

**Antibody Sequence Analysis by Controlled Proteolysis and Ion-Ion  
Chemistry**

Joshua David Hinkle  
Knoxville, Tennessee

B.S. Chemistry, Carson-Newman University, 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, 2019

## Acknowledgements

I would first like to acknowledge Professor Donald F. Hunt and Dr. Jeffrey Shabanowitz for both their help and encouragement over the course of my graduate studies. They have both been invaluable to my growth as a researcher and I would certainly not be where I am without them. I am incredibly grateful for being given the opportunity to work in this lab, and cannot thank them enough.

I would also like to thank Drs. John E. P. Syka and Christopher Mullen for their suggestions and assistance on this work. Their help both in the form of suggestions and technical assistance have been incredibly valuable over the course of this work.

Further, all of my lab mates who have been both my close friends and collaborators over the course of my graduate studies. I particularly appreciate the contributions of Robert D'Ippolito, Emily Zahn, and Elizabeth Duselis who have been particularly involved in working on these projects with me, either through a cooperating or through providing improvements to the analytical methods used.

Finally, I would like to thank my family. They have always stood beside me and supported me every step of the way through my education. I love all of you and could not have done it without you.

## Table of Contents

Acknowledgements.....	i
Table of Contents.....	ii
List of Figures.....	iv
Abbreviations.....	vii
Abstract.....	x
Introduction.....	1
Introduction to Protein Mass Spectrometry .....	2
Liquid Chromatography.....	2
Electrospray ionization .....	4
Overview of Mass Analyzers.....	8
Quadrupoles.....	8
Ion Traps.....	13
Orbitraps .....	17
Ion Dissociations.....	20
Collisional Dissociation.....	20
Electron Transfer Dissociation .....	24
Protein Sequencing .....	29
Modes of Sample Preparation.....	33
Antibodies.....	36
References.....	40
ETD Parallel Ion Parking of Antibody Subunits .....	46
Introduction.....	46
Background.....	47
Quadrupolar Fields .....	47
Ion Stability .....	50
Secular Frequency and Resonant Excitation.....	54
Gas phase reaction kinetics.....	56
Parallel Ion Parking .....	59
Materials and Instrumentation .....	62
Methods .....	64
Apomyoglobin Standard Preparation and Analysis.....	64
NIST mAb Preparation and Analysis .....	65
Results and Discussion .....	66
Parking Parameter Evaluation with Apomyoglobin.....	66
NISTmab Intact Subunit Analysis.....	70

References .....	81
ETD Parallel Ion Parking Coupled to Size-Controlled Proteolysis for Antibody CDR Characterization .....	85
Introduction.....	85
Background.....	86
Conventional Strategies for Generating Larger Peptide .....	86
Aspergillopepsin .....	88
Immobilized Reactor Digestion Complexity .....	92
Diffusion Limitations.....	93
Multipath Diffusion .....	95
Materials and Instrumentation .....	97
Methods .....	98
Aspergillopepsin I Enzyme Reactor Fabrication.....	98
Apomyoglobin Standard Digestions.....	100
Antibody Sample Preparation.....	103
Antibody Digestion Analysis.....	105
Results and Discussion .....	106
Enzyme Reactor Scaffold Modifications.....	106
Antibody Variable Region Analysis.....	110
References.....	117
Unambiguous Antibody Sequence Determination by Shotgun Decision Tree	
Methodology.....	121
Introduction.....	121
Background.....	122
Complementarity of Fragmentation Techniques .....	122
Instrumental Advantages of Orbitrap Fusion .....	124
Materials and Instrumentation .....	127
Methods .....	129
Antibody sample preparation.....	129
Decision-Tree Based Mass Spectrometry.....	130
Results and Discussion .....	133
Unambiguous Antibody Sequence Analysis .....	133
References.....	145
Conclusions.....	148
References.....	153
Appendix.....	154

## List of Figures

### Chapter 1:

Figure 1.1: Proposed Mechanisms for ESI Ionization

Figure 1.2: Applied Voltages to a Quadrupole Mass Filter.

Figure 1.3: Quadrupole Scan Line

Figure 1.5: Schematic of 2D and 3D Ion Traps

Figure 1.6: Voltage application configuration on a linear ion trap

Figure 1.7 Schematic representation of Orbitrap ion injection and analysis.

Figure 1.8: Mechanism for CA+D fragmentation.

Figure 1.9: Mechanism for ETD fragmentation.

Figure 1.10: Front end ETD Injection Scheme on a Linear Ion Trap

Figure 1.11: Nomenclature for denoting peptide fragment ions.

Figure 1.12: Representation of Peptide Sequencing

Figure 1.13: Improved Sequence Coverage by Multiple Fragmentation Methods

Figure 1.14: Antibody Structure Schematic

### Chapter 2: Ion Parking

Figure 2.1: Stability diagram for x-stable and y-stable ion trajectories in Qx space

Figure 2.2: Depiction of sequential ETD reactions and internal fragments

Figure 2.3: Schematic Fourier transform of an ETD parking waveform

Figure 2.4: Depiction of slowed ion kinetics via ion parking with respect to ion stability in a linear ion trap.

Figure 2.5: Instrument schematic for FETD enabled Orbitrap Elite

Figure 2.6: ETD parking only Spectrum of Apomyoglobin<sup>+25</sup>

Figure 2.7: ETD parking/IIPT Spectrum of Apomyoglobin<sup>+25</sup>

Figure 2.8: Condensed Sequence Coverage of pipETD/IIPT of Apomyoglobin<sup>+25</sup>

Figure 2.9: Positive ESI charge-state distribution comparison of unreduced vs reduced/alkylated antibody F(ab')<sub>2</sub> subunits

Figure 2.10: Total Ion Chromatogram of from NIST Subunit Analysis

Figure 2.11: MS1 spectra of NIST Subunits

Figure 2.12: ETD Parking/IIPT MS2 Fragmentation Spectra of NIST Subunits

Figure 2.13: Condensed Sequence Coverage derived from ETD Parking/IIPT Spectra of NIST Subunits

### Chapter 3: Enzyme Reactor Parking

Figure 3.1: Crystal Structure of Aspergillopepsin

Figure 3.2: Schematic representation of controlled digestion by Aspergillopepsin I Enzyme Reactor

Figure 3.3: Representation of Pore Diffusion

Figure 3.4: Representation of Multipath Diffusion (Eddy Diffusion)

Figure 3.5: Mechanism for surface Conjugation by Reductive Amination

Figure 3.6: Comparison of total ion chromatograms from 0.70 ms 20 μm scaffold and 0.53 ms 1 μm scaffold digestion.

