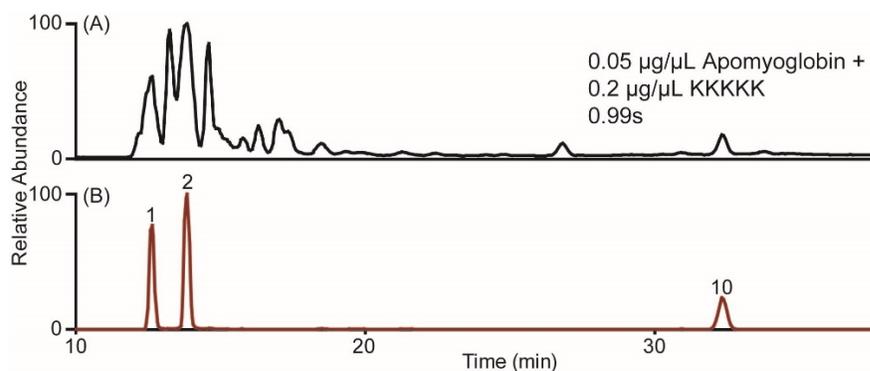


Figure 3.18. KKKKK Treated Apomyoglobin Digestion. A) The total ion current chromatogram of the ~1s digestion of 0.2 µg/µL KKKKK + 0.05 µg/µL apomyoglobin. B) Extracted ion chromatogram of the 12 tracked peptides. Only 3 of the 12 possible peptides were observed which implies that apomyoglobin is still in an externally diffusion limited system.

3.5.3 Protamine as the Ideal Inhibitor

The final additive tested was a mixture of protamine from salmon (also known as salmine). These proteins are ~4 kDa in molecular weight (Table 3.2) and can be purchased from Sigma at low cost (< \$50/g). These proteins contain mainly arginine residues, which should provide multiple preferred cleavage sites to be available for Aspergillopepsin I. Additionally, their high hydrophilicity makes these proteins the first species to elute from a reverse phased column, providing minimal interference from the digestion products of interest.

Table 3.2. Table of Protamines Observed in Salmine.

M+H (Da)	Sequence
4234.59	PRRRRRSSSRPIRRRRRPRASRRRRRGGRRRR
4317.68	PRRRRSSRRPVRRRRRPRVSRRRRRRGGRRRR
4248.61	PRRRRSSSRPVRRRRRPRVSRRRRRRGGRRRR
4062.54	PRRRRASRRIRRRRRRPRVSRRRRRRGGRRRR

Samples of 0.05 $\mu\text{g}/\mu\text{L}$ and 0.02 $\mu\text{g}/\mu\text{L}$ apomyoglobin were treated with a final concentration of 0.2 $\mu\text{g}/\mu\text{L}$ of protamine. The ~ 1 s digestions resulted in chromatograms that had very similar digestion profiles to that of a 0.2 $\mu\text{g}/\mu\text{L}$ apomyoglobin digestion for ~ 1 s (Figure 3.19). Treatment with protamine increased the abundance of high molecular weight products. Additionally, undigested apomyoglobin was detected in the 0.05 $\mu\text{g}/\mu\text{L}$ apomyoglobin treated digestion. Since protamine is very small in comparison to apomyoglobin (~ 4 kDa for protamine vs. ~ 17 kDa for apomyoglobin), protamine has the capacity to diffuse through the diffusion layer at a faster rate. A percentage of active sites were occupied with protamine and prevented apomyoglobin from being continually digested. Treatment of this externally diffusion limited system with protamine resulted in a shift to a catalytic limited system. Therefore, protamine was efficient as a competitive inhibitor.

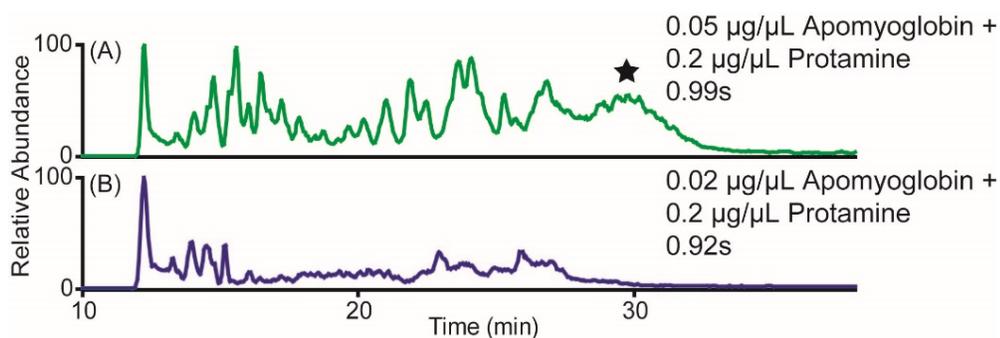


Figure 3.19. Protamine Treated, Externally Diffusion Limited Apomyoglobin Digestions. Total ion current chromatograms for (A) 0.05 $\mu\text{g}/\mu\text{L}$ and (B) 0.02 $\mu\text{g}/\mu\text{L}$. Both digestion times were ~ 1 s with protamine at a concentration of 0.2 $\mu\text{g}/\mu\text{L}$. The star in (A) denotes undigested apomyoglobin while the first major peak at ~ 12 minutes denotes peptides associated with protamine.

Samples of 0.2 $\mu\text{g}/\mu\text{L}$ and 0.1 $\mu\text{g}/\mu\text{L}$ apomyoglobin were treated with a final concentration of 0.2 $\mu\text{g}/\mu\text{L}$ of protamine to determine the effect of protamine on a system that is closer to catalytically limited. The ~ 1 s digestions resulted in chromatograms that had a very large abundance of undigested apomyoglobin (Figure 3.20). Therefore, protamine resulted in the inhibition of the digestion of apomyoglobin even at these concentrations.

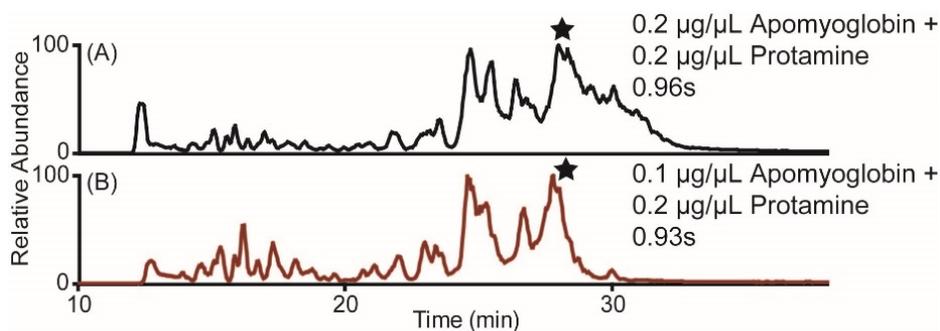


Figure 3.20. Protamine Treated, Catalytic Limited Apomyoglobin Digestions. Total ion current chromatograms for (A) 0.2 $\mu\text{g}/\mu\text{L}$ and (B) 0.10 $\mu\text{g}/\mu\text{L}$. Both digestion times were ~ 1 s with protamine at a concentration of 0.2 $\mu\text{g}/\mu\text{L}$. The star in (A) and (B) denotes undigested apomyoglobin while the first major peak at ~ 12 minutes denotes peptides associated with protamine.

The mass area of a few selected peptides of varying molecular weights were gathered from digestions of apomyoglobin at four different concentrations with and without protamine treatment (Figure 3.21). The abundance of larger peptides decreased with decreasing apomyoglobin concentration while the smaller peptide abundances increased. Treatment with protamine preserved the abundances of the larger peptides while the smaller peptides remained low.

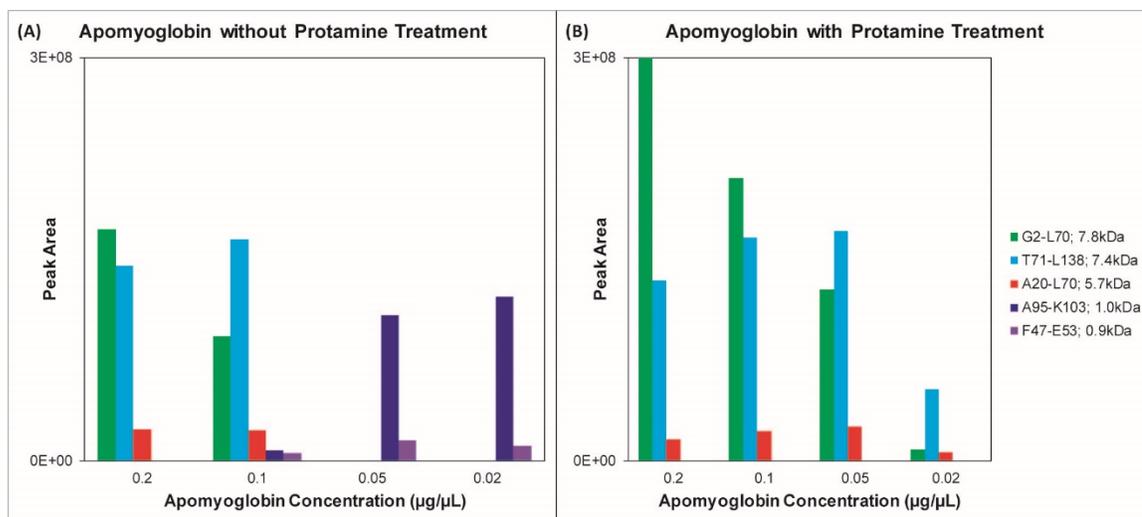


Figure 3.21. Bar Graph Representation of Selected Peptides from Apomyoglobin Digestions. (A) Apomyoglobin digestions without protamine treatment. (B) Apomyoglobin digestions with protamine treatment. All digests for the eight conditions were ~ 1 s.

Since protamine was the additive of choice, 0.2 $\mu\text{g}/\mu\text{L}$ of protamine was digested for ~ 900 ms by itself to evaluate the digestion products of protamine without interference (Figure 3.22). Notably, all species elute as a single, tailing peak. The major species within this peak were undigested protamines while the minor digestion products were observed at about the 10% relative abundance level. Additionally, there was the possibility that a percentage of digestion products were not observed due to extreme hydrophilicity, thus preventing retention on a reverse-phase column.

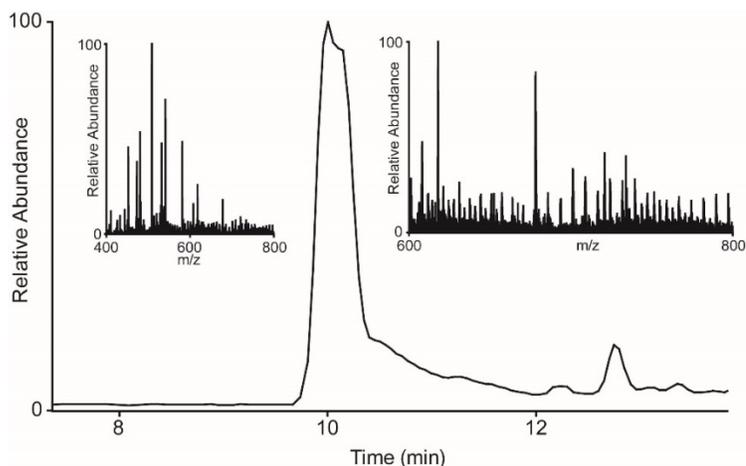


Figure 3.22. Protamine Only Digestion. Total ion current chromatogram of protamine digested for ~ 900 ms. The left inset depicts the mass range in which the protamines elute during the major peak from ~ 9.8 -11.5 minutes. The largest peaks in this window represent undigested protamine. The right inset depicts the mass range of the lower abundance protamine digestion products.

A distinct advantage in using protamine was the early elution time. Since all digestion samples must be washed to remove the 8M urea prior to MS analysis, these protamines can be easily removed during this step. Typically, column washing is carried out using 0% Solvent B. A slight increase to 4% solvent B resulted in the removal of an extensive percentage of the protamines (Figure 3.23). Although the protamines were removed, peptides that elute with less than 4% Solvent B were also removed from the column. However, these peptides were typically very low in molecular weight. Since introducing protamine preserves the large peptides generated, these low molecular weight peptides were covered by larger peptides and did not hinder the ability to accurately sequence the protein.

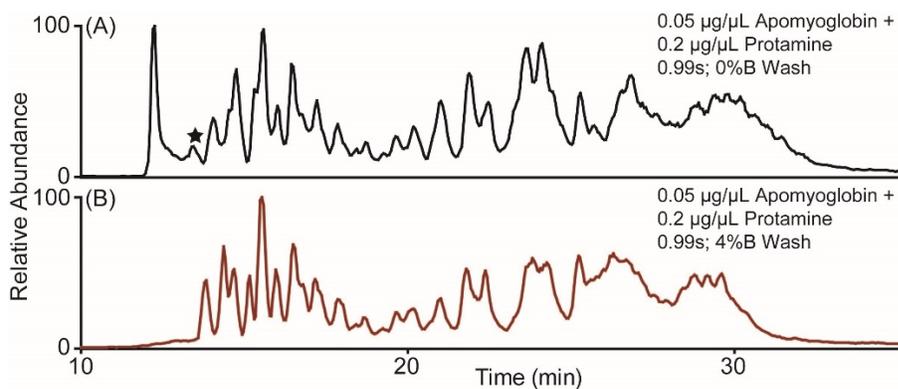


Figure 3.23. Effect of Using Organic Rinse. Total ion current chromatograms of the same ~1 s digestion of 0.05 µg/µL apomyoglobin treated with 0.2 µg/µL protamine. (A) Digestion in which the sample was rinsed for 1 hour with 0% Solvent B. (B) Digestion in which the sample was rinsed for 1 hour with 4% Solvent B. The star in (A) denotes the peptide FDKFKHLKTE. This peptide is an example of peptides that were lost when rinsing with 4% Solvent B.

3.5.4 Reproducibility of Protamine Inhibition

Protamine was successful in limiting the number of observed cleavages of apomyoglobin. Four other protein standards were tested to ensure that protamine would act similarly across multiple proteins (Figure 3.24). These proteins included chicken lysozyme, bovine β-lactoglobulin, soybean trypsin inhibitor, and bovine trypsinogen.

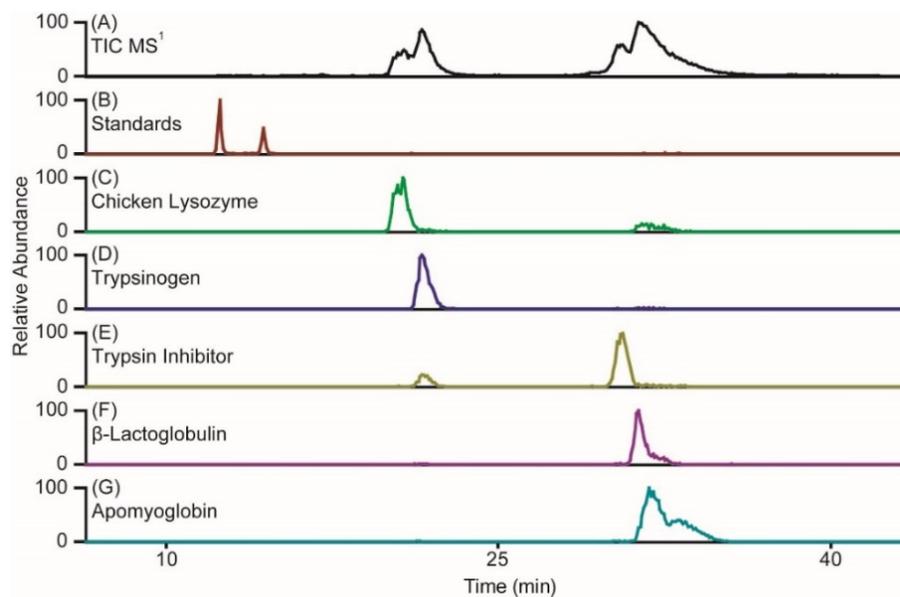


Figure 3.24. Chromatogram of All Proteins Tested for Protamine Treatment. (A) Total ion current of all Orbitrap MS¹ scans. (B) Extracted ion chromatogram of the internal standard peptides vasoactive intestinal peptide and angiotensin I used throughout all analyses. (C) Extracted ion chromatogram of chicken lysozyme. (D) Extracted ion chromatogram of bovine trypsinogen. (E) Extracted ion chromatogram of soybean trypsin inhibitor. (F) Extracted ion chromatogram of bovine β -lactoglobulin. (G) Extracted ion chromatogram of apomyoglobin.

All proteins were digested at 0.2 $\mu\text{g}/\mu\text{L}$, 0.05 $\mu\text{g}/\mu\text{L}$, and 0.02 $\mu\text{g}/\mu\text{L}$ without protamine treatment. When the protein concentration decreased, the number of observed cleavages increased (Figure 3.25). This was indicated by an increase in abundance of low molecular weight peptides, similarly to the low concentration digestion of apomyoglobin.

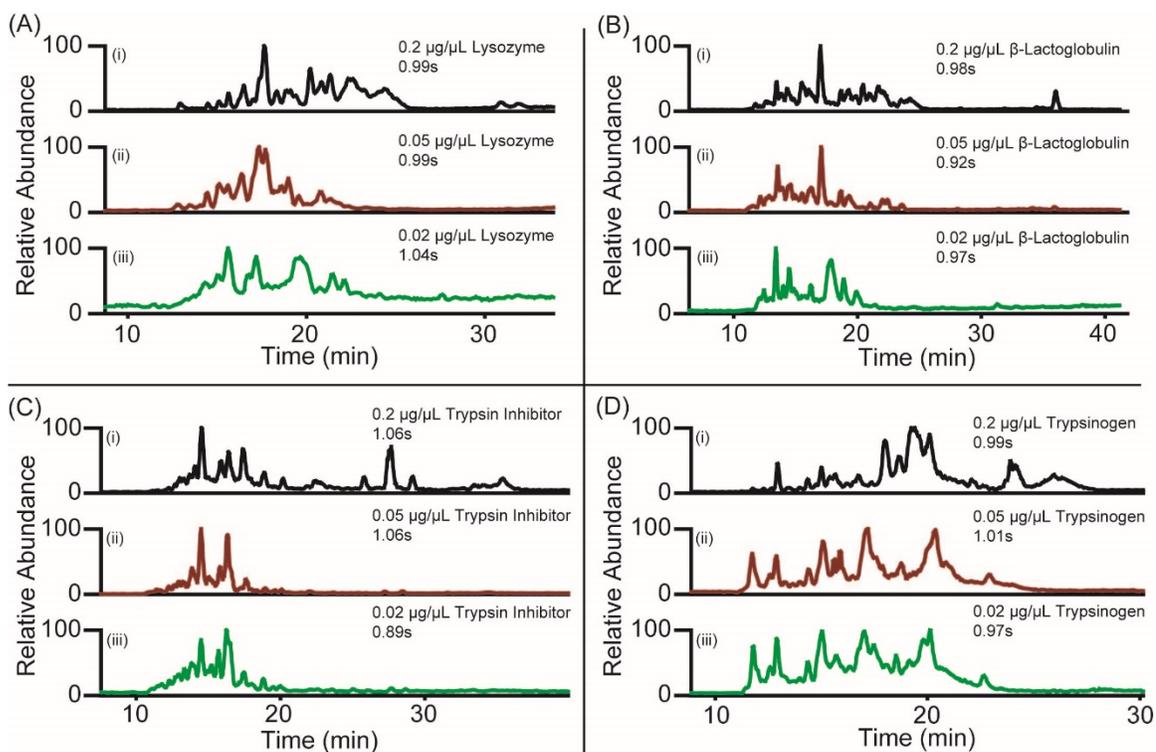


Figure 3.25. Untreated Protein Digestion of Multiple Proteins. Total ion current chromatograms of the digestions of (A) chicken lysozyme, (B) β -lactoglobulin, (C) trypsin inhibitor, and (D) trypsinogen. For each quadrant, (i) represents a $0.2 \mu\text{g}/\mu\text{L}$ digestion, (ii) represents a $0.05 \mu\text{g}/\mu\text{L}$ digestion, and (iii) represents a $0.02 \mu\text{g}/\mu\text{L}$ digestion. All digestion times were ~ 1 s.

The four proteins were digested at $0.05 \mu\text{g}/\mu\text{L}$ and $0.02 \mu\text{g}/\mu\text{L}$ with protamine treatment. Treating the proteins in an externally diffusion limited system with protamine gave similar results to the digestion of apomyoglobin with protamine. The number of observed cleavages was reduced for the same digestion time (Figure 3.26). Selected peptides were tracked across all respective digestions and produced similar trends: large molecular weight products were retained in high abundance while small molecular weight products were minimized (Figure 3.27). Overall, protamine inhibits digestion of larger peptides at low substrate protein concentrations. This enables the digestion of proteins up to an order of magnitude lower in concentration than previous working conditions.

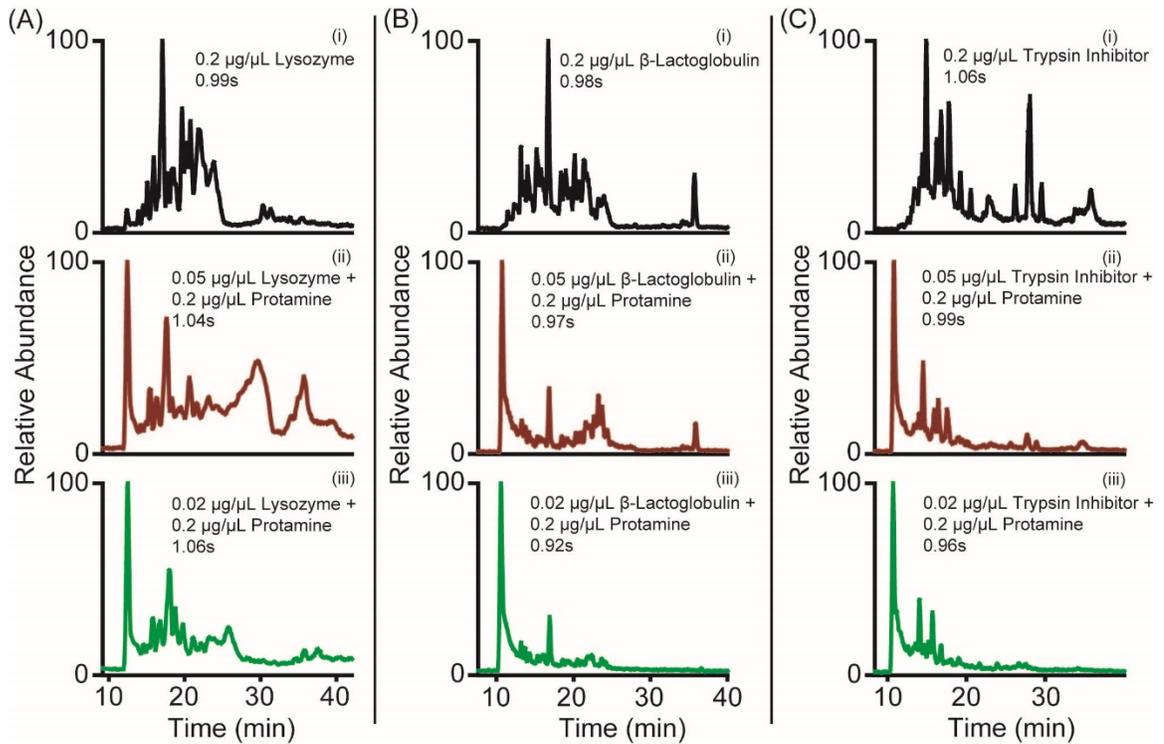


Figure 3.26. Treated Protein Digestion of Multiple Proteins. Total ion current chromatograms of the digestions of (A) chicken lysozyme, (B) β -lactoglobulin, and (C) trypsin inhibitor. For each quadrant, (i) represents an untreated 0.2 $\mu\text{g}/\mu\text{L}$ digestion for comparison, (ii) represents a protamine treated 0.05 $\mu\text{g}/\mu\text{L}$ digestion, and (iii) represents a protamine treated 0.02 $\mu\text{g}/\mu\text{L}$ digestion. All digestion times were ~ 1 s.

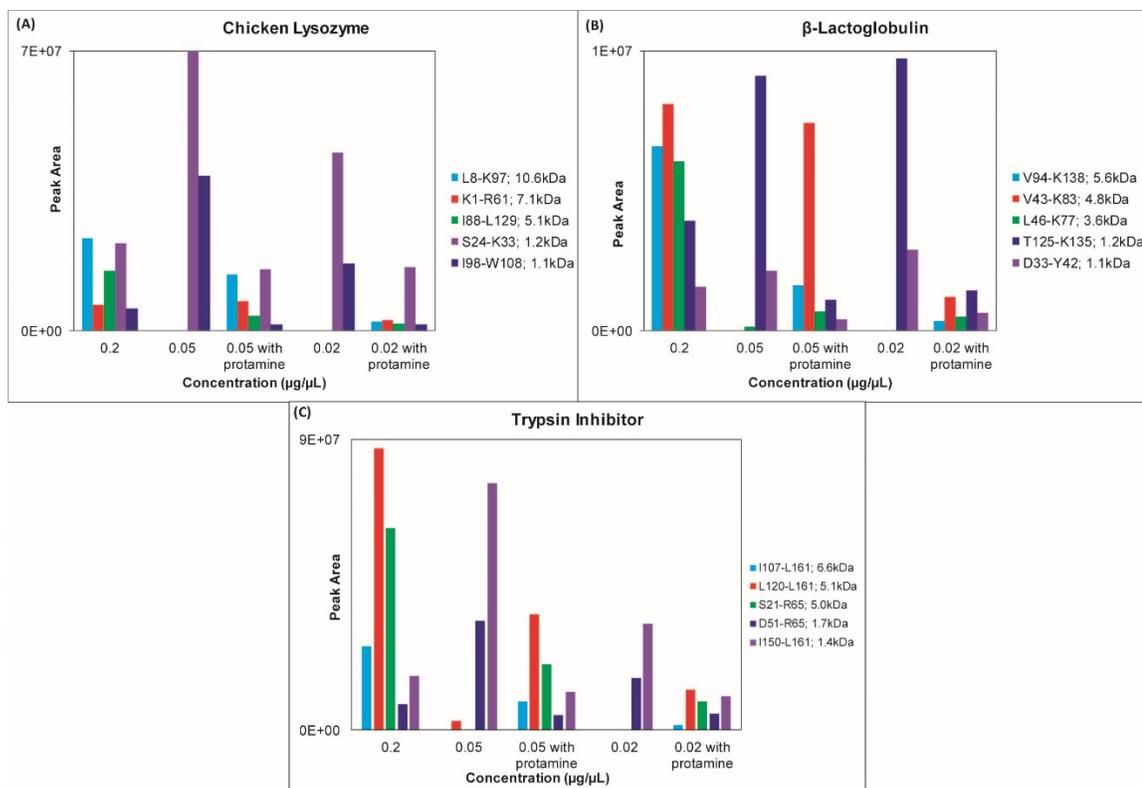


Figure 3.27. Bar Graph Representation of Selected Peptides from Protein Digestions. (A) Chicken lysozyme digestions. (B) β -lactoglobulin digestions. (C) Trypsin inhibitor digestions. All digests were ~ 1 s.

However, bovine trypsinogen, the inactive form of the protease trypsin, did not produce the expected results when treated with protamine. Figure 3.28 shows the chromatogram of a 0.05 $\mu\text{g}/\mu\text{L}$ protamine treated digestion. Notably, the only major peaks present were protamine and a large peak corresponding to undigested trypsinogen. Trypsinogen should not be in the active form of trypsin for multiple reasons: 1) the propeptide of VDDDDK was still incorporated into the sequence, 2) all disulfide bonds were reduced and alkylated to disrupt the tertiary structure and 3) the buffer for digestion was 8M urea at pH ~ 4 , both of which were incompatible with the active form of trypsin. Therefore, protamine treatment of trypsinogen specifically caused an interaction that prevented trypsinogen from being exposed to the active sites of Aspergillopepsin I. This lack of digestion was not reproduced for any protein tested other than trypsinogen. Therefore, protamine treatment has the potential to completely inhibit Aspergillopepsin I digestion for specific proteins.

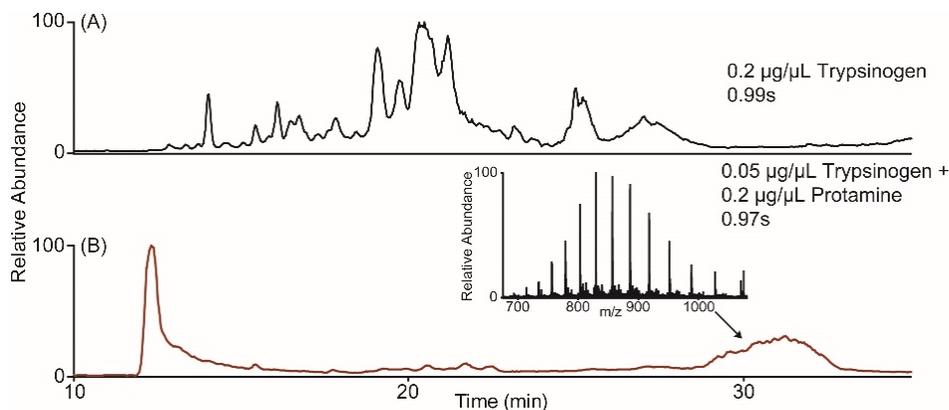


Figure 3.28. Trypsinogen Digestion when Treated with Protamine. Total ion current chromatograms of (A) untreated 0.2 $\mu\text{g}/\mu\text{L}$ digestion of trypsinogen and (B) protamine treated 0.05 $\mu\text{g}/\mu\text{L}$ digestion of trypsinogen. Both digestion times were ~ 1 s. The primary peak in (B) at about 30 minutes was undigested trypsinogen. The MS¹ inset shows the electro spray distribution of the undigested protein.

3.6 Conclusion

Demonstrated here, the initial concentration of the substrate was determined to play a major role in the degree of digestion by Aspergillopepsin I. When concentration decreased, the abundance of low molecular weight species increased while large molecular weight species were not observed. Treating these low concentration samples with protamine mitigated this effect through competitive inhibition of Aspergillopepsin I. Among four standard proteins, three showed a similar reduction in observed cleavages. However, trypsinogen did not show the same result. This particular protein seems to be an outlier to the overall trend. The ability to reduce the number of cleavages through protamine treatment enabled protein samples to be digested at concentrations that are an order of magnitude lower than previous conditions allowed. This provides a potentially useful tool for the digestion of biological samples in which the protein(s) are low in abundance to retain large products. Bispecific antibodies are a perfect test case to demonstrate the utility of the results presented here and will be described in the following chapter.

3.7 References

1. Massolini G, Calleri E (2005) Immobilized trypsin systems coupled on-line to separation methods: Recent developments and analytical applications. *J Sep Sci* 28(1):7–21.
2. Mohamad NR, Marzuki NHC, Buang NA, Huyop F, Wahab RA (2015) An overview of technologies for immobilization of enzymes and surface analysis techniques for

- immobilized enzymes. *Biotechnol Biotechnol Equip* 29(2):205–220.
3. Zhang L, et al. (2016) Analysis of monoclonal antibody sequence and post-translational modifications by time-controlled proteolysis and tandem mass spectrometry. *Mol Cell Proteomics* 15(4):1479–1488.
 4. Li Y, Xu X, Deng C, Yang P, Zhang X (2007) Immobilization of Trypsin on Superparamagnetic Nanoparticles for Rapid and Effective Proteolysis. *J Proteome Res* 6(9):3849–3855.
 5. Pang Y, Wang W-H, Reid GE, Hunt DF, Bruening ML (2015) Pepsin-Containing Membranes for Controlled Monoclonal Antibody Digestion Prior to Mass Spectrometry Analysis. *Anal Chem* 87(21):10942–10949.
 6. Jiang S, Zhang Z, Li L (2015) A one-step preparation method of monolithic enzyme reactor for highly efficient sample preparation coupled to mass spectrometry-based proteomics studies. *J Chromatogr A* 1412:75–81.
 7. Hoarau M, Badiéyan S, Marsh ENG (2017) Immobilized enzymes: understanding enzyme – surface interactions at the molecular level. *Org Biomol Chem* 15(45):9539–9551.
 8. Chaplin MF, Bucke C (1990) *Enzyme technology* (Cambridge University Press).
 9. Wang C, et al. (2000) Integration of immobilized trypsin bead beds for protein digestion within a microfluidic chip incorporating capillary electrophoresis separations and an electrospray mass spectrometry interface. *Rapid Commun Mass Spectrom* 14(15):1377–1383.
 10. Laidler KJ, Bunting PS (1980) [9] The kinetics of immobilized enzyme systems. *Enzyme Kinetics and Mechanism - Part B: Isotopic Probes and Complex Enzyme Systems*, ed Purich DLBT-M in E (Academic Press), pp 227–248.
 11. Shuler ML, Kargi F, DeLisa M (2017) *Bioprocess Engineering: Basic Concepts* (Pearson Education) Available at: <http://www.informit.com/articles/article.aspx?p=2783636&seqNum=3>.
 12. Freije JR, et al. (2005) Chemically Modified, Immobilized Trypsin Reactor with Improved Digestion Efficiency. *J Proteome Res* 4(5):1805–1813.
 13. Cingöz A, Hugon-Chapuis F, Pichon V (2008) Evaluation of various immobilized enzymatic microreactors coupled on-line with liquid chromatography and mass spectrometry detection for quantitative analysis of cytochrome c. *J Chromatogr A* 1209(1–2):95–103.
 14. Toth CA, Kuklennyik Z, Barr JR (2019) Nuts and Bolts of Protein Quantification by Online Trypsin Digestion Coupled LC-MS/MS Analysis BT - Functional Proteomics: Methods and Protocols. eds Wang X, Kuruc M (Springer New York, New York, NY), pp 295–311.
 15. Zhang S, et al. (2015) Integrated platform with a combination of online digestion and 18O labeling for proteome quantification via an immobilized trypsin microreactor. *Analyst* 140(15):5227–5234.
 16. Laskay ÜA, Lobas AA, Srzentić K, Gorshkov M V, Tsybin YO (2013) Proteome Digestion Specificity Analysis for Rational Design of Extended Bottom-up and Middle-down Proteomics Experiments. *J Proteome Res* 12(12):5558–5569.
 17. Tsiatsiani L, Heck AJR (2015) Proteomics beyond trypsin. *FEBS J* 282(14):2612–2626.
 18. Papagianni M (2007) Advances in citric acid fermentation by *Aspergillus niger*: Biochemical aspects, membrane transport and modeling. *Biotechnol Adv* 25(3):244–263.
 19. Show PL, et al. (2015) Overview of citric acid production from *Aspergillus niger*. *Front Life Sci* 8(3):271–283.
 20. Takamine J (1914) Enzymes of *Aspergillus Oryzae* and the Application of Its Amyloclastic

- Enzyme to the Fermentation Industry. *Ind Eng Chem* 6(10):824–828.
21. Machida M, et al. (2005) Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* 438(7071):1157–1161.
 22. Mamo J, Assefa F (2018) The role of microbial aspartic protease enzyme in food and beverage industries. *J Food Qual* 2018.
 23. Segal BH (2009) Aspergillosis. *N Engl J Med* 360(18):1870–1884.
 24. Kurup VP, Shen H-D, Banerjee B (2000) Respiratory fungal allergy. *Microbes Infect* 2(9):1101–1110.
 25. ICHISHIMA E, YOSHIDA F (1965) Molecular Weight of Acid Proteinase of *Aspergillus saitoi*. *Nature* 207(4996):525–526.
 26. Ichishima E (2013) Aspergillopepsin I. *Handb Proteolytic Enzym*:135–141.
 27. Yoshida F, Nagasawa M (1956) Studies on the Proteolytic Enzymes of Black Aspergilli: Part I. Investigation of Strains Producing Superior Proteinase Yields in Black Aspergillus and the Crystallization of Proteolytic Enzyme from *Aspergillus Saitoi* Part II. Several Properties of the Cry. *J Agric Chem Soc Japan* 20(sup1):252–266.
 28. Shintani T, Ichishima E (1994) Primary structure of aspergillopepsin I deduced from nucleotide sequence of the gene and aspartic acid-76 is an essential active site of the enzyme for trypsinogen activation. *Biochim Biophys Acta - Protein Struct Mol Enzymol* 1204(2):257–264.
 29. Shitani T, Kobayashi M, Ichishima E (1996) Characterization of the S1 subsite specificity of aspergillopepsin I by site-directed mutagenesis. *J Biochem* 120(5):974–981.
 30. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242.
 31. Cho SW, Kim N June, Choi M-U, Shin W (2001) Structure of aspergillopepsin I from *Aspergillus phoenicis*: variations of the S1'-S2 subsite in aspartic proteinases. *Acta Crystallogr Sect D Biol Crystallogr* 57(7):948–956.
 32. Garrett RH, Grisham CM (2017) *Biochemistry* (Cengage Learning, Boston). Sixth Edit.
 33. Rombouts I, et al. (2015) Formation and reshuffling of disulfide bonds in bovine serum albumin demonstrated using tandem mass spectrometry with collision-induced and electron-transfer dissociation. *Sci Rep* 5:12210.
 34. Liu S, Moulton KR, Auclair JR, Zhou ZS (2016) Mildly acidic conditions eliminate deamidation artifact during proteolysis: digestion with endoprotease Glu-C at pH 4.5. *Amino Acids* 48(4):1059–1067.
 35. Bennion BJ, Daggett V (2003) The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci U S A* 100(9):5142–5147.
 36. Rocco AG, et al. (2008) Characterization of the Protein Unfolding Processes Induced by Urea and Temperature. *Biophys J* 94(6):2241–2251.
 37. Zhang H-M, et al. (2008) Enhanced Digestion Efficiency, Peptide Ionization Efficiency, and Sequence Resolution for Protein Hydrogen/Deuterium Exchange Monitored by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal Chem* 80(23):9034–9041.
 38. Blanco RM, Calvete JJ, Guisán J (1989) Immobilization-stabilization of enzymes; variables that control the intensity of the trypsin (amine)-agarose (aldehyde) multipoint attachment. *Enzyme Microb Technol* 11(6):353–359.
 39. Hinkle J, et al. (2019) Unambiguous Sequence Characterization of a Monoclonal Antibody in a Single Analysis Using a Nonspecific Immobilized Enzyme Reactor. *Anal Chem* 91(21):13547–13554.
 40. Daneyko A, Höltzel A, Khirevich S, Tallarek U (2011) Influence of the Particle Size Distribution on Hydraulic Permeability and Eddy Dispersion in Bulk Packings. *Anal Chem*

- 83(10):3903–3910.
41. Senko MW, et al. (2013) Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal Chem* 85(24):11710–11714.
 42. Roig MG, Estevez FB, Velasco FG, Cachaza JM (1987) Evaluation of the main factors determining the mode of action of immobilized enzymes. *Biochem Educ* 15(1):33–41.
 43. Lee GK, Lesch RA, Reilly PJ (1981) Estimation of intrinsic kinetic constants for pore diffusion-limited immobilized enzyme reactions. *Biotechnol Bioeng* 23(3):487–497.
 44. Berendsen WR, Lapin A, Reuss M (2006) Investigations of Reaction Kinetics for Immobilized Enzymes—Identification of Parameters in the Presence of Diffusion Limitation. *Biotechnol Prog* 22(5):1305–1312.
 45. Rodrigues RC, Ortiz C, Berenguer-Murcia Á, Torres R, Fernández-Lafuente R (2013) Modifying enzyme activity and selectivity by immobilization. *Chem Soc Rev* 42(15):6290–6307.
 46. Young ME, Carroad PA, Bell RL (1980) Estimation of diffusion coefficients of proteins. *Biotechnol Bioeng* 22(5):947–955.
 47. Bódalo A, et al. (1986) Analysis of diffusion effects on immobilized enzymes on porous supports with reversible Michaelis-Menten kinetics. *Enzyme Microb Technol* 8(7):433–438.
 48. He L, Niemeyer B (2003) A Novel Correlation for Protein Diffusion Coefficients Based on Molecular Weight and Radius of Gyration. *Biotechnol Prog* 19(2):544–548.

Analysis of Bispecific Antibodies Using an Immobilized Aspergillopepsin I Enzyme Reactor

4.1 Abstract

Bispecific antibodies (bsAbs) have been the target of multiple research studies for therapeutic development. These engineered antibodies interact with two different epitopes due to their sequence different variable regions. The large molecular weight and complex structures of antibodies require stringent analyses to confirm not only the primary structure but the higher order structures as well. Presented here is the analysis of two classifications of bsAbs using an immobilized Aspergillopepsin I enzyme reactor: the IgG-like, common light chain bsAb emicizumab and the bispecific T-cell engager (BiTE) EpCAM-CD3. The primary structure was reconstructed from a series of overlapping peptides for each bsAb. The digestion of emicizumab was treated with protamine to preserve large peptide fragments for analysis by parallel ion parking to deepen the analysis of the complimentary determining regions (CDRs). The digestion of EpCAM-CD3 was performed to characterize and localize all disulfide bonds in the molecule while still obtaining full sequence coverage.

4.2 Introduction

4.2.1 Antibodies

Antibodies within the body, also known as immunoglobulins (Ig), are produced by B cells in the adaptive immune system (1, 2). The most common antibody found in the blood, IgG, consists of two identical heavy chains (~50 kDa each) and two identical light chains (~25 kDa each). Four interchain and twelve intrachain disulfide bonds covalently bind the chains together into a “Y”-like structure with an overall molecular weight of ~150 kDa (Figure 4.1) (3, 4). The primary function of these molecules is to bind to their target epitope with a high affinity to produce an immune response.