Figure 3.7: Comparison of peptide abundances by size in 20 um vs 1 um scaffold Apomyoglobin digestions

Figure 3.8: Total ion chromatogram of ~0.9s adalimumab digestion

Figure 3.10: MS1 spectra of adalimumab variable region peptides

Figure 3.11: Parked ETD/IIPT MS2 fragmentation spectra of the adalimumab variable region peptides

Figure 3.12: Sequence coverage from ETD Parking/IIPT spectra of adalimumab variable region peptides.

Figure 3.13: Aspergillopepsin I cleavage site sequence homology comparison of several therapeutic antibodies.

### Chapter 4: Enzyme Reactor Shotgun Method

Figure 4.1: Orbitrap Fusion Schematic

Figure 4.2 Ion injection schemes for standard vs high capacity ETD

Figure 4.3: Orbitrap Fusion Decision Tree Method

Figure 4.4: Total ion chromatogram of ~0.9s enzyme reactor digestion of adalimumab.

Figure 4.5: Map of total peptide abundance from adalimumab enzyme reactor digestion and the peptide subset selected for manual sequence evaluation

Figure 4.6: Composite sequence coverage derived from the annotated subset of peptides from adalimumab subunits

Figure 4.7: Structure for common antibody glycans

Figure 4.8: Extracted ion chromatogram and fragmentation spectra of lysine clipped and unclipped peptides identified within the analysis.

## Abbreviations

μ - micro (1x10<sup>-6</sup>)

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<sup>-15</sup>)

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 ( $1 \times 10^{-3}$ )

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 ( $1 \times 10^{-9}$ )

N, Asn - asparagines

o.d. outer diameter

p - pico ( $1 \times 10^{-12}$ )

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<sub>H</sub> – Heavy Chain Variable Region

V<sub>L</sub> – Light Chain Variable Region

W, Trp tryptophan

Y, Tyr tyrosine

## **Abstract**

The study of proteins represents an invaluable tool for the characterization of biological systems. As proteins are one of the primary functional units of the cell, serving to transmit signals, catalyze reactions, and more, possessing the tools to effectively study them is of particular importance for studying biological systems. Antibodies in particular are a product of significant interest for in-depth characterization given the strict tolerances required when using them as antibody therapeutics. Consequently, better tools to more rapidly and thoroughly analyze these molecules are deeply needed.

Mass spectrometry has become an invaluable tool for the analysis of proteins, achieving a level of precision and sensitivity largely unrivaled by competing methods. However, despite the significant strides made in recent years, several key limitations still face modern mass spectrometric methods. Two of the major limitations still facing the proteomics community are that of complete peptide coverage following proteolytic digestion and complete fragmentation coverage following peptide dissociation. Both of these issues limit the information that can be obtained in a given experiment. To address these concerns, this dissertation presents the combination of precisely controlled digestion and ion-ion reaction strategies for improved sample preparation and fragmentation, respectively. As shown here, these results achieve significantly improve on the depth of analysis for therapeutic monoclonal antibodies, though their applicability likely extends to many other systems.

## **Introduction to the Dissertation – Basics of Protein Mass Spectrometry and Antibodies**

### **1.1 Introduction**

Proteins represent one of the key players in all biological systems across the board. As they drive a plethora of functions within a cell, ranging from cell signaling to energy metabolism (1), the ability to effectively analyze proteins is a necessary pursuit to ultimately understand all biological systems. A protein's functionality is inextricably linked to its chemical structure, and as such it is imperative to be able to analyze proteins for their precise chemical compositions. An entire class of research known as proteomics has developed toward pursuing this endeavor which aims to precisely identify proteins on the basis of their structural features (2).

Fundamentally, proteins are a class of biopolymer composed of repeating  $\alpha$ -amino acids as its foundational substrate (1). These compounds share a common chemical structure. They all consist of both a carboxyl and an amino group which are separated by a single carbon, namely the alpha carbon. Each alpha carbon also possesses a unique side-chain which is ultimately responsible for the chemical diversity between proteins (1). These side chains are typically one of 20 different chemical moieties which together represent the 20 canonical proteinogenic amino acids (1). To form a peptide or protein, many different amino acids are connected to one another by a dehydration reaction between the carboxyl group of one amino acid and the amino group of the second, forming an amide or peptide bond. In order to form a complete protein, hundreds to

thousands of amino acids may be strung together to form the intact molecule. As such, they are generally quite complex and difficult to analyze.

Given both their importance and structural complexity, it becomes necessary to pursue the use of efficacious analytical strategies in order to achieve appropriate characterization of these molecules. Such technology must both enable detailed characterization in order to map the complexity of the molecule, while also being sufficiently sensitive to analyze low level species. Fortunately, protein analysis by mass spectrometry fulfils both of these criteria, and as such are a powerful tool when used to solve biological problems.

## **1.2 Introduction to Protein Mass Spectrometry**

Fundamentally, mass spectrometers use a variety of techniques which measure the mass-to-charge ratio ( $m/z$ ) of ionized analytes (3). This process facilitates their identification both by distinguishing the molecule on the basis of its molecular mass as well as the mass of its constitutive parts following some manner of dissociation (2, 4). This section will discuss the basic operating principles for the most common techniques used in protein characterization by mass spectrometry as they pertain to this work

### ***Liquid Chromatography***

Prior to analysis within a mass spectrometer, proteins are very commonly separated on the basis of their chemical characteristics (2, 5). This applies an additional level of orthogonal temporal separation which strongly complements the mass-based separation enabled by the mass spectrometer, adding an additional level of resolution and thereby improving the sensitivity of the technique; if all proteins or peptides in a sample

were injected at the same time, it is unlikely that all of them could be appropriately resolved from one another nor occur at similar enough intensities that they could all be observed in the same spectrum (2, 6). Although many types of separation techniques are employed in principle, the most commonly used strategy is reverse-phase liquid chromatography.

Reverse-phase liquid chromatography, condensed to only liquid chromatography (LC) for the remainder of this work, achieves protein separation on the basis of their hydrophobicity. Peptides or proteins dissolved in an aqueous solution are passed through a column packed with some kind of hydrophobic stationary phase(7); this phase can range from silica particles with aliphatic carbon chains chemically attached to the surface of particles constructed from hydrophobic polymers like polystyrene(8, 9). Due to their predominantly hydrophobic properties, the protein material then exhibits a higher affinity for the hydrophobic surface than the aqueous solution and almost entirely partitions out of the aqueous solution and onto the stationary phase(7). To liberate these peptides, the organic concentration can be gradually increased over time, decreasing the affinity to the stationary phase as the mobile phase becomes less aqueous(10). The protein material will then more preferentially partition into the mobile phase, causing it to elute from the column, and this partitioning occurs at lower organic concentrations for proteins containing more charged or polar side amino acids. As a result, these more hydrophilic proteins will elute sooner than their more hydrophobic counterparts, causing them to elute from the column earlier and separating the two species in time.