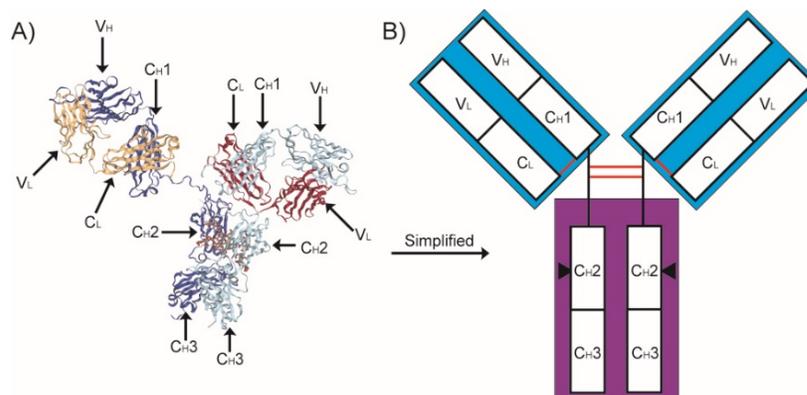


Figure 4.1. Structure of IgG. A) Crystal structure of the 4 disulfide bound chains. The heavy chains are drawn in light and dark blue. The light chains are drawn in yellow and red. Individual subunits are labeled and the glycan tree can be found in the C_{H2} domain. Each individual domain contains an intrachain disulfide bond. B) Simplified depiction of the structure. The 4 interchain disulfide bonds are noted in red and the glycan tree is noted by the triangles. The blue shaded areas correspond to the Fab regions while the purple shaded area corresponds to the Fc region. The section around the two inter-heavy chain disulfide bonds is the hinge region. Crystal structure provided by PDB (ID: 1HZH) (5). Solving of structure by reference (6).

The region of the antibody that targets and binds a specific epitope is the fragment antigen binding (Fab) region (blue areas in Figure 4.1). This region encompasses the variable domain and first constant domain of the light and heavy chains found at the N-terminal portion of the chains. Each antibody has two Fab regions. The variable domains contain three complimentary determining regions (CDRs) in the light and heavy chains (4). It is these CDRs that dictate the affinity to an epitope because they physically bind the epitope. Therefore, any structural changes in the CDRs, such as point mutations or PTMs, can drastically affect antibody binding (7–9).

The remaining portion of the antibody, consisting of the second and third constant domains of the heavy chains, is known as the fragment crystallizable region (Fc) (purple area in Figure 4.1). The function of the Fc is to interact with Fc receptor proteins on immune cells to trigger the appropriate immune response (10, 11). Additionally, Fc interaction with the neonatal Fc receptor (FcRn) plays an important role in determining antibody half-life in the body (12–14). These particular receptors are found within the acidic lysosome of endothelial cells. The FcRn binds the Fc region to shuttle the antibody back into the blood, effectively recycling the antibody (13).

The Fc region contains an N-linked glycan tree found on an asparagine residue (typically N297 in IgG) on each heavy chain (4). The trees are branched glycosylation structures that consist of at least 5 N-acetylglucosamines (GlcNAc) and 3 mannoses. Two additional galactoses, two sialic acids, and a single fucose can also be added to these trees to give larger structures (Figure 4.2) (3, 11, 15). The glycosylation pattern is essential to the function of the Fc region (10, 11). Different variations of glycosylations directly impact the affinity of the antibody to a given receptor.

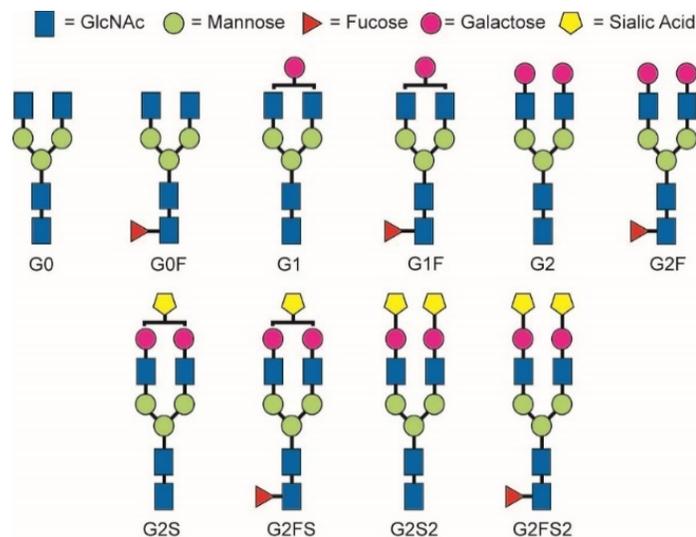


Figure 4.2. Structures of Common Antibody Glycosylation Trees. The G0 tree represents the core glycosylation tree. All other variations build from this core.

The ability to specifically target an epitope with the Fab region while triggering an immune response with the Fc region is an attractive system for developing therapeutics. Since the first approved antibody therapeutic in 1986 (Muromonab), over 80 different antibodies have been approved for consumer use (16). This new avenue for drug development presents the need to quickly and accurately determine the primary structure of the antibody being produced. MS has become the primary instrument to meet this demand. Intact analysis cannot give unambiguous sequence analysis due to the ~150 kDa molecular weight of the molecule (17, 18). Even with specialized proteases like IdeS and GingisKhan, specialized enzymes that cleave the antibody around the hinge region, subunit analysis still falls short in determining the entire sequence (19–22). Typically, multiple specific protease digestions are carried out to provide overlapping peptide

coverage to confirm the primary structure (23–25). However, this method requires multiple LC-MS analyses for sequence reconstruction, making it a time-intensive process.

4.2.2 Bispecific Antibodies

Although monospecific antibodies are produced naturally in the body, bispecific antibodies (bsAbs) are a highly engineered class of molecules (26). These molecules differ from natural antibodies in that they target a different epitope with each Fab region. This enables the ability to target two different epitopes with the same molecule. However, the affinity for a given epitope is lower than that of a mAb since each Fab has different targets.

Since the introduction of the first bsAb Removab in 2009, research into bsAbs has been on the rise (27–30). Although most bsAbs retain the IgG structure of mAbs, there are also different structural forms of this class of molecule (Figure 4.3) (31). Noticeably, the IgG structure is not utilized in all the different possible forms of bsAbs. The commonality between all the different structures is the ability to target two different epitopes with two Fab regions. Two different bsAbs will be investigated in this chapter: a knob-in-hole IgG with a common light chain bsAb and a bispecific T-cell engager (BiTE).

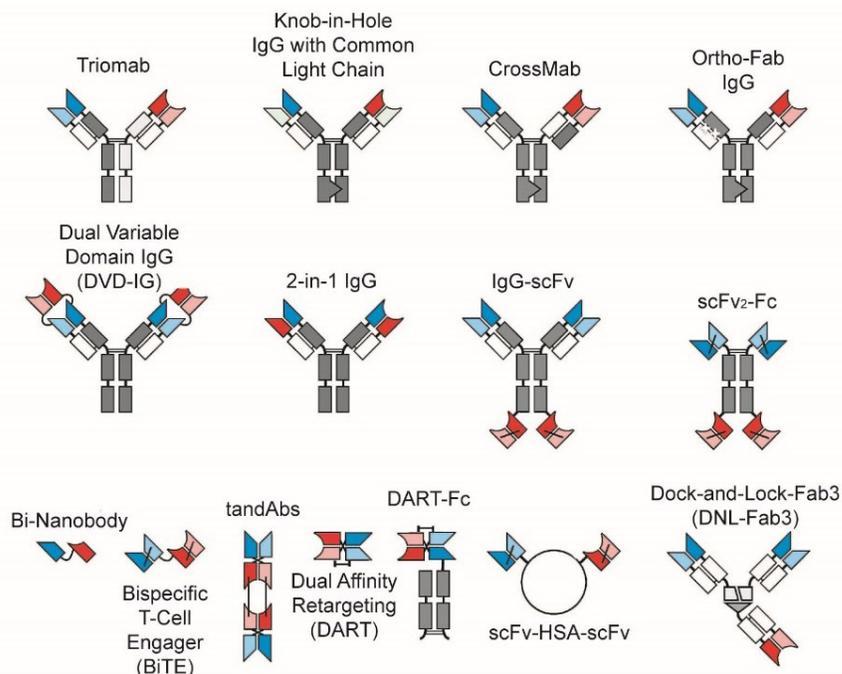


Figure 4.3. Structures of Bispecific Antibodies. Although not all structures maintain the IgG conformation, all bsAbs have the ability to target two different epitopes through at least two variable regions. scFv = single chain fragment variable. HSA = Human Serum Albumin. Figure obtained and modified from reference (31).

4.2.2.1 IgG-Like Bispecific Antibodies

IgG-like bsAbs maintain the ~150 kDa structure of natural IgG molecules: two light chains and two heavy chains bound by 4 interchain and 12 intrachain disulfide bonds. However, bsAbs complicate the structure by pairing two different light chains to two different heavy chains. This presents the problem of introducing 16 different potential confirmations of chain pairing, of which the desired outcome only represents a 12.5% probability of pairing (32).

Point mutations in the third constant region of the Fc region, known as knob-in-hole, are commonly introduced to facilitate correct heavy chain pairing (33). These mutations occur where the two C_H3 regions of the antibody interact. Large, bulky amino acids are introduced to form a protruding knob on one chain while the other chain contains small amino acids to produce a hole. The mutations lead to preferentially forming the knob-hole pairs while minimizing the knob-knob and hole-hole heavy chain pairs. Additionally, one chain could contain negatively charged residues while the other contains positively charged residues to form an electrostatic pairing mechanism

(26). Overall, the heavy chain pairing through the knob-in-hole method is effective at minimizing the number of incorrectly paired antibodies (33).

Light chain mispairing is not as simple to overcome. Point mutations in the constant regions of the light and heavy chains are not enough to produce the desired outcome (34). The most straightforward method to ensure correct light chain pairings entails the production of each half of the antibody in separate cells and then pairing the heavy/light combination halves *in vitro*. A common light chain can also be used to eliminate mispairing altogether, but it can minimize the epitope affinity or the ability to target two distinct epitopes. A novel way to solve the pairing problem is to swap domains between one set of heavy and light chains (known as CrossMab and developed by Roche), the most effective being swapping C_{H1} and C_L (Figure 4.4) (34, 35). Since C_L regions do not pair, the second, correctly formed light chain is not paired with the modified heavy chain and thus mispairing is minimized. Another method to prevent mispairing involves the substitution of C_{H1} and C_L with the C_{H1} and C_L regions of a different antibody type, such as an IgE (36). The C_L region from IgG will not pair with the C_L region from IgE, thus reducing the mispaired light chains.

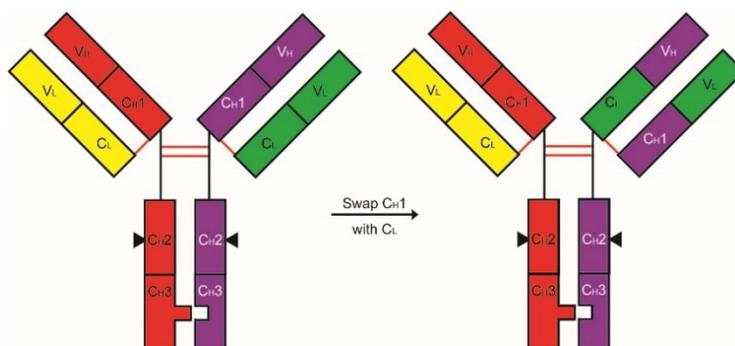


Figure 4.4. Depiction of CrossMab. An IgG-like structure is shown on the left with a knob-in-hole mutation in the C_{H3} region of the heavy chains. By swapping C_{H1} with C_L from one heavy/light chain pair, this preferentially forms the correct pairing. Figure was based on a figure from reference (35).

This chapter details the analysis of the biosimilar of emicizumab, an IgG-like bsAb. Originally produced by Genentech, this drug was FDA approved in 2017 to treat patients with hemophilia A, a genetic condition that results in either defective or absent coagulation factor VIII (37–41). Coagulation factor VIII is an essential protein to clot blood. This bsAb acts similarly to factor VIII by bringing activated factor IX and factor X, two coagulation factors, in close proximity to one another

to continue the blood clotting process. Structurally, the heavy chains are paired using an electrostatic knob-in-hole while utilizing a common light chain.

4.2.2.2 Bispecific T-Cell Engager (BiTE)

Certain bsAbs structures diverge from the structure of an IgG as shown in Figure 4.3. There are a group of bsAbs that eliminate the Fc region of the molecule altogether. One of these types of bsAbs is known as Bispecific T-cell engagers, or BiTEs (Figure 4.5) (42). This molecule removes all chain pairing problems by expressing the molecule as a single peptide chain. Strings of glycines and serines are used to connect the C-terminus of the V_L regions to the N-terminus V_H regions and the C-terminus of the V_H region to the N-terminus of the V_L region. This results in a protein that is ~50-55 kDa in molecular weight.

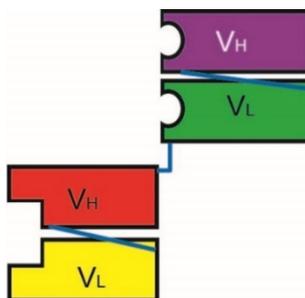


Figure 4.5. Structure of a BiTE. These bsAbs do not have an Fc region. The two Fab regions are held together by glycine/serine stretches, noted by the blue lines.

Functionally, BiTEs are used to bring a T-cell and a cancerous cell in close proximity to trigger an immune response (42). Typically, a T-cell receptor recognizes a peptide presented by an MHC molecule as non-self, triggering the immune system to respond. However, cancerous cells may hinder the presentation of the MHC molecules (43). The BiTE can mimic the MHC molecule and trigger an immune response by targeting T-cell receptors, the most common being CD3, and a protein in high abundance on cancerous cells. Blinatumomab was the first BiTE approved for use in 2014 under the accelerated approval program for acute lymphoblastic leukemia by targeting CD3 on T-cells and CD19, a receptor on B cells responsible for triggering responses to antigens (44, 45).

Although BiTEs are engineered to produce the correct therapeutic molecule, they have a very short half-life. The half-life is roughly 1.25 hours, or ~0.25% of the half-life of a mAb, due to the lack of the Fc region (42). Since there is no Fc region, FcRn cannot recognize and prevent the antibody from being eliminated in the bloodstream. The short half-life requires a continuous IV of the BiTE over the 4-8 weeks of treatment. However, the concentration of BiTE required to produce the desired response is fairly low (46). This eases the production strain of the BiTE since high concentrations are not needed.

This chapter describes the analysis and disulfide bond localization of a BiTE. This particular BiTE targets CD3 and epithelial cell adhesion molecule (EpCAM). EpCAM is primarily located in intercellular spaces of normal cells where tight junctions are formed by epithelial cells. However, EpCAM is homogenous on the surface of cancerous cells. This difference in locality prevents the BiTE from targeting normal cells since they are mainly covered by other cells (47).

4.2.3 Previous Work with Immobilized Aspergillopepsin I for Antibody Analysis

An immobilized Aspergillopepsin I enzyme reactor was implemented to obtain overlapping peptides and unambiguous sequence analysis of a mAb in a single chromatographic analysis. Two generations of enzyme reactors have been developed and applied to mAb analysis. The first generation by Zhang et al. utilized 20 μm porous beads to digest a murine IgG for ~5.7 s (48). 39 peptides were targeted for fragmentation to give 98% sequence coverage of the light chain and 94% sequence coverage of the heavy chain. However, instrumentation constraints inhibited the ability to obtain high quality data for all 39 peptides in a single analysis. Three chromatographic runs were necessary to obtain the necessary data (48). The first analysis determined the molecular weight of all species in the digestion, the second targeted and dissociated the 39 peptides using ETD, and the third targeted the same 39 peptides for dissociation by CAD (48).

The second generation by Hinkle et al. utilized 1 μm solid sphere beads to digest the biosimilar adalimumab in ~0.9 s (49). The introduction of the Thermo Orbitrap Fusion enabled the ability to create a more detailed data-dependent method which greatly benefitted identification and dissociation of the majority of peptide species present in the digestion (50). Only ~5.5% of the

identified peptides were manually inspected to give 100% sequence coverage of both the light and heavy chains. Importantly, all data was collected in a single chromatographic analysis, vastly reducing the instrument time needed to obtain comprehensive coverage.

While improvements in instrumentation and the enzyme reactor resulted in minimizing the time needed to obtain overlapping coverage, each generation presented a unique technique for antibody analysis.

4.2.3.1 Parallel Ion Parking

Hinkle et al. applied parallel ion parking during ETD (pipETD) to extensively characterize two peptides generated by the nonspecific digestion (51). Parallel ion parking, introduced by McLuckey et al. in 2002, is a technique to slow the rate of the ion/ion reaction of specific ions of selected m/z values (52, 53). The rate limiting step of an ion/ion reaction is the formation of a collision complex, resulting in a rate constant for ion/ion reactions as defined by:

$$(4.1) k \propto v \left(\frac{Z_1 Z_2 e^2}{4\pi\epsilon_0 \mu v^2} \right)^2$$

where k is the rate constant, v represents relative velocity of the reactants, Z_1 represents the absolute value of the number of charges from the precursor ion, Z_2 represents the absolute value of the number of charges from the reagent ion, e represents the charge of an electron, ϵ_0 represents the vacuum permittivity constant and μ represents the reduced mass of the two species involved (54). Large peptides generally have higher charge states after ionization, which is ideal for ETD. Therefore, a large Z_1 is expected, thus leading to a fast reaction time. Simplifying Equation 4.1 in terms of Z_1 to the following:

$$(4.2) k \propto Z_1^2$$

shows the rate of reaction is proportional to the square of the charge state of the precursor ions.

Since ETD randomly cleaves along the peptide backbone, the peptide has the potential to asymmetrically cleave. If the charges are distributed evenly along the peptide, then small fragments will have fewer charges and larger fragments have more charges. However, based on Equation 4.2, the highly charged fragment ion can react at roughly the same rate as the precursor. For example, a +35 precursor reacts ~300 times faster than a +2 fragment ion. However, the +35

precursor only reacts ~1.06 times faster than a +34 fragment ion, implying that the +34 will be consumed at roughly the same rate as the precursor.

The ideal scenario to obtain the highest S/N possible for the generated fragment ions involves completely reacting away the precursor using an extended ETD reaction time. However, the extended ETD reaction time results in secondary dissociation of large, highly charged products since they react at approximately the same rate as the precursor. This leads to the generation of internal fragments (Figure 4.6). These fragment ions do not contain either termini and thus are not sequence informative. Parallel ion parking during ETD allows the precursor to be reacted away while minimizing the potential for the first generation fragment ions to undergo additional ETD reactions.

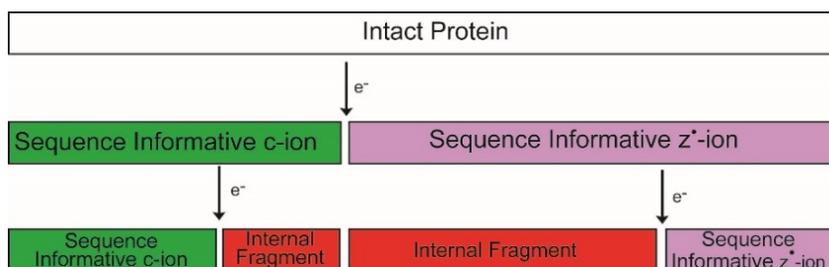


Figure 4.6. Extended ETD Reaction Times Result in the Generation of Internal Fragments.

The rate of reaction for those specific ions must be minimized to preserve the first generation fragment ions. Simplifying Equation 4.1 to the following:

$$(4.3) k \propto \frac{1}{v^3}$$

shows the rate of reaction to be inversely proportional to the relative velocity cubed. Increasing the relative velocity of the cations greatly hinders the formation of the collision complex, thus drastically reducing the rate of the reaction.

Stable ions trapped in the linear ion trap oscillate at defined, m/z dependent frequencies as stated in Section 1.2.4. The application of a supplemental AC potential in which the frequency is equal to the frequency of motion of an ion will increase the velocity of that specific ion similarly to that of CAD. However, the amplitude of the supplemental AC potential is below the energy threshold for fragmentation. Multiple AC frequencies can be applied to resonantly accelerate

multiple ions. The sum of the AC frequencies used for acceleration is known as a waveform. A waveform is applied to accelerate all ions except the reagent and precursor for pipETD (Figure 4.7A). Since the accelerated ions have an increased velocity, the reaction rate is severely diminished. This allows the precursor ions to react quickly while the fragment ions are essentially unreactive such that they appear “parked” (Figure 4.7B). Additionally, the precursor is completely reacted while first generation fragments are prevented from sequential electron transfer.