Chromatographic separation in this work is typically performed inside of a narrow-bore fused silica microcapillary packed with a variety of different packing

materials(11). Stationary phase particles are pressure loaded into these capillaries and the particles are held in place by a silicate frit to form the packed bed(11, 12). Proteinaceous solutions may then be pressure loaded through the column, typically in an acidic, aqueous solution, in order to load the peptides in the column, after which they are gradient eluted with increasing levels of acetonitrile and isopropanol.

While this kind of separation can indeed precede offline (13), one of the primary motivating factors for using liquid chromatography is the simplicity of performing the separation in-line with the mass spectrometer (14). Common solvents used for LC consist of chemicals like water, acetonitrile, isopropanol, and volatile acids like formic acid or acetic acid. As all these components are sufficiently volatile, peptides may be separated in-line and ionized immediately following elution using the most common mode of ionization, electrospray ionization.

### ***Electrospray Ionization***

As mass spectrometers are ultimately analyzing the properties of ionized molecules, analytes must first be converted into an ionized gas prior to analysis. The most commonly used ionization technique for analyzing proteins and peptides is electrospray ionization. Using this strategy, peptides and proteins can be ionized directly out of volatile aqueous and organic solutions, allowing the technique to be easily coupled to powerful separation strategies like reverse phase chromatography or capillary electrophoresis (15).

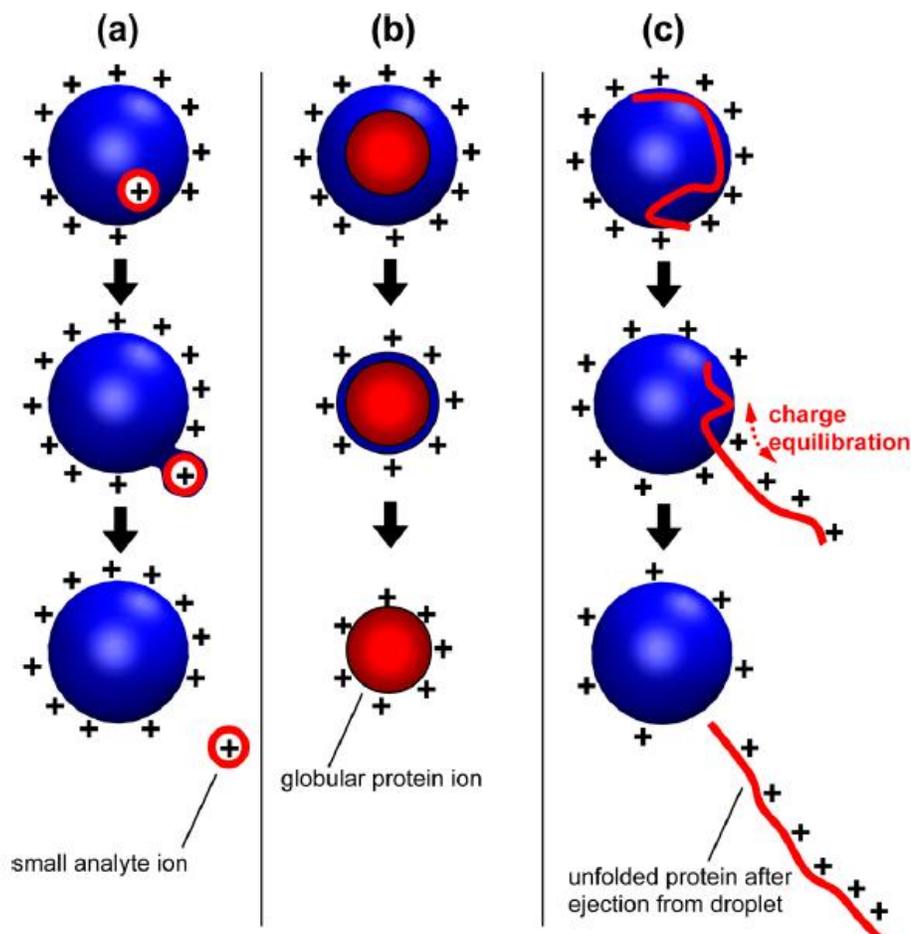
Electrospray ionization involves the application of a high voltage to a liquid analyte in order to drive the ionization process. Typically, this involves the application of

around 2-5 kV to the tip of a narrow bore capillary, often with a narrowed tip pulled at the emitting end of the capillary (16). Sample can be flowed through this capillary at a low flow rate, often ranging from 60 nL/min to 10  $\mu$ L/min (16, 17). The application of the high voltage to the assembly causes a build-up of positive charges at the tip of the emitter. The repulsion resulting from this high concentration of charge distorts the liquid at the tip, creating what is known as a Taylor cone (18), and ions are emitted from the end of the cone in a fine plume (16, 17).

After being emitted from the tip of the Taylor cone, droplets will enter the inlet of the instrument through a heated capillary. The elevated temperature of this capillary drives evaporation of excess solvent from these droplets. As the droplet size decreases, the charges within the droplet become increasingly concentrated and repel each other more strongly. Eventually, the droplet will reach its Rayleigh limit at which point the repulsive effects of the droplet charges outweigh the surface tension of the droplet, and the droplet will erupt into many smaller droplets to disperse the charge (19). These droplets may be orders of magnitude smaller than the initial droplet and will undergo their own process of evaporation and fission until ionization occurs (17).

The actual process of analyte ionization is thought to occur through a number of different mechanisms. They are known as the ion evaporation model (IEM) (20), the charged residue model (CRM) (21, 22), and the chain ejection model (CEM) (17, 23) and are visually depicted in Figure 1.1. The IEM proposes that peptides and small molecules, being fairly surface active agents, will tend to reside at the surface of the droplet. Once the Coulombic repulsion gets high enough, they will be sufficiently repelled off of the droplet and “evaporate,” becoming an independently ionized molecule (20). The chain

ejection model is essentially a refinement of this model for large, unfolded proteins, proposing a stepwise ejection of the long protein chain due to the same Coulombic repulsion that causes small molecules and peptides to evaporate in the IEM (17). These proteins will also tend to reside at the gas-liquid interface because after being denatured, their hydrophobic cores will tend to be solvent exposed. Conversely, the CRM proposes that charged solvent is instead removed from the droplet via the IEM mechanism until the analyte is the only thing remaining in the droplet (21, 22). As the droplet evaporates completely, the charges on the last remaining solvent molecules are instead transferred to the analyte, resulting in an ionized molecule. This is thought to be the primary mechanism for the ionization of proteins in their native conformations because they tend to be more hydrophilic and therefore less surface active.



**Figure 1.1: Various proposed mechanisms for electrospray ionization.** A) depicts the ion evaporation model wherein a small ion evaporates from the surface of the droplet, as is thought to primarily happen to peptides. B) Depicts the charge reduction model, wherein the solvent simply evaporates away to leave residual charge on the analyte, is thought to primarily occur for folded proteins. C) Depicts the chain ejection model wherein a long polymeric chain is sequentially ejected from the droplet as the charge equilibrium of the droplet vs analyte shifts, as is thought to primarily occur in the case of denatured proteins. Adapted from (17)

In the context of mass spectrometric analysis, ESI is particularly useful because it can be coupled directly to peptides being separated chromatographically given that the solvents are compatible with the ESI process (15). As peptides are separated along the length of the LC column, they will eventually progress to the tip of the ESI needle, where they will be ionized in the order of their elution. This allows peptides to be separated on

the basis of their hydrophobicity before being discriminated within the mass spectrometer on the basis of their mass (and constitutive fragment masses). This strong complementarity is one of the primary reasons that LC-MS analysis is often the method of choice for peptide and protein analyses (3).

### **1.3 Overview of Mass Analyzers**

Having converted any analyte proteins into an ionized gas, it becomes necessary to introduce a kind of instrument capable of determining the mass of the ion or its constituent fragment. A wide variety of techniques have been employed to measure the mass of ions, but only those relevant to the work presented in this dissertation will be described here, namely the quadrupole, ion trap, and Orbitrap mass analyzers.

#### *Quadrupoles*

Quadrupole and ion trap mass analyzers operate somewhat similarly in that they both perform mass analysis by taking advantage of mass selective stability (in the case of the quadrupole) or instability (in the case of the ion trap) within their quadrupole fields. The operational theory of these quadrupole fields, particularly in the context of ion traps, will be discussed in more detail in Section 2.2, but a simplified discussion is relevant in this context.

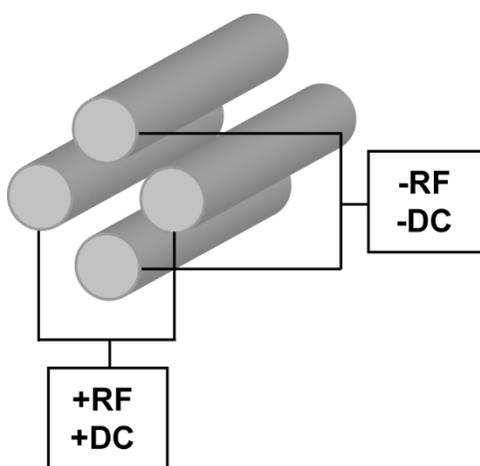
Quadrupole-based mass analyzers, both quadrupole mass filters as well as ion traps, fundamentally manipulate ions on the basis of dynamically oscillating electric fields which keep ions in a constant state of flux in order to confine their position to a particular region (24–26). By rapidly inverting the potentials on the device, ions oscillate

between being attracted to and repelled by the different electrodes in the instrument; this in effect keeps them confined toward the center of the device.

Quadrupole mass analyzers generally consist of four cylindrical rods placed in parallel to generate a close approximation to an ideal quadrupole field (25, 27). Although in principle the rod shape would need to be hyperbolic to generate a quadrupole field without any higher order (e.g. octopolar) field components, a close approximation is achieved through carefully aligned round rods such that the ratio of the rod diameter ( $r$ ) and the field radius ( $r_0$ ) approximates the equation (27):

$$r/r_0 = 1.128$$

The quadrupole field is generated by a  $\sim 1.145$  MHz radiofrequency (RF) alternating current which is applied  $180^\circ$  out of phase on the x-rods as compared to the y-rods of the quadrupole, as depicted in Figure 1.2. This results in the rods remaining at equivalent but opposing polarities during the operation of the device. Further, an additional direct current (DC) component can be applied to these rods, creating a static potential offset between the set of rods.



**Figure 1.2: Applied Voltages to a Quadrupole Mass Filter.** The opposite polarity RF (for trapping) and DC (for selective instability) are both applied to the same sets of rods.

The two primary experimental features that are modified when operating a quadrupole are the amplitude of the drive RF and the amplitude of the additional DC offset (24, 25). When only serving to simply transmit ions, quadrupoles are typically operated over a variety of RF amplitudes but maintain a DC offset of 0V. The range of RF amplitudes that will effectively confine an ion is based in part on its  $m/z$ ; an intuitive way to envision this is that the RF amplitude needs to be high enough that to effectively repel an ion within the field given that its momentum increases with mass. At the same time, it must be low enough that exceptionally small ions are pushed too hard by the RF potential, out of the bounds of the quadrupole field. Ions meeting both of these conditions will be stably transmitted from one end of the quadrupole to the other, where they may be detected or passed further into the instrument.

In order to more selectively transmit ions, the DC offset placed between the two sets of rods is increased. This, in effect, acts as a destabilizing force which causes otherwise stable ions to adopt unstable trajectories. Most typically, the y-rods are held at a slightly more negative potential and x-rods at a positive potential, although this decision is largely arbitrary (24). Large ions, as noted previously, are very weakly confined in the quadrupole field owing to their increased size. The application of the DC offset ultimately destabilizes (otherwise stable) large ions in the y-axis as a result of the negative attractive potential (28). While the x-rods will act as a repulsive force, causing the ion to oscillate between them in that dimension, the y-rods possess a negative, attractive potential which will then pull these larger ions with a greater force than they are confined by the RF field. Conversely, for ions that are nearing the border of being too small, the positive x-rod potential destabilizes them by providing extra repulsion

reinforcing the already substantial magnitude of the trapping RF (28). This leads to the large RF amplitude more easily driving these ions into the x-rods as the DC offset is increased.

In this way, increasing the DC offset effectively acts as a band-pass filter, narrowing the range of masses transmitted through the quadrupole by destabilization in one of the two axes until ultimately a very narrow range of  $m/z$  passes the full length. This can be employed as a mass filter, isolating a very narrow  $m/z$  window for further experimentation (e.g. fragmentation). However, it may also be employed as an independent mass analyzer (29). Sweeping both the RF and DC components at a consistent ratio allows a narrow  $m/z$  transmission window to be effectively swept across the mass range. As the RF amplitude is increased in a manner that would destabilize certain  $m/z$  ions, the DC offset is equivalently increased to compensate, causing the  $m/z$  value of the ions being stably transmitted to increase with time (24). These ions will then sequentially hit a detector at the end of the quadrupole, and the resulting signal spikes can be correlated with the quadrupole conditions at the time of detection to determine all of the  $m/z$  being generated from an ion source at a given time. This is often referred to as a scan line and can be represented with respect to the stable RF and DC voltage amplitudes at which different  $m/z$  ions are stable, as depicted in Figure 1.3.