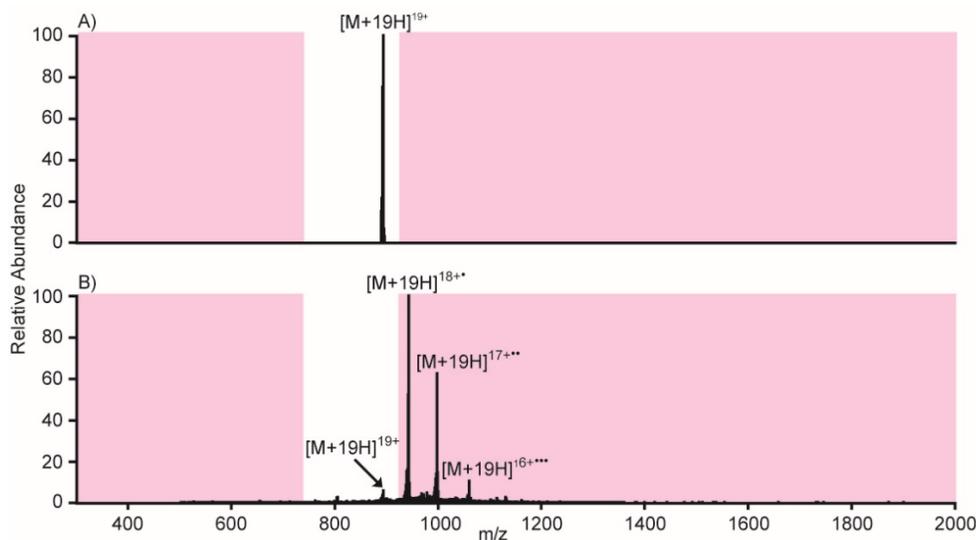


Figure 4.7. Example pipETD Performed on the Thermo Orbitrap Elite. A) Scan of the +19 precursor ions of equine apomyoglobin prior to pipETD. Fragment ions that fall into the pink boxes will be resonantly excited by the supplemental waveform. Ions within these boxes are “parked.” The white area between the parking regions is the precursor notch. Any ion that falls within this window will *not* be resonantly excited. B) The result of 40ms of pipETD on the +19 precursor. The majority of the fragment ions fall within the excitation window and do not fragment further. Noticeably, the precursor ions and the 3rd charge reduced species are low in abundance, implying that the maximum amount of first generation product ions were generated and retained.

Notably, the unreacted precursor ions are low in abundance (<10%). Typically, the efficiency of parallel ion parking is determined by maximizing the first charge reduced species while minimizing the second charge reduced species since the second charge reduced species can only be generated if two ETD reactions occur. However, the 2,2'-biquinoline used as the ETD reagent on the pipETD enabled instrument behaves differently. This reagent appears to undergo two reactions with a single molecule as shown by the observation of a charge-inverted cation. The reagent anion transfers electrons and acquires protons to form cations at 257 and 258 m/z (work

not published). Therefore, this reagent is transferring electrons while simultaneously acquiring protons. The third charge reduced species should be minimal when using this reagent since this species cannot be formed by a single reagent anion. The peak shown in Figure 4.7B is indeed minimized at ~15% relative abundance, demonstrating sufficiently efficient parking.

The ability to maximize the consumption of precursor ions while maintaining the first generation fragment ions is clearly visible in Figure 4.7. However, the majority of the fragment ions are overlapping in the region from ~900-1000 m/z (Figure 4.8A). This greatly inhibits the ability to accurately distinguish fragment ions and prevents the sequence analysis of this protein. A subsequent ion/ion reaction, known as ion/ion proton transfer (IIPT), can be used to overcome this problem (55–58). This reaction involves the co-trapping of positively charge fragment ions and the negatively charged radical reagent anion perfluoro(methyldecalin). However, this reagent does not dissociate the fragment ions. The reagent instead takes a proton from the multiply charged fragment ion to reduce the net charge of the fragment ion by 1 (Figure 4.8B). Since this changes the z value of a given fragment ion, the observed m/z of the ion is increased. The reaction ultimately enables the dispersion of fragment ions over the entire usable mass range of the instrument and eases the identification of fragment ions observed (Figure 4.8C).

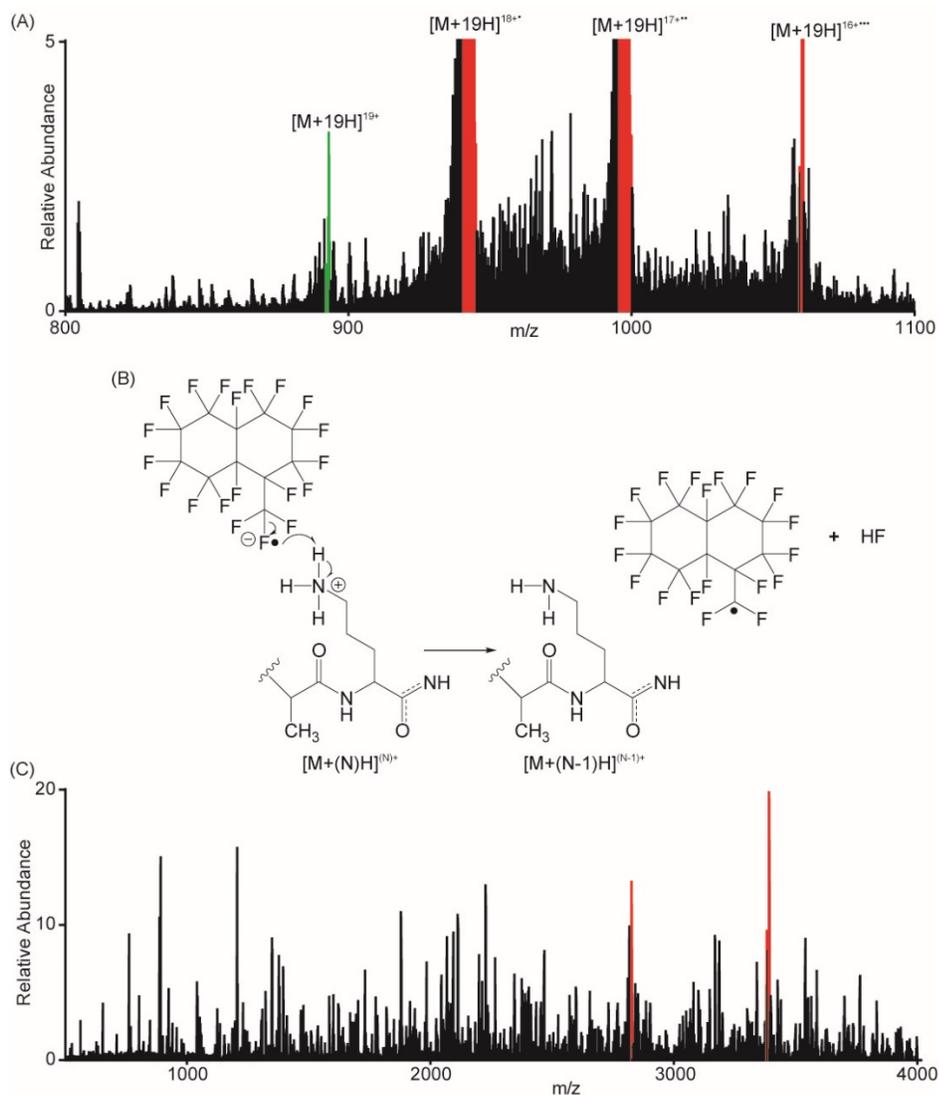


Figure 4.8. IIPT Improves the Ability to Distinguish Fragment Ions. (A) Zoom in on 800-1100 m/z of the 40ms pipETD of the +19 charge state of equine apomyoglobin from Figure 4.7B. The y-axis has been zoomed in on the low abundance species to denote the high density of fragment ions found in the region from 900-1000 m/z . The precursor ions are in green while the charge reduced species are in red. (B) Reaction mechanism for IIPT. The reagent anion, perfluoro(methyldecalin), abstracts a proton from the fragment ions without dissociation to reduce the overall number of charges by 1, thus increasing the m/z of the fragment ion. (C) The ions generated in (A) were reacted with perfluoro(methyldecalin) for 37ms to produce this spectrum. All of the fragment ions that were overlapping in a tight region in (A) were dispersed over the workable mass range of up to 4000 m/z . The charge reduced species are noted in red.

Hinkle et al. utilized parallel ion parking during ETD to extensively characterize two peptides generated by the Aspergillopepsin I digestion of a mAb (51). These peptides were the result of a single cleavage in the V_L and V_H domains and contained all 3 CDRs for each chain. The

determination of the CDR residues with minimal digestion is ideal to give a more confident analysis of the CDRs since this region maintained its connectivity. Sequence coverage of these two peptides were 85% for the light chain fragment and 79% for the heavy chain fragment (51).

4.2.3.2 Disulfide Bond Localization

Zhang et al. sought to localize disulfide bonds in the native state of a murine IgG antibody (48). These covalent bonds link the side chains of two cysteine residues and are imperative for stability, folding, and potential linkage of multiple polypeptide chains (59). Localization of these bonds is important to antibody analysis to ensure proper folding of the therapeutic. NMR and X-ray crystallography can effectively localize disulfides, but these methods require highly concentrated samples. Mass spectrometry is not limited by sample concentration and has become common for disulfide localization (60–62).

Dissociation of disulfide bonded peptides is readily carried out through HCD and ETD. The resulting fragmentation spectrum will have ions corresponding to the two disulfide bound peptides (61, 62). However, ETD also has the unique ability to preferentially break the S-S bond, resulting in ions corresponding to the molecular weights of the previously bound peptides (63, 64). This provides additional information to confirm the sequences of the two disulfide bound peptides.

Proteolytic digestion of the disulfide bound protein is standard procedure (60). Currently, two methods are commonly used to determine disulfide linkages. The first employs digestion with a specific protease followed by LC-MS/MS analysis (60, 61). This straightforward method relies on the ability to accurately dissociate the disulfide bound peptides to give the localization of disulfide bonds. The second method also employs digestion with a specific protease. However, the results of this digestion are split into two samples. The first sample is analyzed by LC-MS/MS analysis to give a non-reduced LC profile. The second sample is reduced to break the disulfide bonds and then analyzed by a separate LC-MS/MS analysis (60, 61). The two LC runs are then compared to determine which peaks are found only in the nonreduced sample. These peaks contain a disulfide bond and can be related back to one another (60, 61).

The ideal pH for most proteases (pH ~8) can lead to disulfide shuffling (61, 65). This artifact involves the rearrangement of natural disulfide bonds to unnatural disulfide bonds at alkaline pHs due to the pKa of the side chain of cysteine (pKa ~8.3), particularly if there are any free cysteines present in the molecule (66). The thiolate anion is prone to shuffle disulfide bonds (Figure 4.9). Therefore, digestions need to be carried out in slightly acidic conditions to prevent any disulfide shuffling. Although using specific proteases at lower pHs with a much lower digestion efficiency is commonplace for digestion, utilizing proteases that are functional in acidic conditions, such as Aspergillopepsin I, ensures that no disulfide shuffling will take place (65).

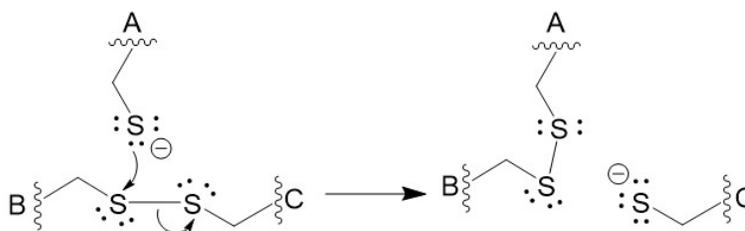


Figure 4.9. Disulfide Shuffling. Free cysteines at alkaline pHs will form thiolate anions. These anions readily react with natural disulfide bonds to shuffle the disulfide bond location. A, B, and C represent three cysteine residues in a theoretical protein. A natural disulfide bond formed between B and C is disrupted by the thiolate anion from A to produce an unnatural disulfide bond between A and B with C leftover as a thiolate anion.

When Zhang et al. digested the native murine IgG with the Aspergillopepsin I enzyme reactor, the antibody maintained its tightly folded structure (48). Four digestion times of 12 s, 93 s, 260 s, and 740 s were necessary to observe peptides that contained each of the 17 unique disulfide bonds. These long digestion times allowed sufficient time for Aspergillopepsin I to cleave residues from the termini toward the hinge region of the molecule. All disulfide bond pairs were successfully identified despite the long and multiple digestions (48).

The primary structure of two bsAbs will be determined in this chapter using the Aspergillopepsin I enzyme reactor. The first bsAb was an IgG-like biosimilar of emicizumab employing a knob-in-hole heavy chain with a common light chain. Recall from Section 3.5.2, the individual protein concentration determines if Aspergillopepsin I will digest the protein in a diffusion or catalytic manner. Since the heavy chains are different, they are effectively at half the concentration of the light chain, and thus are more extensively digested than the light chain.

Emicizumab was treated with protamine to preserve large peptide products sufficient for parallel ion parking. The second bsAb, a BiTE, was digested in a diffusion limited manner to localize the disulfide bonds. Although this protein does not have interchain disulfide bonds like an IgG-like bsAb, this bsAb serves as a good model to determine how lowering the concentration to a diffusion limited system enables full disulfide characterization in a single analysis.

4.3 Materials

Agilent Technologies (Palo Alto, CA)

1100 Series high performance liquid chromatograph

1100 Series vacuum degasser

PLRP-S 3 μ m diameter reverse phase packing material

Anaspec, Inc. (Fremont, CA)

Vasoactive Intestinal Peptide; Human, porcine, rat; > 95% purity

Cobbold Laboratory (Harvard University, Cambridge, MA)

EpCAM-20G6 BiTE

Creative Biolabs (Shirley, NY)

IgG-like bispecific antibody (anti-Factor IX and anti-Factor X), >90% purity

Eppendorf (Hauppauge, NY)

5414R Benchtop centrifuge

Genovis (Cambridge, MA)

IdeS (FabRICATOR), lyophilized

Honeywell (Morristown, NJ)

Burdick and Jackson® Acetonitrile, LC-MS grade

Labconco Corporation (Kansas City, MO)

Centrivap centrifugal vacuum concentrator

Molex (Lisle, IL)

Polymicro Technologies™ polyimide coated fused silica capillary,

360 μ m o.d. x 75 μ m i.d.

Oakwood Chemical (Estill, SC)

Perfluoro(methyldecalin)

Phoenix S&T (Chadds Ford, PA)

Pencil Column Heater

Column Heater Controller

PQ Corporation (Valley Forge, PA)

Kasil – Potassium silicate solution

Protein Metrics (Cupertino, CA)

Byonic™

Sigma Aldrich (St. Louis, MO)

2,2'-Biquinoline, 98%

2-propanol, LC-MS grade

Angiotensin I acetate salt hydrate, 99% purity (human)

Glacial acetic acid, ≥99.99% trace metal basis

Fluoranthene, >99% purity

Formamide, ≥99.5% (GC)

N-(2-Aminoethyl)maleimide trifluoroacetate salt, ≥95% (HPLC), ≥98% (T)

Protamine sulfate salt from salmon, Grade X, amorphous powder (acquired by Dr. Juan Ausió)

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

Thermo Fisher Scientific (San Jose, CA/Bremen, Germany)

Aldehyde/Sulfate Latex Beads, 4% w/v, 1.0 μm

Formic Acid, LC-MS Grade

Orbitrap Elite™ hybrid mass spectrometer, custom modified with front-end ETD

Orbitrap Fusion™ Tribrid™ mass spectrometer

Pierce® Water, LC-MS Grade

Proteome Discoverer™ 2.2

Urea (>99.0%)

4.4 Methods

4.4.1 Aspergillopepsin I Bead Conjugation

Immobilization of Aspergillopepsin I onto 1 μm latex beads was carried out in the same manner as Section 3.4.1.

4.4.2 Bispecific Antibody Preparation

IgG-Like Common Light Chain Bispecific

$\sim 8.5 \mu\text{g}$ of the IgG-like bsAb ACE-910 (biosimilar of emicizumab) was first digested with the enzyme IdeS in 100 mM ammonium bicarbonate (pH ~ 6) to cleave the antibody below the hinge region into the $\text{F}(\text{ab}')_2$ and $\text{Fc}/2$ subunits (19). The cleaved bsAb was evaporated to dryness. The bsAb was then reduced with TCEP and alkylated with NAEM in the same manner as Section 3.4.2 (Figure 4.10). The final volume was brought to 42 μL with digestion buffer (same digestion buffer as Section 3.4.2) to give a concentration of $\sim 0.2 \mu\text{g}/\mu\text{L}$ ($\sim 1.3 \text{ pmol}/\mu\text{L}$).

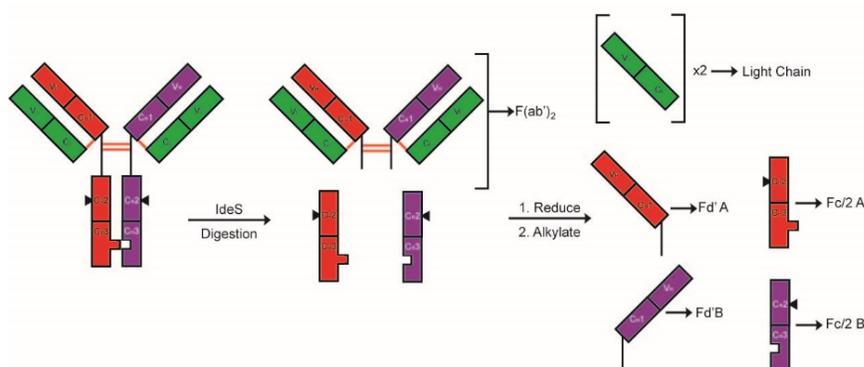


Figure 4.10. Sample Preparation of the Common Light Chain bsAb Prior to Digestion.

The same procedure was followed for the protamine treated sample with the exception that the final volume was brought to 41 μL with digestion buffer. This solution was treated with 0.8 μL of 10 $\mu\text{g}/\mu\text{L}$ protamine in water to give a final concentration of $\sim 0.2 \mu\text{g}/\mu\text{L}$ of antibody and $\sim 0.2 \mu\text{g}/\mu\text{L}$ of protamine.

Non-Reduced BiTE

$\sim 4 \mu\text{g}$ of EpCAM BiTE was diluted to a final volume of 80.4 μL of digestion buffer to give a final concentration of $\sim 0.05 \mu\text{g}/\mu\text{L}$ ($\sim 0.91 \text{ pmol}/\mu\text{L}$). This sample was not treated with protamine.

4.4.3 Bispecific Antibody Digestion with Immobilized Enzyme Reactor

Dried Aspergillopepsin I conjugated beads were suspended in water as described in Section 3.4.3. The column bed length was packed to ~2 mm for digestion of the IgG-like bsAb. The column bed length was packed to ~3 mm for digestion of the BiTE. The beads were washed with water for 30 minutes at 500 psi to compact them. Digestion buffer was flowed for 5 minutes at 200 psi to prevent the denatured antibodies from coming in contact with non-denaturing conditions. The sample to be digested was flowed over the bed at a given flow rate to achieve the desired digestion time as given by Equation 3.2. Samples were collected for a calculated amount of time to give ~10 μL of digestion products.

4.4.4 Liquid Chromatography and Mass Spectrometry

An Agilent Technologies 1100 Series binary HPLC system was coupled to a Thermo Scientific Orbitrap Fusion tribrid mass spectrometer operating in intact protein mode for analysis of all digestions. ~400 nL of each digestion was pressure-loaded onto an analytical column packed with 10 cm of 3 μm diameter, 300 \AA PLRP-s packing material within a 360 μm O.D. x 75 μm I.D. fused silica capillary. The electrospray tip was integrated ~5 cm from the end of the packing material to allow for the use of a column heater. 100 fmol of the internal standards vasoactive intestinal peptide and angiotensin I were also loaded onto the column. The column was washed for 1 hour with 0.3% formic acid in water (Solvent A) to remove all salts. Peptides were then gradient eluted at ~100 nL/min with a gradient of 0-25-50-100%B (72% ACN, 18% IPA, 10% water, 0.3% formic acid) in 0-5-80-85 minutes at 50°C.

A tiered decision tree was used to fragment the eluting peptides in a data-dependent manner based on size (Figure 4.11). A 60,000 resolution Orbitrap MS^1 scan was taken to obtain intact molecular weights of all eluting species. Precursor ions were selected based on this scan for both ETD and collisional fragmentation for three seconds before the next Orbitrap MS^1 scan. First priority was given to peptides with charge states ≥ 10 within 500-925 m/z for ETD fragmentation and charge states ≥ 8 within 1100-1500 m/z for HCD fragmentation. These scans were taken at 120,000 resolution in the Orbitrap with precursor targets of $1e6$ and an additional microscan before

converting the frequency analysis to m/z. Second priority was given to peptides with charge state from 5-9 within 300-925 m/z for ETD and charge states 4-7 for HCD. These scans were taken at 60,000 resolution in the Orbitrap with precursor targets of $4e5$. Third priority was given to peptides with charge states 3-4 for ETD and 2-4 for CAD. These scans were taken in the ion trap at normal scan speed with precursor targets of $2e4$. Since these scans were taken in the ion trap, fragmentation analysis of these small peptides occurred in parallel to scans taken in the Orbitrap due to the unique instrument structure of the Thermo Fusion Tribrid (50). Individual dynamic exclusion lists were used for each of the six unique fragmentation categories. First priority scans were placed on an exclusion list after 4 replicates, second priority scans after 3 replicates, and third priority after 2 replicates. All ETD scans used the calibrated reaction times in high capacity mode (67–69). All HCD scans used stepped normalized collision energies of 22-25-28%. All CAD scans used 30% normalized collision energy for 10 ms.