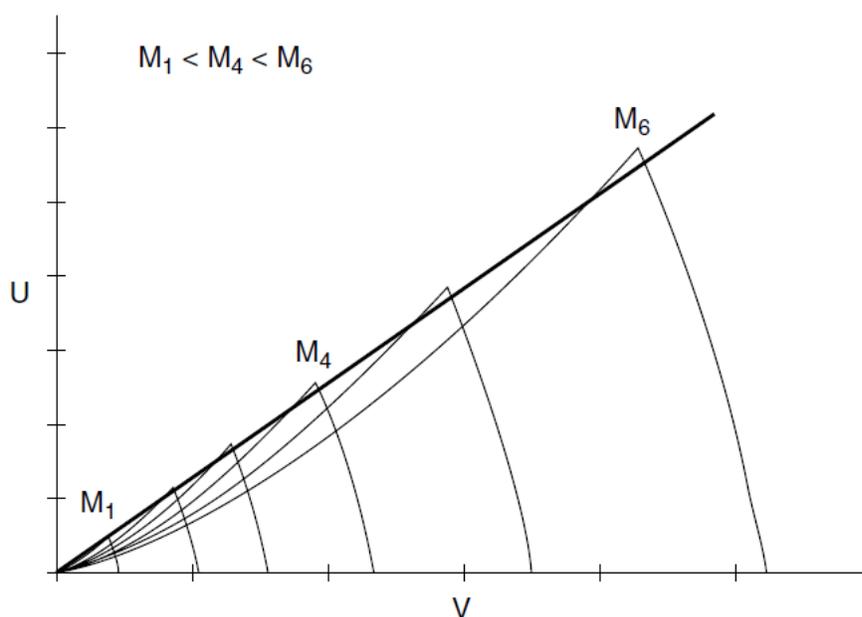


Figure 1.3: Scan line used for quadrupole mass analysis. Each labeled segment represents the stable conditions at which an ion may reside with respect to RF ( $V$ ) and DC ( $U$ ) amplitudes. Notably, by ramping these values at a fixed ratio, sequentially increasing masses can be selectively transmitted through the instrument. Adapted from (24).

The nature of the scan modality used in quadrupoles offers them some significant advantages. For instance, because they are commonly paired with electron multipliers as end-point detectors, quadrupoles are typically very sensitive instruments (30). Further, they offer very robust quantitation by making effectively instantaneous measurements of ion flux so long as a particular ion of interest continues to be allowed to transmit through the quadrupole (31). This requirement to narrow in on a single ion does create a slight limitation, however. When performing a scan, the majority of the ions emitted from the source actually go unobserved at any point in time. Were these ions injected together, the fill time necessary to product an appreciable signal across the entire mass range would be significantly smaller. Fortunately, this weakness is somewhat offset by the quadrupole's rapid voltage transitions and high sensitivity.

Quadrupoles are often used for additional ion manipulations by using them in tandem. A common configuration for quadrupoles is the so-called triple quadrupole, containing 3 quadrupoles in tandem (29). The first and third quadrupole are typical resolving quadrupoles, wherein they can effectively isolate any particular ion being transmitted. However, the second quadrupole is filled with gas and is used as a collision cell. Ions can thus be isolated in the first quadrupole, fragmented in the second one, and the fragment ion masses determined with the third quadrupole. This enables quadrupoles to better functionally characterize ions beyond simply their molecule weight.

### ***Ion Traps***

Ion trap operation behaves similarly to the quadrupole in that it relies upon quadrupole fields to manipulate ions (24), but the actual principles of operation differ somewhat, leading to both advantages and disadvantages of the two devices. Ion traps, or more specifically quadrupole ion traps or Paul traps, fundamentally confine ions based on a quadrupole field, just as is the case with resolving quadrupoles. However, unlike resolving quadrupoles, they are more typically fabricated with a hyperbolic surface rather than round rods in order to generate a more ideal quadrupole field. They generally come in 2 varieties, namely 3D ion traps(32), which make use of a ring electrode and two end-caps to confine ions in 3 dimensions, and 2D or linear ion traps which only use RF potentials to confine ions in 2 dimensions (33). Both variants are pictured in Figure 1.4.

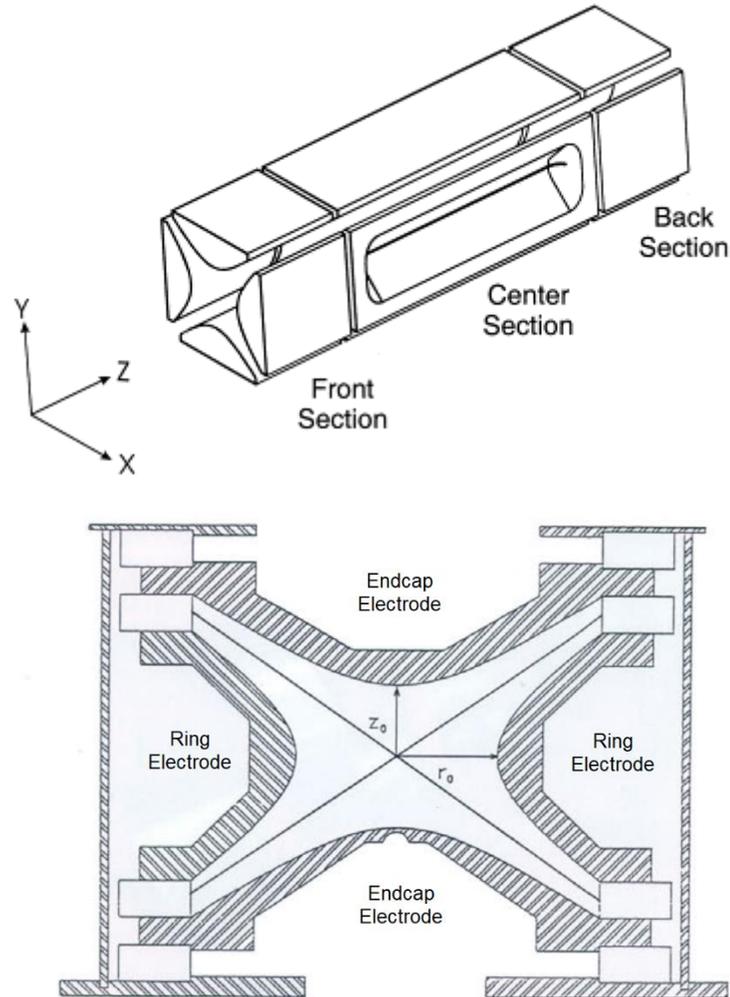


Figure 1.4: Representation of a 2D/linear ion trap (top) and a 3D ion trap (bottom). The ring electrode on the 3D trap penetrates through the page, forming a symmetrical ring in the x-dimension. Images adapted from (33) and (32), respectively