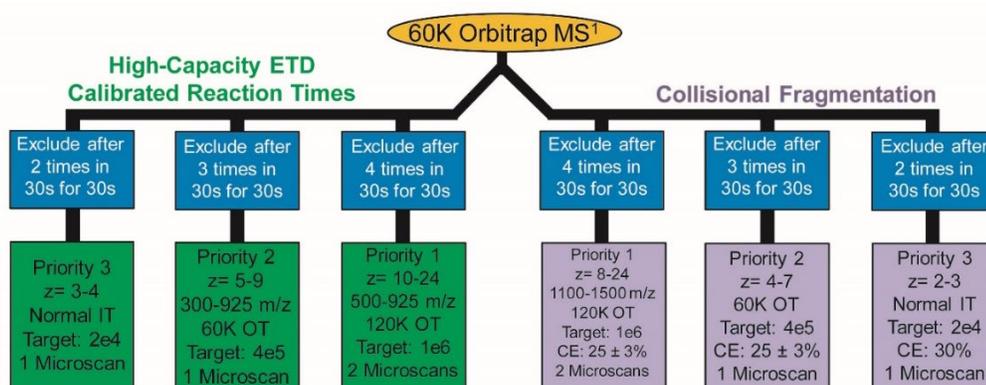


Figure 4.11. Data-Dependent Tiered Decision Tree.

The HPLC system was coupled to an in-house modified Thermo Scientific Orbitrap Elite hybrid mass spectrometer to employ targeted pipETD of large peptides from the IgG-like bsAb (70). The same amount of the IgG-like digestion treated with protamine was loaded and washed on the same analytical column. The peptides were gradient eluted using a gradient of 0-31-37-100%B in 0-5-65-70 minutes at 50°C.

The parking waveform developed by Duselis et al. was employed with the q value of the ETD reagent set to 0.75 (71). Ions that fall within the parked values were excited to diminish the rate of secondary ETD reactions. A region about 90-120 m/z around the precursor was not excited to allow

the precursor to react at the normal rate. Additionally, the reagent was excited slightly to reduce the overall reaction rate. An in-house script was written in the ion trap control language (ITCL) to allow for multiple precursor targets. Two precursors were targeted for the first 40 minutes of analysis before switching to two different targets for the remaining time.

Two separate analyses were performed. The first analysis was for pipETD only while the second was for pipETD followed by IIPT. The first two targets for fragmentation were 820.90 m/z and 773.73 m/z with isolation windows of ± 4 m/z. The light chain I21-E105 and heavy chain 1 F68-K153 peptides were both fragmented within the 820.90 m/z window. These peptides were fragmented with 95 ms of pipETD. The heavy chain 1 V5-R67 peptide was fragmented with 125 ms of pipETD. The second two targets for fragmentation were 823.78 m/z and 872.24 m/z with isolation windows of ± 4 m/z. Both of these masses targeted the heavy chain 2 V20-K149 peptide at different charge states. Each window was fragmented with 40 ms of pipETD. The second analysis employed 33 ms of IIPT following the designated pipETD times. Five multiple fills of the C-trap with fragment ions were employed prior to analysis by the Orbitrap (70).

4.4.5 Data Analysis

The data files were searched using Byonic (version 3.3.11) within Proteome Discoverer (version 2.2.0.386) for the IgG-like bsAb digestions performed on the Thermo Fusion Orbitrap. Proteome Discoverer was used to split the .raw file into two separate searches. Both searches used a 10 ppm molecular weight tolerance against a nonspecific digestion of the 5 individual subunit sequences. The first search was limited to the high resolution Orbitrap data with a fragment mass tolerance of 15 ppm and a maximum precursor mass of 20,000 Da. The second search was limited to the low resolution ion trap data with a fragment mass tolerance of 0.35 Da and a maximum precursor mass of 6,000 Da. Both searches included a fixed modification for NAEM on C, common modifications of oxidation at M, common pyro-Gln at N-terminal Q, pyro-Glu at N-terminal E, and rare modifications of HexNAc at S and T, and Hexose at S and T. For glycan tree analysis of the Fc, the HexNAc total was held constant at 4 while the Hexose total could range from 3-5 and the fucose total could range from 0-1. The peptide abundances from the treated and untreated

digestions of the IgG-like bsAb were compared in Proteome Discoverer. The mass areas of each Byonic identified peptide was calculated using a 10 ppm precursor tolerance and a minimum S/N ratio of 5. The ratios of the treated area to the untreated area were calculated.

Manual annotation of the parked pipETD/IPT MS² scans from the protamine treated IgG-like bispecific was carried out. The total number of scans within the chromatographic peak width were averaged using Qual Browser.

An initial search of the disulfide bound BiTE was performed in Proteome Discoverer as described above against a database containing the single protein sequence and the common modifications of oxidation at M and pyro-Glu at N-terminal E. This search was used to identify all peptides that did not contain a cysteine. The Byonic node in Proteome Discoverer does not have the option to search for disulfide bonds. Therefore, the .raw file was converted using in-house software into 4 MGF files consisting of 1) Orbitrap ETD scans, 2) Orbitrap HCD scans, 3) ion trap ETD scans, and 4) ion trap CAD scans. Each MGF file was searched individually using the stand-alone Byonic program with the disulfide option selected and a cleavage specificity of C-terminal to K,L,R,D,P,E and N-terminal to V,L,S,A,T,I. Although Aspergillopepsin I is nonspecific, the search software fails if more amino acids are added to the specificity due to the exponential increase in potential peptide combinations to search. However, since the number of missed cleavages is set to unlimited, the search will still perform a nonspecific search. The ion trap data file searches contained a maximum precursor mass of 6,000 Da while the Orbitrap data file searches contained a maximum precursor mass of 10,000 Da. Selected fragmentation scans were chosen for manual annotation to reconstruct the sequence and characterize the disulfide bonds.

4.5 Results and Discussion

4.5.1 Complete Sequence Coverage of an IgG-Like Bispecific Antibody

Our lab recently demonstrated the ability to characterize a mAb using the enzyme reactor in ~0.9 s at an overall concentration of 0.2 µg/µL (49). Each chain of a mAb has the same primary structure and represents 1/3 of the overall concentration (0.067 µg/µL each). However, when shifting to IgG-like bsAbs with knob-in-hole mutations and different light chains, each subunit

contains a different primary structure, effectively halving their concentration ($0.033 \mu\text{g}/\mu\text{L}$ each) in comparison to a mAb. As shown in Chapter 3, decreasing the concentration of a given protein resulted in an increase of observed cleavages by Aspergillopepsin I. Therefore, it is expected that bsAb chains will be more highly digested than a mAb. The biosimilar of emicizumab was used to demonstrate the effects of reducing the concentration for digestion by Aspergillopepsin I in a biological system. This particular bsAb was chosen because it has a simplified structure containing a common light chain. Therefore, the concentration of the light chain is double that of the Fd' and Fc/2 subunits.

The bsAb was digested for ~ 1 s at an overall concentration of $0.2 \mu\text{g}/\mu\text{L}$ (Figure 4.12). Qualitatively, the majority of the ion current lies early in the gradient (~ 22 - 35 minutes). This implies the generation of primarily small to medium sized peptides (~ 0.8 - 5.0 kDa). These results agree with the expected results: reducing the subunit concentration results in an increase in observed cleavages and thus a higher abundance of small to medium peptides.

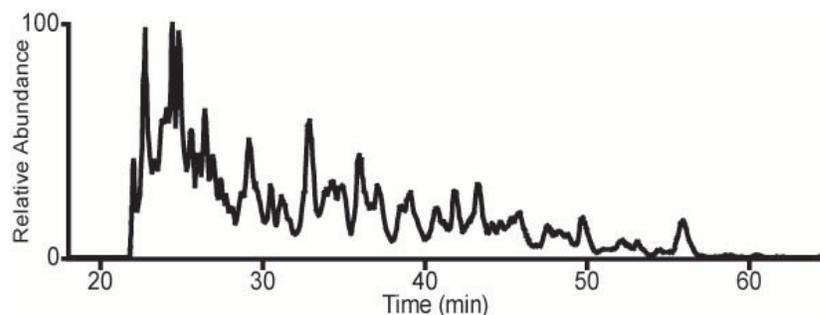


Figure 4.12. Total Ion Current Chromatogram of an ~ 1 s Digestion of Emicizumab.

Although the majority of the peptides were low in molecular weight, enough overlapping peptides were observed to give comprehensive sequence coverage. A database search against the known sequence identified 1,091 unique peptides using a 1% FDR cutoff. Peptides generated from the C_{H1} domains of the heavy chain and the majority of each Fc/2 chain maintain the exact same primary structure. Therefore, these peptides cannot be attributed to one particular chain over the other. Only peptides that contain portions of the V_H region of the Fd' and the knob-in-hole point mutations can be definitively attributed to a corresponding chain.

Although the identified peptides can be used to obtain near complete sequence coverage, the high abundance of small peptides is not ideal for analysis. The largest peptide observed in this analysis was the ~7.8 kDa peptide N₅₀-K₁₀₂ of the Fc/2 region containing the G1F glycan tree modification contributing 1.6 kDa of the total molecular weight. Larger peptides are desired such that large portions of the protein will be found in a single peptide, providing high confidence in the connectivity of the primary structure. It is particularly desirable to maintain connectivity of the 3 CDRs on a single peptide. The bsAb was treated with protamine prior to digestion to efficiently generate larger peptides.

Introducing protamine to a sample for digestion will decrease the number of observed cleavages because protamine competitively inhibits Aspergillopepsin I as demonstrated in Chapter 3. This treatment should allow for the retention of large, desirable peptides for analysis. The IdeS digested bsAb was reduced, alkylated, and treated with protamine such that both the bsAb and protamine had a concentration of 0.2 µg/µL. The results of an ~1 s digestion are shown in Figure 4.13. Qualitatively, the ion current is more evenly distributed throughout the gradient in comparison to the untreated digestion (Figure 4.12). Importantly, the ion current later in the gradient (~30 min onwards) is higher in relative abundance, indicative of maintaining the larger peptides through the digestion.

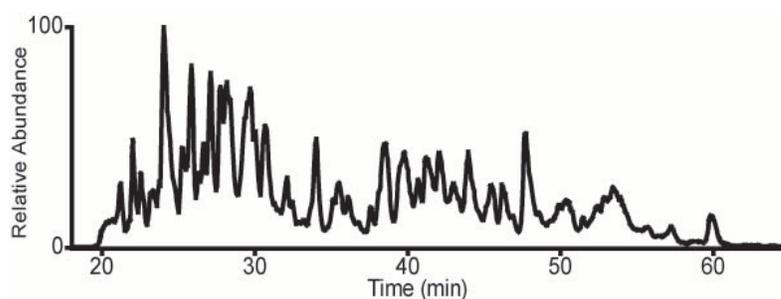


Figure 4.13. Total Ion Current Chromatogram of an ~1s Digestion of Emicizumab Treated with Protamine

Proteome Discoverer was used to identify peptides in both analyses and determine their respective mass areas. The ratio of the mass areas was calculated to show the change in abundance (Figure 4.14). Since these analyses contained only one data point each, a wide tolerance of greater than 4 times the difference in abundance was chosen as a threshold for

considering a change to be significant. Minor differences in the experimental setup, such as differing ion suppression due to lack of peptides eluting at a given time and slight changes in the flow rate, could have influenced the abundance of a given peptide. ~61% of the peptides observed in both analyses had similar abundances using this wide tolerance. When comparing the remaining ~40% of identified peptides, considerably more peptides were found in the treated digestion. ~12% of identified peptides were unique and ~18% of identified peptides were greater than 4 times in abundance in the treated digestion while ~2% of identified peptides were unique and ~8% were greater than 4 times in abundance in the untreated digestion. The increase in abundance of the 18% of peptides found in the treated sample had additional benefits. The time needed to reach the desired target for dissociation was reduced, resulting in reduced cycle time which allowed lower level peptides to be chosen for fragmentation. The increase in signal also generates a higher quality fragmentation spectrum.

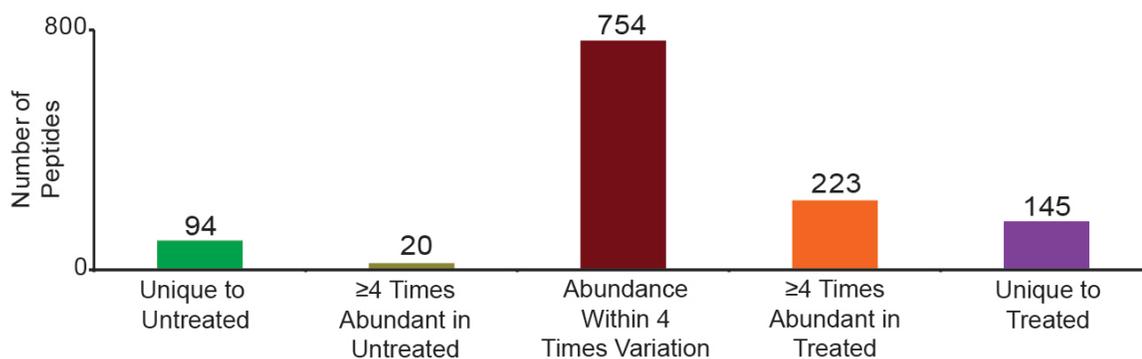


Figure 4.14. Histogram of Peptides Identified in the Digestion of Emicizumab Grouped by Change in Abundance.

The log of the ratio was plotted against the peptide molecular weight to give Figure 4.15. The point colors in Figure 4.15 match the colors of the histogram bars in Figure 4.14, and the dashed lines show the boundaries of the four-fold threshold. The goal of treating the sample with protamine was to increase the abundance of large molecular weight peptides. Figure 4.15 clearly shows that this goal was achieved. 42.8% and 29.6% of the 145 unique peptides and 223 peptides that significantly increased in abundance were greater than 5 kDa in molecular weight.

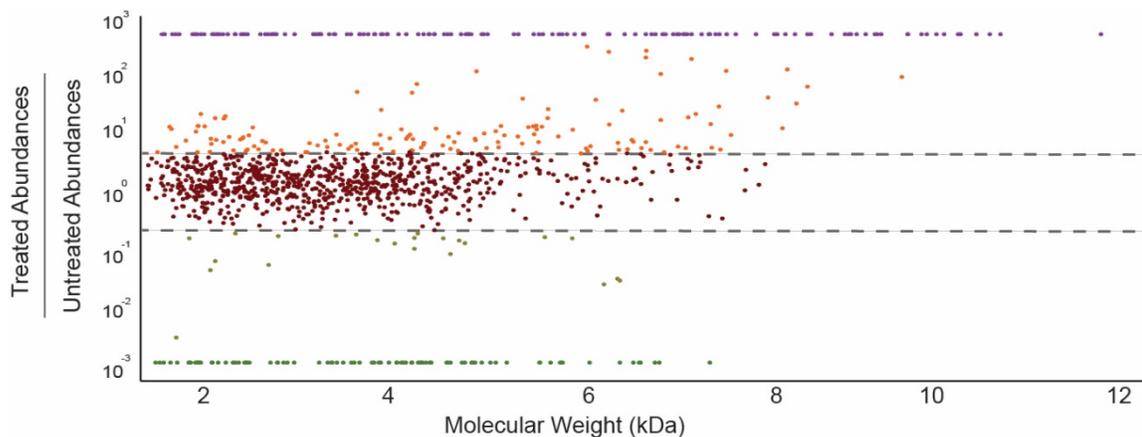


Figure 4.15. Scatter Plot of the Identified Peptides in the Digestions of Emicizumab versus the Molecular Weight. The colors of the points correspond to the colors of the grouping in the histogram depicted in Figure 4.14.

A corresponding fragment ion must be observed between each amino acid in the primary structure for unambiguous sequence coverage of a protein. When using a single specific protease, it is impossible to have a fragment ion at the enzymatic cleavage site. Additionally, low mass ions are typically not observed because they fall below the mass range. Aspergillopepsin I lacks specificity and thus provides multiple cleavage points across the molecule, allowing for multiple peptides to be used to obtain confirmatory fragment ions that may be absent in other peptides.

A total of 58 unique peptides were used to confirm the sequence, representing only ~5.4% of the total identified peptides. These included 14 light chain peptides, 17 Fd' A peptides (12 unique), 14 Fc/2 A (2 unique), 14 Fd' B peptides (9 unique), and 16 Fc/2 peptides (4 unique). Total sequence coverage of ~99.5% was obtained since five cleavages out of a possible 1,101 cleavages did not have a corresponding fragment ion (see Figure 4.16 for peptide map and Figure 4.17 for observed fragment ions). Two of the five missed cleavages lacked observed peptides at enzymatic cleavage sites: D₁₅₁/N₁₅₂ of the light chain and D₂₉/V₃₀ of the Fc/2. Since there are no peptides overlapping these cleavage points, there must be a strong preference for cleavage by Aspergillopepsin I at these sites. Overall, the treatment with protamine prior to digestion gave the expected results: large peptides are retained due to the decrease in the number of enzymatic cleavages.

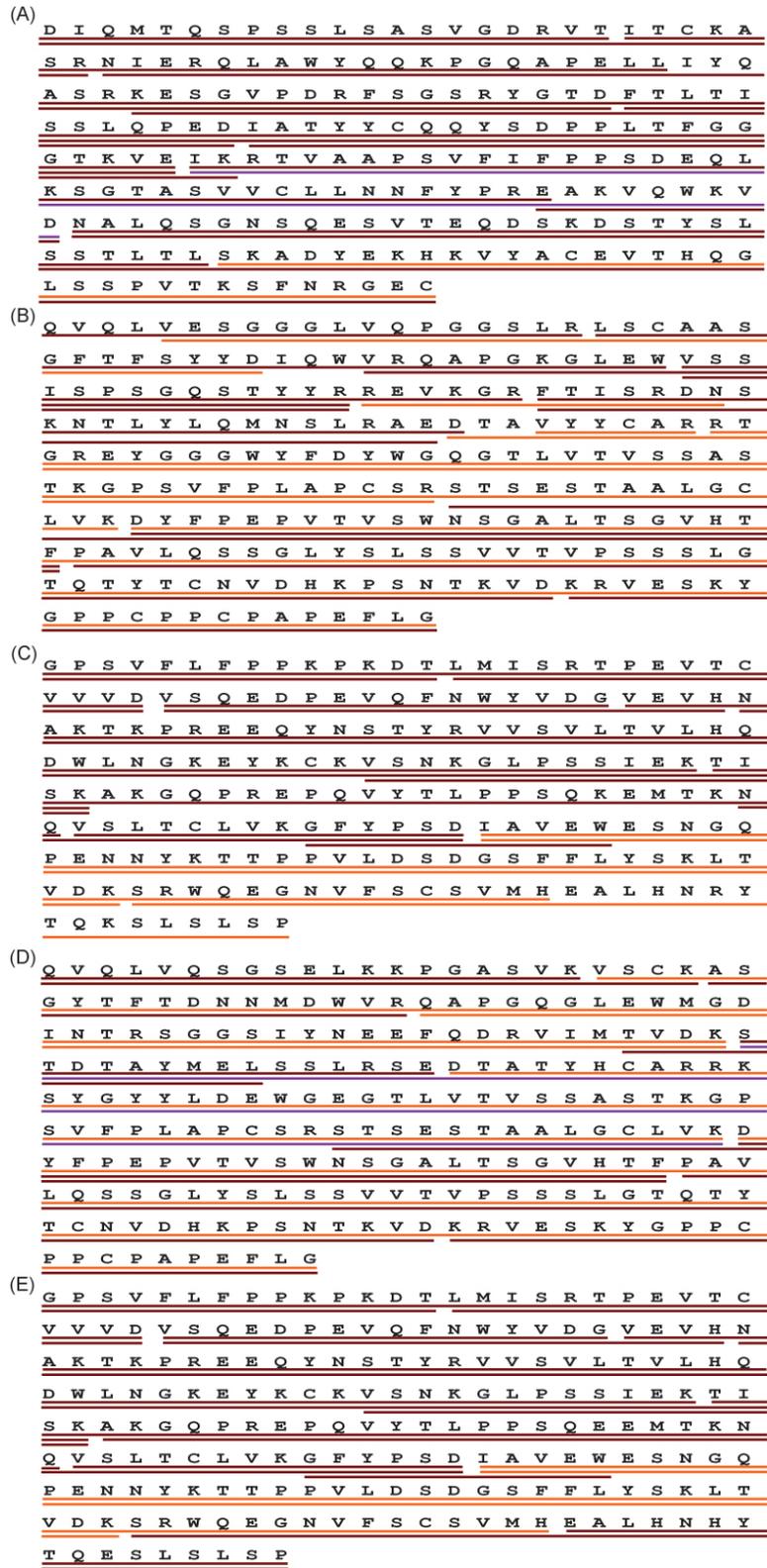


Figure 4.16. Peptide Map Used for Fragment Ion Analysis from the Digestion of Emicizumab. A) Light chain. B) Fd' A. C) Fc/2 A. D) Fd' B. E) Fc/2 B. The lines indicate the peptides used, and the color of the lines correspond to the colors of the ratios shown in Figures 4.14 and 4.15.

(A) D I Q M T Q S P S S L S A S V G D R V T I T C K A
S R N I E R Q L L A W Y Q Q K P G Q A P E L L I Y Q
A S R K E S G V P D R F S G S R Y G T D F T L T I
S S L Q P E D I A T Y Y C Q Q Y S D P P L L T F G G
G T K V E I K R I T V A A P S V F I F P P S D E Q L
K S G T A S V V C L L N N F Y P R E A K V Q W K V
D N A L L Q S G N S Q E S V T E Q D S K D S T Y S L
S S T L T L S K A D Y E K H K V Y A C E V T H Q G
L S S P V T K S F N R G E C

(B) Q V Q L L V E S G G G L V Q P G G S L R L S C A A S
G F T F S Y Y D I Q W V R Q A P G K G L E W V S S
I S P S G Q S T Y Y R R E V K G R F T I S R D N S
K N T L Y L Q M N S L R A E D T A V Y Y C A R R T
G R E Y G G W Y F D Y W G Q G T L V T V S S A S
T K G P S V F P L A P C S R S T S E S T A A L G C
L V K D Y F P E P V T V S W N S G A L T S G V H T
F P A V L Q S S G L Y S L S S V V T V P S S S L G
T Q T Y T C N V D H K P S N T K V D K R V E S K Y
G P P C P P C P A P E F L G

(C) G P S V F L F P P K P K D T L M I S R T P E V T C
V V V D V S Q E D P E V Q F N W Y V D G V E V H N
A K T K P R E E Q Y N S T Y R V V S V L T V L H Q
D W L N G K E Y K C K V S N K G L P S S I E K T I
S K A K G Q P R E P Q V Y T L P P S Q K E M T K N
Q V S L T C L V K G F Y P S D I A V E W E S N G Q
P E N N Y K T T P P V L D S D G S F F L Y S K L T
V D K S R W Q E G N V F S C S V M H E A L H N R Y
T Q K S L S L S P

(D) Q V Q L L V Q S G S E L L K K P G A S V K V S C K A S
G Y T F T D N N M D W V R Q A P G Q G L E W M G D
I N T R S G G S I Y N E E F Q D R V I M T V D K S
T D T A Y M E L S S L R S E D T A T Y H C A R R K
S Y G Y Y L D E W G E G T L V T V S S A S T K G P
S V F P L A P C S R S T S E S T A A L G C L L V K D
Y F P E P V T V S W N S G A L T S G V H T F P A V
L L Q S S G L Y S L S S V V T V P S S S L G T Q T Y
T C N V D H K P S N T K V D K R V E S K Y G P P C
P P C P A P E F L G

(E) G P S V F L F P P K P K D T L M I S R T P E V T C
V V V D V S Q E D P E V Q F N W Y V D G V E V H N
A K T K P R E E Q Y N S T Y R V V S V L T V L H Q
D W L N G K E Y K C K V S N K G L P S S I E K T I
S K A K G Q P R E P Q V Y T L P P S Q E E M T K N
Q V S L T C L V K G F Y P S D I A V E W E S N G Q
P E N N Y K T T P P V L D S D G S F F L Y S K L T
V D K S R W Q E G N V F S C S V M H E A L H N H Y
T Q E S L S L S P

Figure 4.17. Observed Fragment Ions from the Digestion of Emicizumab. A) Light chain. B) Fd' A. C) Fc/2 A. D) Fd' B. E) Fc/2 B. Red hash marks represent c- and z'-ions while blue hash marks represent b- and y-ions. The CDRs are underlined in (A), (B), and (D).

4.5.2 Using Protamine to Generate Peptides with all CDRs for Further Investigation Using Parallel Ion Parking

Digesting an IgG-like bsAb treated with protamine for ~1 s was an excellent method to obtain complete sequence coverage of the molecule. However, this digestion time was insufficient for generating peptides that maintain continuity across all 3 CDRs in each respective chain. Previous work by Hinkle et al. on mAbs provided evidence that these ideal peptides for the light and heavy chains are the result of a single cleavage at E₁₀₅/I₁₀₆ and K₁₄₉/D₁₅₀, respectively (51). These residues were also present in emicizumab. Therefore, the generation of these large peptides containing all 3 CDRs on a single peptide should be obtainable. An attempt was made to generate these peptides by reducing the enzyme reactor digestion time. Reducing the digestion time decreased the interaction time between the sample and Aspergillopepsin I. This should consequently reduce the number of cleavages, generating larger peptides.

Sample constraints limited this experiment since only a single digestion time could be obtained with the remaining sample of emicizumab treated with protamine. The digestion time acquired was ~800 ms and produced the chromatogram shown in Figure 4.18. Although the chromatogram has similar features to the ~1s digestion (Figure 4.13), large peptide products are found in the ~800 ms digestion that were not present in the ~1 s digestion.

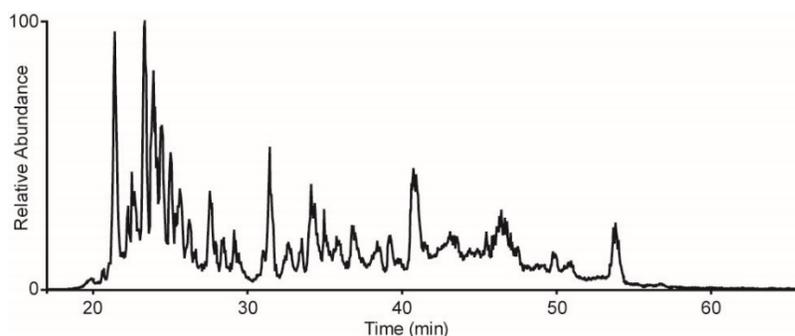


Figure 4.18. Total Ion Current Chromatogram of an ~800 ms Digestion of Emicizumab Treated with Protamine.

Unfortunately, the ideal single cleavage peptides at the predicted cleavage points were not observed. However, two peptides were created as a result of two enzymatic cleavages, including one from the expected cleavage site. The peptides did maintain all 3 CDRs on a single peptide for

both the light chain and heavy chain B for further investigation by parallel ion parking. A single peptide was not observed that contained all 3 CDRs for heavy chain A. This chain had multiple preferred cleavage sites within the variable region. However, two peptides were investigated further by parallel ion parking. The first peptide contained CDRs 1 and 2 while the second peptide contained CDR 3. Although not ideal, two out of the three CDRs maintained continuity. All four peptides were targeted for parallel ion parking during ETD (pipETD) in a chromatographic analysis. Figure 4.19 shows the chromatogram of the analysis with filters for each targeted peptide and its respective MS¹.

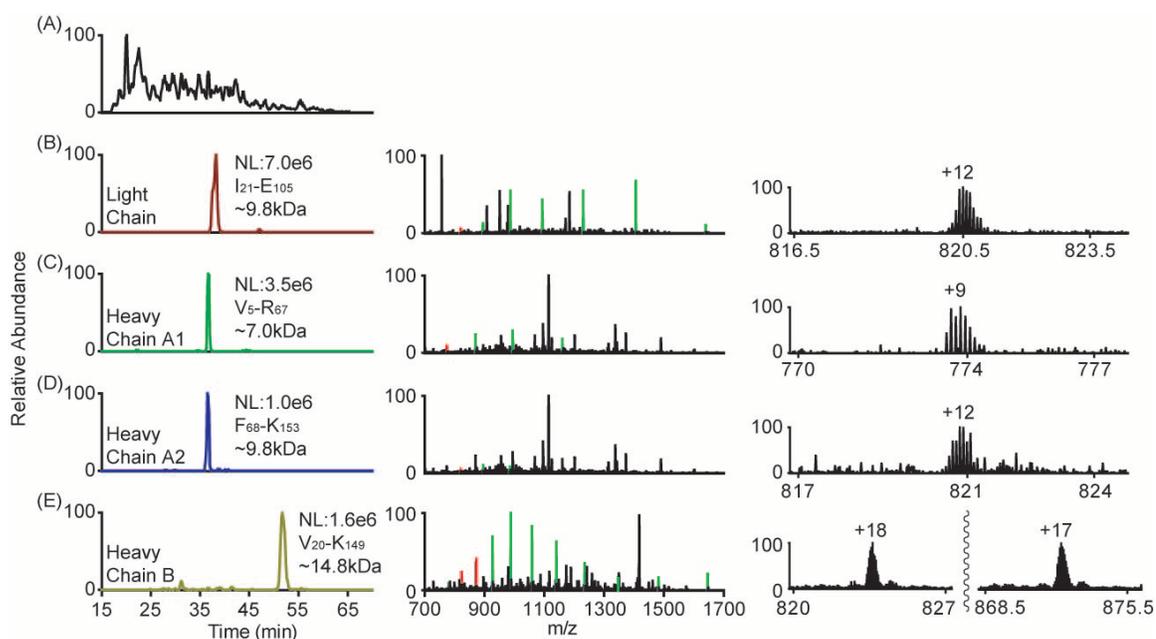


Figure 4.19. Peptides Targeted for Further Investigation by pipETD. The first column corresponds to the chromatogram. The second column corresponds to the MS¹ at the time that the peptides of interest elute. The green peaks represent the charge states of the peptide of interest while the red peaks represent the targeted peak for dissociation. The third column is a magnification of the peak targeted for dissociation. The charge state of the peak is noted. (A) Total ion current chromatogram. (B) Mass filter for the peptide of interest from the light chain. (C) Mass filter for the first peptide of interest from heavy chain A. (D) Mass filter for the second peptide of interest from heavy chain A. (E) Mass filter for the peptide of interest from heavy chain B.

A single charge state was targeted for pipETD of the light chain and the two heavy chain A peptides. Since the heavy chain B peptide eluted later in the gradient, the cycle time of the reactions allowed for two charge states to be targeted. The supplemental waveform as described in section 4.2.3.1 was applied during ETD to ensure that only a single electron transfer event occurred (Figure

4.20A). All products underwent IIPT to spread the signal across the mass range (Figure 4.20B). The pipETD products underwent enough IIPT such that minimal charge reduced species were maintained in the mass range, allowing the C-trap to be filled with more sequence informative fragment ions before reaching the maximum capacity of the C-trap ($\sim 1e6$ charges). These scans were manually inspected for sequence coverage (see Figure 4.21 for fragmentation maps and Table 4.1 for percentage analysis).

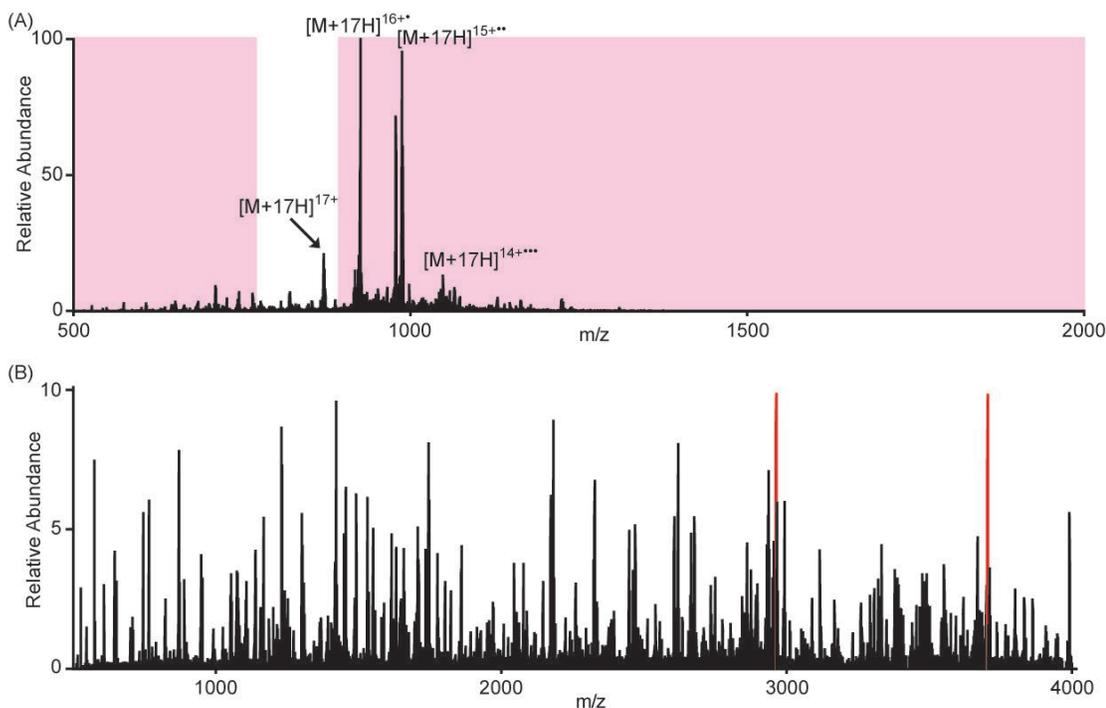


Figure 4.20. pipETD of a Large Peptide Containing all 3 CDRs. (A) The heavy chain B +17 charge state underwent 40ms of pipETD. The m/z ranges that were resonantly excited by the supplemental waveform are located in the pink boxes. The precursor ions were not excited to ensure maximum consumption, as noted by the low relative abundance of the $[M+17H]^{17+}$ peak.

The products underwent minimal secondary ETD reactions as noted by the $<10\%$ relative abundance of the $[M+17H]^{14+}$ peak. (B) All of the products generated in (A) underwent 33ms of IIPT to spread the fragment ions over the usable mass range to give the MS^2 shown. Notice the y-axis scale is zoomed in on the lower 10% of the spectrum. The most abundant species in the spectrum are the charge reduced species, noted as the red peaks in this scan. These peaks set the scale level of the scan, while the fragment ions are below 10% relative abundance.

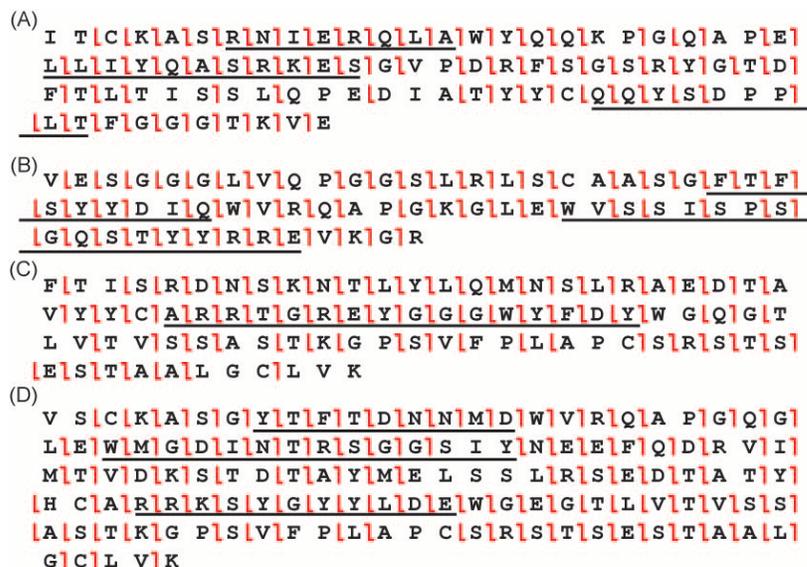


Figure 4.21. Sequence Coverage of the Four Peptides of Interest. (A) Light chain. (B) Heavy chain A peptide 1. (C) Heavy chain A peptide 2. (D) Heavy chain B peptide. The CDRs for each chain are underlined.

Table 4.1. Sequence Coverage for the pipETD followed by IIPT Data of the Four Peptides of Interest.

Peptide	Sequence Coverage	Observed c-ions	Observed z'-ions	CDR 1 Coverage	CDR 2 Coverage	CDR 3 Coverage
Light Chain	85%	76%	61%	8/8 (100%)	11/11 (100%)	6/8 (75%)
Heavy Chain A1	89%	76%	74%	8/9 (89%)	14/17 (82%)	N/A
Heavy Chain A2	82%	65%	68%	N/A	N/A	16/16 (100%)
Heavy Chain B	87%	71%	62%	9/9 (100%)	12/14 (86%)	11/11 (100%)

Table 4.1 depicts the sequence coverage obtained for the light chain, heavy chain A1, heavy chain A2, and heavy chain B peptides. The peptide coverages were 85%, 89%, 82%, and 87% respectively. Although not 100%, in-depth coverage was obtained for the desired CDR regions across the peptides. Eight out of a possible 103 cleavages within the 9 CDRs were not observed in the fragmentation scans. However, three of these cannot be observed because ETD does not fragment at proline residues.

Although the theoretical peptides resulting from a single cleavage were not maintained after ~800 ms of digestion, other peptides were maintained that gave the same intended result. These few peptides were only observed because protamine inhibited the rate of Aspergillopepsin I digestion. If more sample was available, then a ~500 ms digestion time could be tested to potentially generate the theoretical peptides of interest. However, this experiment provides ample evidence for the application of protamine treatment to further investigate low concentration proteins. Protamine treatment inhibits the rate of digestion and results in an increase in large molecular weight products which can be targeted by pipETD to maintain greater connectivity of the protein.

4.5.3 Complete Sequence Coverage of a BiTE with Disulfide Bond Localization in a Single Chromatographic Analysis

Although generating large peptide products is beneficial for maintaining connectivity, there are other instances where smaller peptides are better suited for analysis. One such case is the analysis of disulfide bound peptides. Zhang et al. localized 17 disulfide bonds in a murine IgG antibody using the first generation of Aspergillopepsin I enzyme reactor (48). However, it took four different digestion times (12 s, 93 s, 260 s, and 740 s) to successfully generate the peptides needed to confidently localize these bonds. The ability for Aspergillopepsin I to cleave was hindered by the higher order structure of the antibody. The terminal regions were the most exposed, causing these regions to be cleaved first. The center of the antibody became more accessible as the enzyme cleaved from the termini. The hinge region required the longest digestion time to obtain a peptide of reasonable size for successful analysis.

The previous work by Zhang et al. for disulfide analysis was carried out at a protein concentration of 0.2 $\mu\text{g}/\mu\text{L}$ (48). Additionally, selected disulfide bound peptides were targeted for fragmentation after comparison to the reduced analysis of the same digest. Improvements implemented in the second generation of enzyme reactor by Hinkle et al. provides ample room for improvement on the methodology presented by Zhang et al. Additionally, reducing the concentration of a protein for analysis will increase the number of observed cleavages by Aspergillopepsin I as described in Chapter 3. Therefore, the digestion of a disulfide bound protein

into manageable sized peptides can be carried out in a single digestion at a lower concentration. Additionally, leveraging the instrument control software of the Thermo Orbitrap Fusion enables the ability to fragment the majority of the digestion products in a data-dependent manner, allowing for the analysis of a disulfide bound protein in a single analysis.

A BiTE was used to determine whether reducing the protein concentration and implementing the tiered decision tree method was adequate for analyzing disulfide bound peptides and sequence reconstruction. Although this molecule is different than the murine IgG analyzed by Zhang et al., it provides a simpler set of parameters to determine if the methodology is sufficient for this type of analysis. Ideally, a successful digestion will generate peptides that are <10 kDa in molecular weight to ensure that only a single cysteine residue is present on a given peptide to better localize the disulfide bond.

The BiTE was first diluted to 0.05 $\mu\text{g}/\mu\text{L}$ in digestion buffer and then successfully digested by the enzyme reactor in ~ 1.9 s. Figure 4.22 shows the chromatogram of the digestion products. Qualitatively, the majority of the ion current, and thus digestion products, are within the first ~ 12 minutes of the analysis signifying the generation of small-medium peptides. However, there is still significant ion current for the remainder of the chromatogram. The chromatogram reflects the generation of more manageable sized peptides for disulfide analysis.

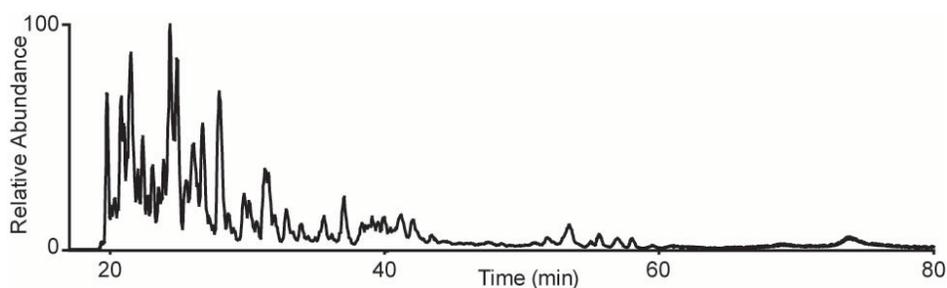


Figure 4.22. Total Ion Current Chromatogram of the ~ 1.9 s Digestion of Nonreduced BiTE.

Figure 4.23 depicts an example peptide that was disulfide bound. This peptide was dissociated by both HCD and ETD using the tiered decision tree data-dependent method on the Thermo Orbitrap Fusion (Figure 4.11). When dissociating disulfide bound peptides, two ion series were present in the resulting MS^2 spectrum. Respectable coverage was generated by the HCD scan with

the exception of the residues around the disulfide bound cysteine (Figure 4.23D). Corresponding fragment ions were not observed, preventing the primary structure from being confirmed. However, the disulfide bond can be deduced due to the clear presence of two ion series without interference in the MS¹ (Figure 4.23C). Figure 4.23E depicts the ETD dissociation of the same peptide. ETD has an added benefit over HCD in that the disulfide bond can be broken, resulting in two peaks corresponding to the molecular weight of the two bound peptides (noted by the blue peaks in Figure 4.23E). ETD obtains coverage throughout each peptide, and localizes one of the two cysteine residues.