The nature of the manipulations in an ion trap are quite distinct from a quadrupole. The intricacies of these manipulations will be discussed in section 2.2, but a brief description will be offered here. Rather than implementing a DC offset to impact ion stability, most ion traps are operated in an RF-only modality without any additional DC offset between the sets of rods (26). Indeed, rather than using this DC offset to mass-selectively stabilize ions, ion traps actually detect ions on the basis of mass selective instability within the trap (34). Put simply, ions will adopt a particular frequency of

oscillation within the ion trap on the basis of their mass-to-charge ratio as well as a variety of modifiable instrument parameters (24, 35). Ions can be manipulated via resonant excitation, wherein an additional dipolar potential is applied at the same frequency at which the ion is moving, constructively interfering and imparting energy to that molecule in a frequency-specific manner. This dipolar potential is most commonly applied in a single axis with opposing rods possessing opposing polarities (unlike the main trapping RF where they possess the same polarity) (36). The voltage configuration to achieve this kind of excitation is pictured in Figure 1.5.

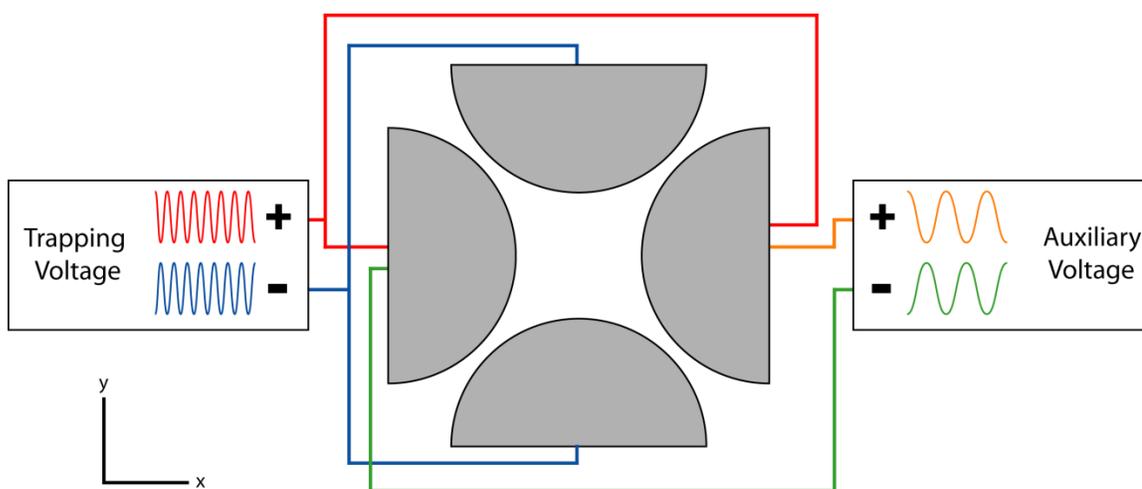


Figure 1.5: Voltage application configuration on a linear ion trap

In order to perform a scan, a specific frequency is excited with a high enough amplitude to eject an ion at that potential from the trap (34). Typically, the selected frequency is close to the upper bound of frequencies which can be stably contained within the trap, although this is not necessarily the case. Experimental parameters are then ramped such that ions sequentially adopt this frequency and are sequentially ejected from the trap on the basis of their mass-to-charge ratio (34). The trap is designed with an opening in the axis of excitation to enable ions to leave the trap without colliding with a

rod. In the case of the linear ion trap, these slits are present in the x-rods of the ion trap (33). Upon leaving the confines of the trap through these openings, the ions then impinge upon an electron multiplier which generates a signal when struck by an ion. As such, ions are ultimately detected on the basis of a particular ion's mass-selective instability.

The resonant excitation used for ejection of ions may also be used for additional ion manipulations as well. Ions can be isolated for further processing using mass-selective stability like a quadrupole, but this is instead performed by applying a broadband waveform which resonantly excited all ions except for the ion of interest (34, 37). Low amplitude excitation can also be used for a variety of applications. For example, ions can be excited at amplitudes insufficient to eject them from the trap, but high enough to result in numerous collisions with the helium bath gas (24). These collisions which convert the translation energy imparted by the excitation into vibration energy which eventually leads to a buildup of internal energy, resulting in collision induced fragmentation (37).

These ion manipulation techniques provide ion traps the tools to fill a lot of the same niches as a quadrupole, albeit with some significant advantages and drawbacks. Ion traps have significantly better sensitivity when performing MS1 profile scans due to accumulating all ions emitted from the source at the same time rather than sequentially checking each mass one by one. This is a big advantage when viewing all ions at once, but if an experiment is aimed at observing a particular peak (known as single ion monitoring or SIM scans), the quadrupole has an advantage due to additional manipulation beyond adjusting the RF/DC values in order to observe only that ion.

Quadrupoles are typically superior for isolating ions for the same reason, though ion traps can operate perfectly well in this capacity, albeit slower.

Advantageously, ion traps massively expand experimental flexibility. For example, as the ions are always confined in the same space and simply being manipulated for the purposes of isolation, fragmentation, etc., ion traps are capable of arbitrarily many ion manipulations, allowing for as many  $MS^n$  experiments as desired. Meanwhile, a quadrupole setup is generally restricted to the number of quadrupole segments used and the number of collision cells, assuming multiple fragmentation events is desired. Further, due to stably confining ions in a specific region of space, ion traps can much more easily be used as reaction vessels for either ion-ion reactions(38, 39) or photoactivation (40). As such, ion traps enable a much wider variety of experiments than a solely quadrupole based instrument.

### *Orbitraps*

Conversely, Orbitrap mass spectrometers are designed to make ensemble measurements of ions in a mixed population. At base, Orbitrap mass spectrometers are refined Kingdon traps where the refinement of the structure enables mass spectrometric measurements to be made (41). In their simpler state, Kingdon traps are a form of ion trap constructed from a metal wire surrounded by a cylindrical electrode (42). The combination of the potential generated by these two components results in a radial logarithmic field which keeps ions radially orbiting the central wire (41, 42). The ions are finally trapped axially by an end-cap electrode, causing them to be repelled when sufficiently close, keeping them confined toward the center of the trap.

However, in a traditional Kingdon trap, the ion motion along the axial direction is not particularly well defined. To combat this problem, the Orbitrap instead operates using a specially shaped center spindle and outer electrodes to create an additional quadrupolar field component in the axial direction, similar to the fields used in a quadrupole or ion trap. Originally illustrated by Knight, the addition of this axial quadrupolar potential allows ions to be confined in the axial domain in a manner that causes them to adopt characteristic harmonic oscillations in that axis, analogous to their behavior in the quadrupolar fields of an ion trap (41, 43). These oscillations are then recorded in the form of image current, wherein the ion motion induces a current, known as an image current, in the outer electrodes which is characteristic of their frequency.

Notably, this kind of detection differs from that of linear ion traps or quadrupoles in that it is an ensemble measurement rather than sequential stabilization or destabilization of ions. Ions are injected into the Orbitrap using a bent flat-a-pole known as a C-trap (44). During this injection, the ions are coalesced into a small window with a narrow range of energies during injection into the Orbitrap. Once inside, all ion packets will tend to form discrete rings around the central spindle electrode (due to rapid ion dephasing in the radial dimension) and oscillated in the axial quadrupolar field at their  $m/z$ -dependent frequency (41, 43), as depicted in Figure 1.6. The aggregate frequencies of all ions injected are collected together, but these signals are ultimately decoupled by performing a Fourier transform on the data and extracting the individual ion frequencies. These frequencies can then be assigned to a given  $m/z$ , generating the complete mass spectrum.

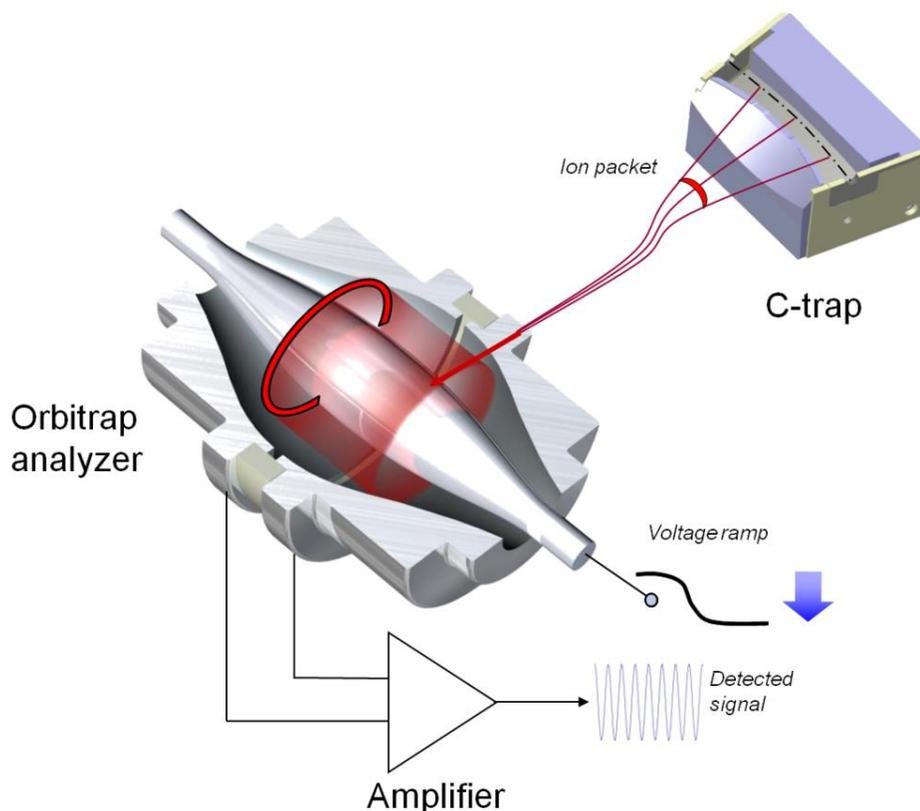


Figure 1.6 Schematic representation of Orbitrap ion injection and analysis. Image produced by Thermo Scientific (Bremen)

The biggest advantage offered by Orbitrap mass spectrometers is the high resolution enabled by the device. While the typical resolution of an ion trap, for example, typically caps out at around 7,000 resolution for the slower (~200 ms), high resolution scan full scan modes, an Orbitrap mass analyzer can easily achieve 60,000 resolution in the same timeframe (45, 46). The analyzer still has limitations, of course. Orbitraps are far less sensitive than ion traps or quadrupoles by virtue of using image current detection rather than electron multipliers; Orbitrap scans often use ~10 times as many ions as a similar scan performed in an ion trap. They are also limited in their capacity for non-analysis ion manipulations and must be paired with a quadrupole or ion trap to perform all but the most rudimentary fragmentation techniques. Nevertheless, the high resolution

enabled by this instrument is imperative for the precise analysis of molecules large enough to necessitate high resolution measurements for accurate results, of which proteins are a prime example.

## **1.4 Ion Dissociations**

While the ability to distinguish molecules by their intact mass can be of some utility for identifying a particular species, the combinatorial complexity of proteins becomes significant enough that it becomes impossible to confidently delineate unknown proteins on the basis of their mass alone; the possibility of isobaric species or amino acid inversions is too great. As such, techniques are necessary in order to effectively discriminate ions based on their constituent parts, thus determining the sequence of their amino acids. To this end, a wide variety of dissociation techniques have been developed in order to divide peptides into their more basic elements, ranging from dissociation by light to electron-based chemistry. Within the scope of this work, two dissociation strategies are primarily used, namely collisional-based dissociation and electron-based dissociation.

### ***Collisional Dissociation***

Fundamentally, collision-based fragmentation techniques rely on ion collisions with an inert bath gas to drive the fragmentation process. These collisions can be driven by a variety of experimental approaches, as will be discussed later, but the basic chemical behavior remains the same (47). As ions are imparted with translational energy from an external source, the collisions with the inert bath gas will convert some of this

translational energy into vibrational energy within the molecule. Once this vibrational energy becomes sufficiently high, the molecule is driven to undergo a molecular rearrangement (47, 48). Under most working conditions for multiply charged protein cation, this rearrangement is driven by mobile protons present on the peptide backbone oxygens promote nucleophilic cyclization reactions from nearby backbone oxygens or amino acid side-chains, thereby breaking the peptide bond and liberating b/y fragment ions (47, 49), as illustrated in Figure 1.7.