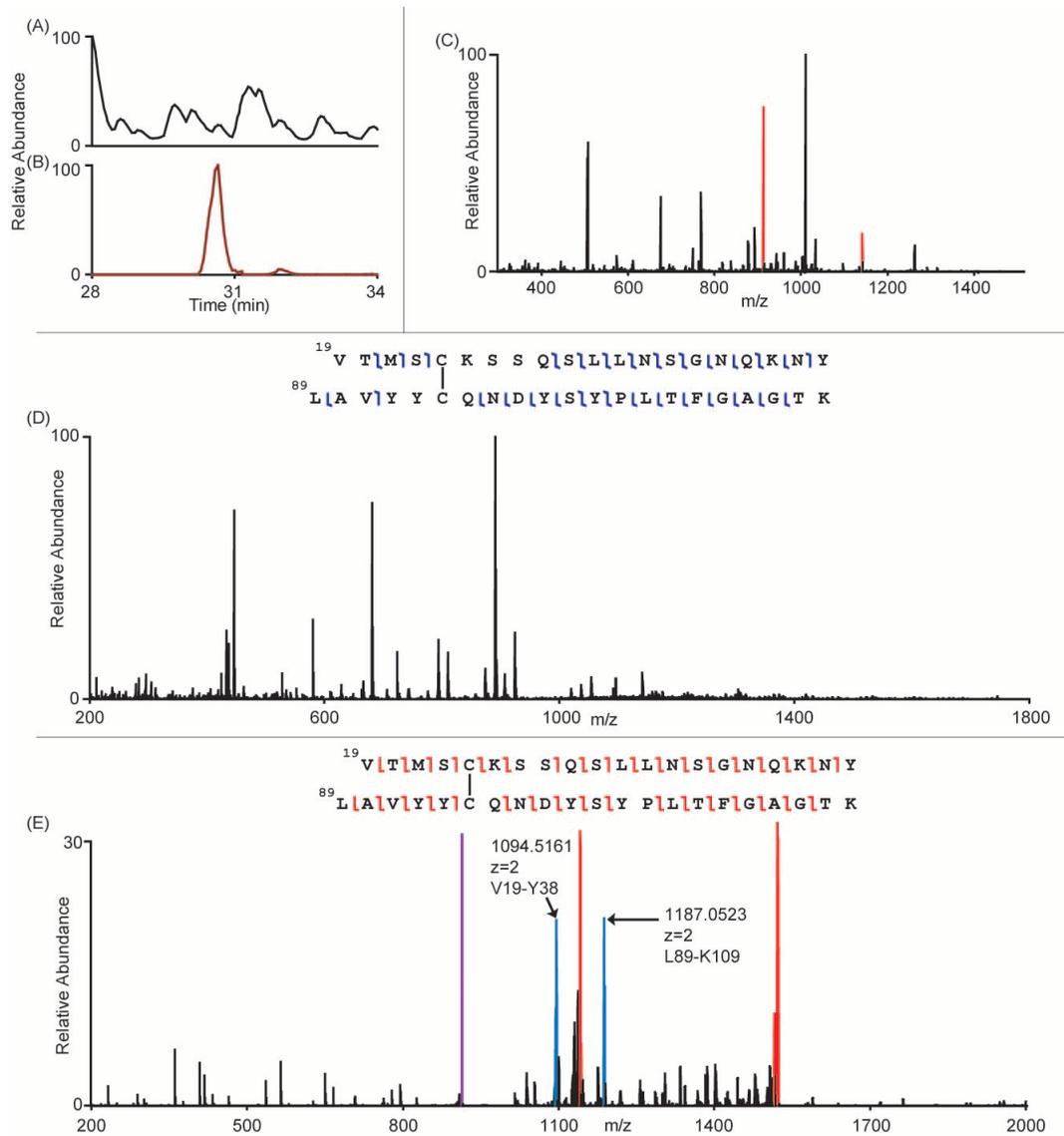


Figure 4.23. Example Disulfide Peptide. (A) Zoom in on the total ion chromatogram from 28-34 minutes. (B) Extracted ion chromatogram of the disulfide bound peptides VTMSCKSSQSLNLSGNQKNY-LAVYYCQNDYSYPLTFGAGTK. (C) MS¹ taken at 30.5 minutes corresponding to the elution of disulfide bound peptide. The red peaks show the z=5 and z=4 charge states of the example peptide. (D) HCD MS² of the z=5 example peptide. Observed fragment ions are shown above the spectrum with blue hash marks corresponding to b- and y-ions. (E) ETD MS² of the z=5 example peptide. The unreacted precursor is in purple while the charge reduced species are in red. The two peaks in blue represent the cleavage of the disulfide bond resulting in the molecular weights of the individual peptides. Observed fragment ions in the spectrum are shown above the spectrum with red hash marks corresponding to c- and z'-ions.

Aspergillopepsin I was successful in generating peptides that covered the entire protein while maintaining the disulfide bond structure. These peptides were dissociated to confidently sequence

the primary structure as well as characterize the disulfide bonds. Figure 4.24 shows 34 peptides that were manually annotated for analysis of these key features. Enzymatic access to some potential cleavage points may have been favored because the BiTE maintained tertiary structure during digestion. One such example is at D₃₄₀/V₃₄₁. No peptide contains both of these residues, implying that this is one of the preferred sites of initial cleavage by Aspergillopepsin I.

```

E L V M T Q S P S S L T V T A G E K V T M S C K S 25
26 S Q S L L N S G N Q K N Y L T W Y Q Q K P G Q P P 50
51 K L L I Y W A S T R E S G V P D R F T G S G S G T 75
76 D F T L T I S S V Q A E D L A V Y Y C Q N D Y S Y 100
101 P L T F G A G T K L E I K G G G G S G G G G S G G 125
126 G G S E V Q L L E Q S G A E L V R P G T S V K I S 150
151 C K A S G Y A F T N Y W L G W V K Q R P G H G L E 175
176 W I G D I F P G S G N I H Y N E K F K G K A T L T 200
201 A D K S S S T A Y M Q L S S L T F E D S A V Y F C 225
226 A R L R N W D E P M D Y W G Q G T T V T V S S G G 250
251 G G S D I V M T Q T P L S L S V T P G Q P A S I S 275
276 C K S S Q S L V H N N G N T Y L S W Y L Q K P G Q 300
301 S P Q S L I Y K V S N R F S G V P D R F S G S G S 325
326 G T D F T L K I S R V E A E D V G V Y Y C G Q G T 350
351 Q Y P F T F G S G T K V E I K G E G T S T G S G G 375
376 S G G S G G A D Q V Q L V E S G G G V V Q P G R S 400
401 L R L S C A A S G F T F T K A W M H W V R Q A P G 425
426 K Q L E W V A Q I K D K S N S Y A T Y Y A D S V K 450
451 G R F T I S R D D S K N T L Y L Q M N S L R A E D 475
476 T A V Y Y C R G V Y Y A L S P F D Y W G Q G T L V 500
501 T V S S H H H H H H

```

Figure 4.24. Manually Annotated Peptide Map from the Nonreduced BiTE Digestion.

Figures 4.25 and 4.26 show the collisional dissociation (both HCD and CAD) and ETD residue cleavages observed, respectively. The collisional fragmentation resulted in 88% residue cleavages while ETD only obtained 76% residue cleavages. This BiTE in particular has stretches of amino

acids that lack a basic residue, preventing the ability to obtain high charge states preferable to ETD. The lack of the charge-adding alkylation agent NAEM further inhibited the ability to increase the charge density. However, when these two fragmentation maps were combined (Figure 4.27), 98% of the cleavages were observed. Only ten out of a possible 509 cleavages were not observed, resulting in near complete sequence coverage. Additionally, all cysteines were accounted for in the disulfide map (Figure 4.28). The four disulfide bonds, unique to each variable region of the BiTE, were observed with no disulfide scrambling.

```

E L[V]M[T]Q[S]P[S]S[L]T[V]T[A]G[E]K[V]T[M]S[C]K[S] 25
26[S]Q[S]L[L]N[S]G[N]Q[K]N[Y]L[T]W[Y]Q[Q]K[P]G[Q]P[P] 50
51[K]L[L]I[Y]W[A]S[T]R[E]S[G]V[P]D[R]F[T]G[S]G[S]G[T] 75
76 D[F]T[L]T[I]S[S]V[Q]A[E] D L[A]V[Y]Y[C]Q[N]D[Y]S[Y]100
101 P[L]T[F]G[A]G[T]K L[E]I[K]G[G]G[G]S[G]G[G]G[S]G[G]125
126[G]G[S]E[V]Q[L]L[E]Q[S]G[A]E[L]V[R]P[G]T[S]V[K]I S 150
151 C K A[S]G[Y]A[F]T[N]Y[W]L[G]W[V]K[Q]R[P]G[H]G[L]E]175
176[W]I[G]D[I]F[P]G[S]G[N]I[H]Y[N]E[K]F[K]G[K]A[T]L]T]200
201 A[D]K[S]S[S]T]A]Y[M]Q[L]S[S]L]T]F]E]D]S]A]V]Y]F]C] 225
226 A R L R[N]W]D]E]P]M]D]Y]W]G]Q]G]T]T]V]T]V]S]S]G]G]250
251[G]G[S]D]I]V]M]T]Q]T]P]L]S]L]S]V]T]P]G]Q]P]A]S]I]S] 275
276 C[K]S[S]Q[S]L[V]H[N]N[G]N]T]Y]L]S]W]Y]L]Q]K]P]G]Q]300
301[S]P]Q[S]L]I]Y]K]V]S]N]R]F]S]G]V]P]D]R]F]S]G]S]G]S] 325
326 G T D[F]T[L]K]I]S]R]V]E]A]E]D]V]G]V]Y]Y]C]G]Q]G]T] 350
351[Q]Y]P]F]T]F]G]S]G]T]K]V]E]I]K]G]E]G]T]S]T]G]S]G]G]375
376[S]G]G]S]G]G]A]D]Q]V]Q]L]V]E]S]G]G]G]V]V]Q]P]G]R]S]400
401[L]R]L]S]C]A]A]S]G]F]T]F]T]K]A]W]M]H]W]V]R]Q]A]P]G]425
426[K]Q]L]E]W]V]A]Q]I]K]D]K]S]N]S]Y]A]T]Y]Y]A]D]S]V]K]450
451[G]R]F]T]I]S]R]D]D]S]K]N]T]L]Y]L]Q]M]N]S]L]R]A]E]D] 475
476 T A V Y Y C R G V Y Y A L S P F D Y W G Q G T L V 500
501[T]V]S]S]H]H]H]H]H]H]

```

Figure 4.25. Collisional Sequence Coverage Obtained from the Analysis of a Nonreduced BiTE. 447/509 residue cleavages were obtained resulting in 88% sequence coverage.

```

E|L|V|M|T|Q|S P|S|S|L|T|V|T|A|G|E|K|V|T|M|S|C|K|S 25
26|S|Q|S|L|L|L|N|S|G|N|Q|K|N|Y|L|T|W|Y|Q|Q|K P|G|Q P P| 50
51|K|L|L I|Y|W|A|S|T|R|E|S|G|V P|D|R|F T G|S G|S|G|T| 75
76|D F T|L|T|I|S|S|V|Q|A|E|D L A|V|Y|Y|C Q|N|D|Y|S|Y 100
101 P|L|T|F|G|A|G|T|K L E|I|K|G|G|G|G|S|G|G|G|S|G|G| 125
126|G|G|S|E|V|Q|L|L|E|Q|S|G|A|E|L|V|R P|G|T|S|V|K I S 150
151 C|K|A|S|G|Y|A|F|T|N|Y|W|L|G|W|V|K|Q|R P|G|H|G|L|E| 175
176|W|I|G|D|I|F P|G|S|G|N|I|H|Y|N|E|K F|K|G|K A T L T 200
201 A D K S S S T A Y M Q|L S S|L T|F|E|D S A|V|Y|F C 225
226 A|R|L|R N W D E P M D Y W G Q G T T V T V S S G G 250
251 G G S D I V M T Q T P L S L S|V|T P G|Q P A|S I|S 275
276 C|K|S|S|Q|S|L|V|H N N G N T Y L|S|W|Y|L|Q|K P|G|Q| 300
301|S P Q|S|L I Y K|V|S|N|R|F|S|G|V P|D|R|F|S|G|S|G|S| 325
326|G|T|D|F|T|L|K|I|S|R|V|E A E D V|G|V|Y|Y C G|Q|G|T| 350
351|Q|Y P|F|T|F|G|S|G|T|K V|E I|K|G|E|G|T|S|T|G|S|G|G| 375
376|S|G|G|S|G|G|A|D|Q|V Q|L V|E|S|G|G|G|V|V|Q P|G|R|S| 400
401|L|R|L|S|C|A|A|S|G|F|T|F|T|K A|W|M|H|W|V|R|Q|A|P G 425
426|K|Q|L|E|W|V|A|Q|I|K|D|K|S|N|S|Y|A|T|Y|Y|A|D|S|V|K| 450
451|G|R|F|T|I|S|R|D|D|S|K N T L Y|L|Q|M|N|S|L|R A E|D| 475
476 T|A|V|Y|Y|C|R|G|V|Y|Y|A|L|S P|F|D|Y|W|G|Q|G|T|L|V| 500
501|T|V|S|S|H|H|H|H|H|H|
    
```

Figure 4.26. ETD Sequence Coverage Obtained from the Analysis of a Nonreduced BiTE. 387/509 residue cleavages were obtained resulting in 76% sequence coverage. Coverage by ETD suffers in the ~200-275 region due to a lack of basic residues. This leads to charge depleted peptides which do not dissociate well by ETD.

```

E L V M T Q S P S S L L T V T A G E K V T M S C K S 25
26 S Q S L L L N S G N Q K N Y L T W Y Q Q K P G Q P P 50
51 K L L I Y W A S T R E S G V P D R F T G S G S G T 75
76 D F T L T I S S V Q A E D L A V Y Y C Q N D Y S Y 100
101 P L T F G A G T K L E I K G G G G S G G G G S G G 125
126 G G S E V Q L L E Q S G A E L V R P G T S V K I S 150
151 C K A S G Y A F T N Y W L G W V K Q R P G H G L E 175
176 W I G D I F P G S G N I H Y N E K F K G K A T L T 200
201 A D K S S S T A Y M Q L S S L T F E D S A V Y F C 225
226 A R L R N W D E P M D Y W G Q G T T V T V S S G G 250
251 G G S D I V M T Q T P L S L S V T P G Q P A S I S 275
276 C K S S Q S L V H N N G N T Y L S W Y L L Q K P G Q 300
301 S P Q S L I Y K V S N R F S G V P D R F S G S G S 325
326 G T D F T L K I S R V E A E D V G V Y Y C G Q G T 350
351 Q Y P F T F G S G T K V E I K G E G T S T G S G G 375
376 S G G S G G A D Q V Q L L V E S G G G V V Q P G R S 400
401 L R L S C A A S G F T F T K A W M H W V R Q A P G 425
426 K Q L E W V A Q I K D K S N S Y A T Y Y A D S V K 450
451 G R F T I S R D D S K N T L Y L Q M N S L R A E D 475
476 T A V Y Y C R G V Y Y A L S P F D Y W G Q G T L V 500
501 T V S S H H H H H H H
    
```

Figure 4.27. Composite Sequence Coverage Obtained by Collisional Fragmentation and ETD from the Analysis of a Nonreduced BiTE. 499/509 residue cleavages were obtained resulting in 98% sequence coverage. The red hash marks signify ETD observed fragment ions while blue hash marks signify collisional (HCD and CAD) observed fragment ions. All 12 CDRs are underlined.

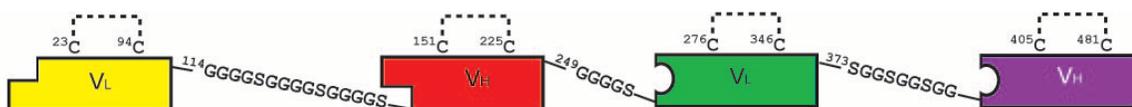


Figure 4.28. Pictorial Representation of Identified Disulfide Bonds. A disulfide bonded pair was present in each region of the BiTE. The linker sequences are noted between each region.

The results presented here show that disulfide localization can be performed in a single analysis while maintaining high sequence coverage of the molecule. However, there was a significant drawback to this experiment: the time needed to search the data. Since Aspergillopepsin I can cleave between any two amino acids, thousands of potential peptides can be generated. The average search time for a fully reduced, nonspecific search is ~1-2 hours. However, adding disulfide bonds drastically changes the necessary search time. Each termini of disulfide bound peptides can be cleaved anywhere, resulting in the exponential increase in possibilities for the search algorithm to consider and requiring an increase in the search time needed.

Controlling the maximum precursor mass to consider limits the amount of time taken to search the data. Table 4.2 gives the search time for three data points when searching this data set. A minimal increase in maximum precursor mass resulted in a substantial increase in search time. Since the search was limited by the molecular weight maximum value, high molecular weight species were unaccounted for in the search. There were peptides observed that were >10 kDa in molecular weight in the analyzed digestion. However, these peptides were unaccounted for in the search algorithm and thus were left unknown without further manual analysis.

Table 4.2. Table of Search Times Required for a Nonspecific Disulfide Analysis

<u>Maximum Precursor Mass (kDa)</u>	<u>Search Time (Hours)</u>
8	6.7
9	22.0
10	60.5

The same issue occurs when adding additional PTMs to the search. Only methionine oxidation and pyro-glutamic acid at the protein N-terminus were considered for the search. Increasing the number of possible combinations of PTMs will also increase the search time needed. Therefore, if a potentially highly modified nonreduced protein is digested and analyzed, then the resulting search should be tailored such that the PTMs can be identified on smaller peptides.

Both the maximum precursor mass and the number of potential PTMs to search greatly inhibit the ability to obtain the full picture of a nonreduced digestion. Searching the data clearly was rate limiting step. However, the ability to obtain near complete sequence coverage and localize the disulfide bonds without scrambling was achieved for the analysis of a novel bsAb. Although the disulfide network analyzed here was simple (Figure 4.27), the results show the current limitations in searching the data. This will be particularly true when the disulfide network of a given protein is more complex, such as bovine trypsinogen (Figure 4.29). The increase in disulfide complexity can potentially lead to more disulfide bound peptides and thus increase the search complexity. Improvements to data search algorithms will greatly impact depth of analysis for a given data set.

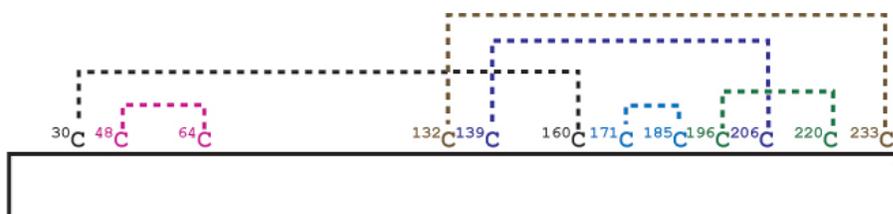


Figure 4.29. Complex Disulfide Bond Network from Bovine Trypsinogen. Unlike in the BiTE, this disulfide network contains nested, crossing, and spatially close disulfide bonds. This can potentially make the resulting search extremely complex.

4.6 Conclusions

Protein concentration affects Aspergillopepsin I digestion, and the ability to harness and manipulate this variable allows control over the performance of the enzyme reactor and the resulting peptides. This is made abundantly clear when digesting and analyzing bsAbs. Larger or smaller peptides may be more desirable depending on experimental needs. Leveraging protein concentration with or without protamine can drastically impact the ability to achieve the desired result.

The concentration of each respective chain of IgG-like bsAbs is effectively half that in comparison to a mAb. This resulted in the further digestion of each chain as demonstrated using a common light chain IgG-like bsAb. Adding protamine to the sample for digestion successfully inhibited Aspergillopepsin I and preserved large peptide products. Since these large products maintained amino acid connectivity, particular peptides can be further investigated for extensive characterization by parallel ion parking, such as peptides containing all 3 CDRs.

Some experiments, such as localized disulfide bonds, require smaller peptides for analysis. Reducing the concentration of a BiTE in the reactor promotes additional cleavages by Aspergillopepsin I. This results in the creation of suitable sized disulfide bound peptides while still allowing high sequence coverage of the entire molecule. Although this was a simple disulfide network, it showed the limitations in the search algorithm. The ability of the search algorithm to quickly and accurately search the data will need to be addressed before moving into more complex disulfide networks in which more than 2 peptides can be disulfide bound.

Tailoring the digestion to better serve experimental needs is readily met by reducing the rate of digestion with protamine or increasing the rate of digestion by decreasing the protein concentration. The enzyme reactor has the powerful ability to obtain near complete sequence coverage for both bsAbs. The results demonstrated here can be further applied to different biologically relevant proteins, such as very large proteins that cannot be accurately identified and dissociated in the mass spectrometer (>30 kDa).

4.7 References

1. Berlot G, Rossini P, Turchet F (2014) Biology of immunoglobulins. *Transl Med @ UniSa* 11:24–27.
2. Janeway C, Travers P, Walport M, Schlomchik MJ (2001) Immunobiology, the Immune System in Health and Disease (Garland, New York).
3. Vidarsson G, Dekkers G, Rispens T (2014) IgG subclasses and allotypes: from structure to effector functions. *Front Immunol* 5:520.
4. Schroeder Jr HW, Cavacini L (2010) Structure and function of immunoglobulins. *J Allergy Clin Immunol* 125(2 Suppl 2):S41–S52.
5. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242.
6. Saphire EO, et al. (2001) Crystal Structure of a Neutralizing Human IgG Against HIV-1: A Template for Vaccine Design. *Science (80-)* 293(5532):1155 LP – 1159.
7. Habegger M, et al. (2014) Assessment of chemical modifications of sites in the CDRs of recombinant antibodies: Susceptibility vs. functionality of critical quality attributes. *MAbs* (Taylor & Francis), pp 327–339.
8. Herold EM, et al. (2017) Determinants of the assembly and function of antibody variable domains. *Sci Rep* 7(1):12276.
9. Igawa T, et al. (2011) Engineering the variable region of therapeutic IgG antibodies. *MAbs* (Taylor & Francis), pp 243–252.
10. Nimmerjahn F, Ravetch J V (2008) Fcγ receptors as regulators of immune responses. *Nat Rev Immunol* 8(1):34.
11. Jennewein MF, Alter G (2017) The Immunoregulatory Roles of Antibody Glycosylation. *Trends Immunol* 38(5):358–372.
12. Booth BJ, et al. (2018) Extending human IgG half-life using structure-guided design. *MAbs* 10(7):1098–1110.
13. Hinton PR, et al. (2006) An Engineered Human IgG1 Antibody with Longer Serum Half-Life. *J Immunol* 176(1):346 LP – 356.
14. Ghetie V, Ward ES (2000) Multiple Roles for the Major Histocompatibility Complex Class I– Related Receptor FcRn. *Annu Rev Immunol* 18(1):739–766.
15. Mimura Y, et al. (2018) Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein Cell* 9(1):47–62.
16. Kaplon H, Reichert JM (2019) Antibodies to watch in 2019. *MAbs* 11(2):219–238.
17. Bondarenko P V, Second TP, Zabrouskov V, Makarov AA, Zhang Z (2009) Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *J Am Soc Mass Spectrom*

- 20(8):1415–1424.
18. Mao Y, Valeja SG, Rouse JC, Hendrickson CL, Marshall AG (2013) Top-Down Structural Analysis of an Intact Monoclonal Antibody by Electron Capture Dissociation-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry. *Anal Chem* 85(9):4239–4246.
 19. von Pawel-Rammingen U, Johansson BP, Björck L (2002) IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G. *EMBO J* 21(7):1607–1615.
 20. Moelleken J, et al. (2017) GingisKHANDTM protease cleavage allows a high-throughput antibody to Fab conversion enabling direct functional assessment during lead identification of human monoclonal and bispecific IgG1 antibodies. *MAbs* 9(7):1076–1087.
 21. Fornelli L, Ayoub D, Aizikov K, Beck A, Tsybin YO (2014) Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. *Anal Chem* 86(6):3005–3012.
 22. He L, et al. (2017) Analysis of monoclonal antibodies in human serum as a model for clinical monoclonal gammopathy by use of 21 tesla FT-ICR top-down and middle-down MS/MS. *J Am Soc Mass Spectrom* 28(5):827–838.
 23. Sen KI, et al. (2017) Automated Antibody De Novo Sequencing and Its Utility in Biopharmaceutical Discovery. *J Am Soc Mass Spectrom* 28(5):803–810.
 24. Choudhary G, Wu S-L, Shieh P, Hancock WS (2003) Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS. *J Proteome Res* 2(1):59–67.
 25. Swaney DL, Wenger CD, Coon JJ (2010) Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* 9(3):1323–1329.
 26. Brinkmann U, Kontermann RE (2017) The making of bispecific antibodies. *MAbs* 9(2):182–212.
 27. Krishnamurthy A, Jimeno A (2018) Bispecific antibodies for cancer therapy: A review. *Pharmacol Ther* 185:122–134.
 28. Thakur A, Huang M, Lum LG (2018) Bispecific antibody based therapeutics: Strengths and challenges. *Blood Rev* 32(4):339–347.
 29. Labrijn AF, Janmaat ML, Reichert JM, Parren PWHI (2019) Bispecific antibodies: a mechanistic review of the pipeline. *Nat Rev Drug Discov* 18:585–608.
 30. Suurs F V., Lub-de Hooge MN, de Vries EGE, de Groot DJA (2019) A review of bispecific antibodies and antibody constructs in oncology and clinical challenges. *Pharmacol Ther* 201:103–119.
 31. Kontermann RE, Brinkmann U (2015) Bispecific antibodies. *Drug Discov Today* 20(7):838–847.
 32. Krah S, Kolmar H, Becker S, Zielonka S (2018) Engineering IgG-Like Bispecific Antibodies—An Overview. *Antibodies* 7(3):28.
 33. Ridgway JBB, Presta LG, Carter P (1996) ‘Knobs-into-holes’ engineering of antibody CH3 domains for heavy chain heterodimerization. *Protein Eng Des Sel* 9(7):617–621.
 34. Klein C, et al. (2012) Progress in overcoming the chain association issue in bispecific heterodimeric IgG antibodies. *MAbs* 4(6):653–663.
 35. Schaefer W, et al. (2011) Immunoglobulin domain crossover as a generic approach for the production of bispecific IgG antibodies. *Proc Natl Acad Sci U S A* 108(27):11187–92.
 36. Cooke HA, et al. (2018) EFab domain substitution as a solution to the light-chain pairing problem of bispecific antibodies. *MAbs* 10(8):1248–1259.
 37. Sampei Z, et al. (2013) Identification and Multidimensional Optimization of an Asymmetric

- Bispecific IgG Antibody Mimicking the Function of Factor VIII Cofactor Activity. *PLoS One* 8(2). doi:10.1371/journal.pone.0057479.
38. Kitazawa T, et al. (2012) A bispecific antibody to factors IXa and X restores factor VIII hemostatic activity in a hemophilia A model. *Nat Med* 18:1570.
 39. Muto A, et al. (2014) Anti-factor IXa/X bispecific antibody ACE910 prevents joint bleeds in a long-term primate model of acquired hemophilia A. *Blood* 124(20):3165 LP – 3171.
 40. Kitazawa T, et al. (2017) Factor VIIIa-mimetic cofactor activity of a bispecific antibody to factors IX/IXa and X/Xa, emicizumab, depends on its ability to bridge the antigens. *Thromb Haemost* 117(07):1348–1357.
 41. Lenting PJ, Denis C V, Christophe OD (2017) Emicizumab, a bispecific antibody recognizing coagulation factors IX and X: how does it actually compare to factor VIII? *Blood* 130(23):2463.
 42. Huehls AM, Coupet TA, Sentman CL (2015) Bispecific T-cell engagers for cancer immunotherapy. *Immunol Cell Biol* 93(3):290–296.
 43. Ellerman D (2019) Bispecific T-cell engagers: Towards understanding variables influencing the in vitro potency and tumor selectivity and their modulation to enhance their efficacy and safety. *Methods* 154:102–117.
 44. Nagorsen D, Baeuerle PA (2011) Immunomodulatory therapy of cancer with T cell-engaging BiTE antibody blinatumomab. *Exp Cell Res* 317(9):1255–1260.
 45. Przepiorka D, et al. (2015) FDA Approval: Blinatumomab. *Clin Cancer Res* 21(18):4035 LP – 4039.
 46. Bargou R, et al. (2008) Tumor Regression in Cancer Patients by Very Low Doses of a T Cell–Engaging Antibody. *Science* (80-) 321(5891):974 LP – 977.
 47. Munz M, Baeuerle PA, Gires O (2009) The Emerging Role of EpCAM in Cancer and Stem Cell Signaling. *Cancer Res* 69(14):5627 LP – 5629.
 48. Zhang L, et al. (2016) Analysis of monoclonal antibody sequence and post-translational modifications by time-controlled proteolysis and tandem mass spectrometry. *Mol Cell Proteomics* 15(4):1479–1488.
 49. Hinkle J, et al. (2019) Unambiguous Sequence Characterization of a Monoclonal Antibody in a Single Analysis Using a Nonspecific Immobilized Enzyme Reactor. *Anal Chem* 91(21):13547–13554.
 50. Senko MW, et al. (2013) Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal Chem* 85(24):11710–11714.
 51. Hinkle J (2019) Antibody Sequence Analysis by Controlled Proteolysis and Ion-Ion Chemistry. Dissertation (University of Virginia, Chemistry - Graduate School of Arts and Sciences, PHD (Doctor of Philosophy), 2019, Charlottesville, VA). Available at: <https://search.lib.virginia.edu/catalog/tb09j630g>.
 52. McLuckey SA, Reid GE, Wells JM (2002) Ion Parking during Ion/Ion Reactions in Electrodynamic Ion Traps. *Anal Chem* 74(2):336–346.
 53. Chrisman PA, Pitteri SJ, McLuckey SA (2005) Parallel Ion Parking: Improving Conversion of Parents to First-Generation Products in Electron Transfer Dissociation. *Anal Chem* 77(10):3411–3414.
 54. Stephenson JL, McLuckey SA (1996) Ion/Ion Reactions in the Gas Phase: Proton Transfer Reactions Involving Multiply-Charged Proteins. *J Am Chem Soc* 118(31):7390–7397.
 55. McLuckey SA, Stephenson JL, Asano KG (1998) Ion/Ion Proton-Transfer Kinetics:

- Implications for Analysis of Ions Derived from Electrospray of Protein Mixtures. *Anal Chem* 70(6):1198–1202.
56. Coon JJ, et al. (2005) Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci U S A* 102(27):9463 LP – 9468.
 57. Anderson LC, et al. (2016) Analyses of histone proteoforms using front-end electron transfer dissociation-enabled Orbitrap instruments. *Mol Cell Proteomics* 15(3):975–988.
 58. Ugrin SA, et al. (2019) Ion-Ion Proton Transfer and Parallel Ion Parking for the Analysis of Mixtures of Intact Proteins on a Modified Orbitrap Mass Analyzer. *J Am Soc Mass Spectrom* 30(10):2163–2173.
 59. Zhang L, Chou CP, Moo-Young M (2011) Disulfide bond formation and its impact on the biological activity and stability of recombinant therapeutic proteins produced by *Escherichia coli* expression system. *Biotechnol Adv* 29(6):923–929.
 60. Tsai PL, Chen S-F, Huang SY (2013) Mass spectrometry-based strategies for protein disulfide bond identification. *Rev Anal Chem* 32(4):257–268.
 61. Lakhub JC, Shipman JT, Desaire H (2018) Recent mass spectrometry-based techniques and considerations for disulfide bond characterization in proteins. *Anal Bioanal Chem* 410(10):2467–2484.
 62. Guan X, Zhang L, Wypych J (2018) Direct mass spectrometric characterization of disulfide linkages. *MAbs* 10(4):572–582.
 63. Mikesh LM, et al. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta - Proteins Proteomics* 1764(12):1811–1822.
 64. Zubarev RA, et al. (1999) Electron Capture Dissociation of Gaseous Multiply-Charged Proteins Is Favored at Disulfide Bonds and Other Sites of High Hydrogen Atom Affinity. *J Am Chem Soc* 121(12):2857–2862.
 65. Gorman JJ, Wallis TP, Pitt JJ (2002) Protein disulfide bond determination by mass spectrometry. *Mass Spectrom Rev* 21(3):183–216.
 66. Berg JM, Tymoczko JL, Stryer L, Stryer L (2007) *Biochemistry* (W.H. Freeman, New York).
 67. Riley NM, et al. (2016) Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. *J Am Soc Mass Spectrom* 27(3):520–531.
 68. Compton PD, Strucl J V, Bai DL, Shabanowitz J, Hunt DF (2012) Optimization of electron transfer dissociation via informed selection of reagents and operating parameters. *Anal Chem* 84(3):1781–1785.
 69. Rose CM, et al. (2015) A calibration routine for efficient ETD in large-scale proteomics. *J Am Soc Mass Spectrom* 26(11):1848–1857.
 70. Earley L, et al. (2013) Front-End Electron Transfer Dissociation: A New Ionization Source. *Anal Chem* 85(17):8385–8390.
 71. Duselis E (2019) Ion/Ion Reactions in Primary Structure Interrogation of Intact Proteins by Mass Spectrometry. Dissertation (University of Virginia, Chemistry - Graduate School of Arts and Sciences, PHD (Doctor of Philosophy), 2019, Charlottesville, VA). Available at: <https://search.lib.virginia.edu/catalog/np193971q>.

Conclusion

This dissertation describes the powerful analytical technique of mass spectrometry for the identification of protein primary structure and post-translation modification. Each chapter discusses different challenges for protein analysis. The mass spectrometer has the sensitivity and selectivity to accurately detect and dissociate complex samples.

Chapter 2 details the successful identification of a novel C-terminal modification of di-aminopropane from highly basic protamines isolated from *Marchantia polymorpha*. These findings provide crucial evidence linking protamines to their histone H1 ancestor. The discovery of a novel C-terminal modification was achieved using the sequential dissociation capability of the instrument. The structure of the modification was confidently determined by targeting an ETD MS² fragment for HCD MS³ analysis. The mechanism of addition and the biological implication of C-terminal di-aminopropane requires further biological investigation. However, it is hypothesized that the generation of di-aminopropane originates from the oxidation of spermidine and spermine, and the addition to the C-terminus removes the only possible negative charge on protamines.

Chapter 3 focuses on the novel enzyme Aspergillopepsin I. This non-specific protease is capable of generating overlapping peptides in seconds. Previous work required relatively high protein concentrations of 0.2 µg/µL to utilize this technology. This concentration may not be feasible for biologically relevant samples. Therefore, the concentration effects were characterized for low concentration samples. Reducing concentrations resulted in an increase in the number of observed cleavages by Aspergillopepsin I. Introduction of protamines from salmon decreased the rate of digestion of the sample through competitive inhibition. This enabled working concentrations available for digestion an order of magnitude lower than previous work.

Sample concentration is only one parameter that affects protein digestion in the Aspergillopepsin I enzyme reactor. Other parameters, such as the dynamic range of the system, may be studied in the future. This will be necessary for the detection of post-translational modifications. Proteins can be modified with PTMs such as phosphorylation, acetylation, methylation, etc. (1, 2). These modifications are typically low in abundance in comparison to the

unmodified protein (3). Specific protease digestions ensure the modification is contained on a few predictable peptides. Aspergillopepsin I may be a viable alternative if a specific protease fails to generate the peptide(s) of interest. However, significantly more modified peptides would be generated due to the lack of specificity of Aspergillopepsin I. Determining the limit of detection for low abundance modified peptides is necessary to confirm that PTM localization can be carried out using Aspergillopepsin I.

Chapter 4 describes an applied usage of the protocol detailed in Chapter 3. Protamine is used to mitigate the effects of low sample concentration in a bispecific antibody. The first section confirmed the results of Chapter 3 in regard to sample concentration. The common light chain, IgG-like bsAb contained heavy chains at half the concentration relative to the identical light chains. Initial digestion by Aspergillopepsin I showed a high abundance of low molecular weight peptides. Using protamines as a competitive inhibitor reduced the number of observed cleavages to increase the abundance of large molecular weight products. The sequence was confirmed using a series of 58 overlapping peptides with 99.4% observed fragment ions.

A single digestion and a single LC-MS/MS analysis greatly reduced the amount of time needed to confirm the sequence of a bsAb in comparison to multiple, overnight specific protease digestions and multiple LC-MS/MS analyses. This rapid identification can benefit antibodies very early in development. For instance, the primary structure of an antibody generated by a new DNA construct could be confirmed. Additionally, laboratories could confirm the primary structure of antibodies in use for various protocols.

The implementation of parallel ion parking followed by IIPT is demonstrated in the second section of Chapter 4. Ideal peptides for analysis contained all 3 CDRs to maintain connectivity. This guaranteed that related CDRs remain attached to each other for analysis and interpretation. A reduced digestion time of ~800 ms with competitive inhibition by protamines produced the target peptides at sufficient abundance for analysis. Although standard ETD reaction times can provide initial identification, parallel ion parking generates high sequence coverage by maintaining first generation fragment ions.

This technique can be extended to the analysis of large proteins (>30 kDa). Intact analysis of proteins in this molecular weight range is difficult due to low S/N ratios (4). Limiting the digestion time (< 750 ms) with protamine inhibition can generate large peptides that maintain a significant portion of the protein suited for parallel ion parking. For example, a 100 kDa protein can be digested into four ~25 kDa peptides. Parallel ion parking can then be used to obtain extensive sequence information with minimal disruption to the primary structure of these four peptides.

The disulfide network of a BiTE is characterized in the final section of Chapter 4 with no disulfide scrambling. Previous disulfide localization experiments required multiple digestion times up to 740 s to obtain peptides containing only a single disulfide bond (5). It was demonstrated in Chapter 3 that reducing the protein concentration for digestion by Aspergillopepsin I will increase the number of observed cleavages. Therefore, additional cleavages were induced by digesting the BiTE at a concentration of 0.05 $\mu\text{g}/\mu\text{L}$ for ~2 s. A high number of low molecular weight peptides with up to one disulfide bond were present in the digestion. The sequence was confirmed with a series of 34 overlapping peptides with 98% observed fragment ions and complete disulfide bond localization. Commercial search algorithms limit this methodology. Search times of over 60 hours were needed to obtain a majority of the information presented in the analysis. Complex disulfide networks require advancements in search algorithms to reduce the necessary time of completion.

The use of Aspergillopepsin I would advance the analysis of membrane proteins. These proteins represent roughly 25% of the entire proteome and over 60% of therapeutic targets (6, 7). However, sequencing membrane proteins is difficult. Membrane proteins lack basic residues preventing effective trypsin digestions (8). These residues are typically located in the extracellular or intracellular domains. The resulting tryptic peptides would ultimately generate peptides containing primarily the hydrophobic transmembrane domain. These peptides are difficult to solubilize which lead to missed peptides when analyzed by LC-MS/MS.

Nonspecific proteases have been used previously for membrane protein analysis. Wu et al. developed methodology to “shave” the extracellular and intracellular regions using the nonspecific protease Proteinase K at high pH (9). The high pH disrupts the spherical membrane of cells to generate membrane sheets. Maintaining some membrane structure provides solubility for the

membrane proteins without harsh, denaturing detergents. Proteinase K, which is functional up to pH 12, can then cleave the exposed residues (10). However, this methodology cannot determine the transmembrane domains. Aspergillopepsin I should be advantageous for determining the entire membrane protein. A percentage of membrane proteins are effectively solubilized in 8M urea and acidic conditions suited for Aspergillopepsin I digestion. However, solubility of membrane proteins vary in 8 M urea (6, 11). Finding a buffer that solubilizes all membrane proteins without disrupting Aspergillopepsin I will enhance the analysis of entire membrane proteins by mass spectrometry. This would provide valuable sequence information of membrane proteins and determine if any modifications occur within the transmembrane domain.

Additionally, alternatives to separation methods prior to mass spectrometric analysis will need to be explored. The highly hydrophobic peptides generated from the transmembrane domain are difficult to elute from conventional, long-alkyl chain reversed phase columns, such as C₁₈. An alternative reversed phase stationary phase consists of short-alkyl chains, such as C₃. Additionally, different separation techniques can be explored, such as size-exclusion chromatography and hydrophilic interaction liquid chromatography (HILIC).

To summarize, this research shows the strength and capabilities of mass spectrometric based analyses to gather valuable information from complex samples. The discovery of novel PTMs drives further experimentation into their biological relevance. Novel sample preparation techniques push the boundaries of the proteins accessible for analysis. The research presented here shows the untapped potentials for existing mass spectrometric technologies and that an even deeper knowledge of proteins can be obtained.

References

1. Nørregaard Jensen O (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* 8(1):33–41.
2. Mann M, Jensen ON (2003) Proteomic analysis of post-translational modifications. *Nat Biotechnol* 21(3):255–261.
3. Olsen J V, Mann M (2013) Status of Large-scale Analysis of Post-translational Modifications by Mass Spectrometry. *Mol & Cell Proteomics* 12(12):3444 LP – 3452.
4. Compton PD, Zamdborg L, Thomas PM, Kelleher NL (2011) On the scalability and requirements of whole protein mass spectrometry. *Anal Chem* 83(17):6868–6874.
5. Zhang L, et al. (2016) Analysis of monoclonal antibody sequence and post-translational

- modifications by time-controlled proteolysis and tandem mass spectrometry. *Mol Cell Proteomics* 15(4):1479–1488.
6. Prados-Rosales RC, et al. (2019) Strategies for Membrane Protein Analysis by Mass Spectrometry BT - *Advancements of Mass Spectrometry in Biomedical Research*. eds Woods AG, Darie CC (Springer International Publishing, Cham), pp 289–298.
 7. Calabrese AN, Radford SE (2018) Mass spectrometry-enabled structural biology of membrane proteins. *Methods* 147:187–205.
 8. Wu CC, Yates JR (2003) The application of mass spectrometry to membrane proteomics. *Nat Biotechnol* 21(3):262–267.
 9. Wu CC, MacCoss MJ, Howell KE, Yates JR (2003) A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* 21(5):532–538.
 10. EBELING W, et al. (1974) Proteinase K from *Tritirachium album* Limber. *Eur J Biochem* 47(1):91–97.
 11. Plucinsky SM, Root KT, Glover KJ (2018) Efficient solubilization and purification of highly insoluble membrane proteins expressed as inclusion bodies using perfluorooctanoic acid. *Protein Expr Purif* 143:34–37.