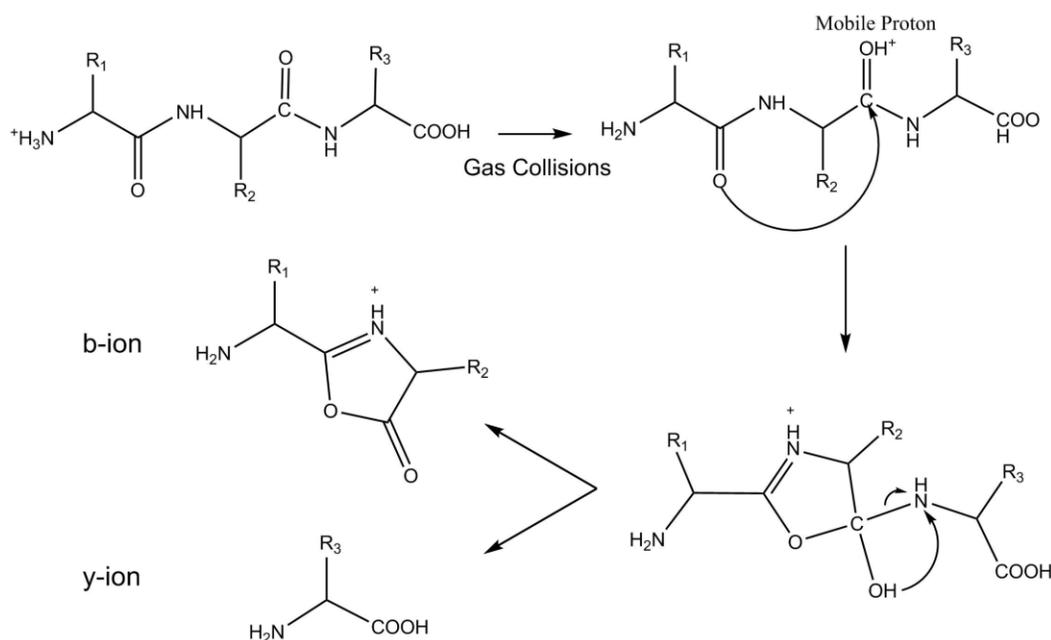


Figure 1.7: Mechanism for CAD fragmentation. Scheme adapted from (47, 49).

Collisional based techniques conducted on current instrumentation can be broadly segmented into trap-type and beam-type collisional fragmentation (48). Although the underlying chemical rearrangement underpinning these two techniques does not differ, the experimental parameters under which these two types of experiments are performed contrast in ways that result in slightly different experimental outcomes. As such, both

variations come with their own strengths and weaknesses, which may be appropriately capitalized on.

Trap-type collisional dissociation achieves deposition of energy into the fragmented ion on the basis of a kind of “slow heating” achieved through sustained resonant excitation of an ion at a particular  $m/z$  (34, 48). By making use of the process of resonant excitation, as described in Section 1.3, an ion within the trap is driven at its particular frequency of oscillation within the ion trap, imparting translational energy to it. As the goal of this experiment is to fragment the ion and eject it, the actual amplitude of excitation is quite limited, typically to around  $\sim 10$  eV per charge on the analyte (37). Therefore, ions undergo a low amplitude, sustained excitation for an extended period of time (several milliseconds) while undergoing thousands of collisions with the bath gas until they finally pass the energy threshold for fragmentation and undergo an intramolecular rearrangement. At this point, the precursor is segmented into two different fragment ions, and as these fragment ions possess their own distinct  $m/z$ , energy deposition ceases. Thus, the process of trap-type activation injects energy just sufficient to result in fragmentation, and the resulting ions are then allowed to relax down to lower energy states and avoid further fragmentation.

Trap-type dissociation comes with a variety of benefits. For one, the extent of activation for different protein precursors tends to be quite extensible between different analytes. As analyte ions are continuously excited until achieving dissociation, trap type CID (performed with sufficient activation energy) tends to generate fragment-rich spectra, achieving fragment ion yields of nearly 100%. Continuous energy deposition until the point of fragmentation ensures that the precursor ions sustain only the necessary

amount of energy to fragment and typically little more, improving fragmentation spectrum sensitivity. Further, trap-type CID is likely to produce highly consistent fragmentation spectra across a range of analytes. This is similarly because ions will be excited until they achieve sufficient energy to fragment and then cease to be excited, so variations in fragmentation threshold tend to have a less significant impact.

Beam-type CID differs from trap-type CID in that rather than a particular precursor  $m/z$  being continuously activated at a resonant frequency, ions are instead collisionally activated as an entire population at a largely fixed input of energy. Instead of being resonantly excited within a neutral bath gas, ions are instead passed between two pieces of ion optics with a much greater offset voltage difference (48, 50, 51). In essence, this makes the voltage gradient steeper, imparting more energy to ions as they pass between components. While this procedure can technically be performed anywhere within the instrument, it is most typically performed by passing ions into a dedicated collision cell. These collision cells are filled with an inert buffer gas as in trap-type CID, but commonly the buffer gas used involves a heavier gas like Nitrogen or even Argon, increasing the energy of a given collision (48).

The most pertinent distinction between beam-type and trap-type CID is the fact that trap type CID imparts a single, large amount of energy to the precursor ions rather than the slow-heating involved in trap-type CID. This largely produces two effects. Given that ions in beam-type CID begin the process with a fixed amount of energy, fragment ions may still remain sufficiently excited following fragmentation to continue to undergo further collisions with energy sufficient to fragment a second time (52). This has the capacity to dilute or destroy useful fragment ion signal through secondary fragmentations

which may either generate a neutral loss (e.g. water loss) or fragment into successively smaller fragment ions, respectively. It also risks leaving some intact precursor for ions which fail to undergo enough collisions for dissociation. However, successive, higher energy collisions conditions may bring the advantage of increasing energy deposition prior to fragmentation, leading to better mobilization of protons and thereby increasing the randomness of cleavages along the backbone (48). While this effect is likely less substantial for smaller peptides, which contain comparatively few bonds, it becomes significant in the context of large peptides and intact proteins which can often exhibit a more limited subset of fragmentation pathways when analyzed using trap-type CID (48).

### ***Electron Transfer Dissociation***

Electron based modes of dissociation represent an attractive alternative to collisionally driven modalities by taking advantage of radical-based chemistry rather than vibrationally induced rearrangements, as used in CAD. At base, ETD is a gas phase chemical reaction between a negatively charged reagent ion and a positively charged analyte ion wherein a single electron is transferred and initiates a fragmentation event. Fragmentation via ETD proceeds via a radically driven rearrangement which predominantly initiates an N-C $\alpha$  bond cleavage (as opposed to a peptide bond cleavage as in CAD), forming even-electron c-type and odd-electron z-type product ions. The exact molecular mechanism remains a topic of debate, with several different models proposed by groups at Cornell (53), as well as Washington and Utah (54, 55). However, precise molecular mechanisms aside, the basic principles remain largely similar. An electron is either transferred into or migrates by some mechanism to the carbonyl carbon of a

backbone peptide bond and this carbonyl carbon abstracts a proton from a nearby basic residue. This radical then facilitates a rearrangement, resulting in segmentation of the N-C $\alpha$  and creating the distinct c and z fragment ions (38). This mechanism is depicted in Figure 1.8. Although backbone cleavages are indeed the most dominant reaction pathway in ETD, non-dissociated ETD events (ETnoD) and several side-chain fragmentations can also result from an electron capture, appearing as charge-reduced precursors and neutral mass losses relative to the charge reduced precursor, respectively (56).

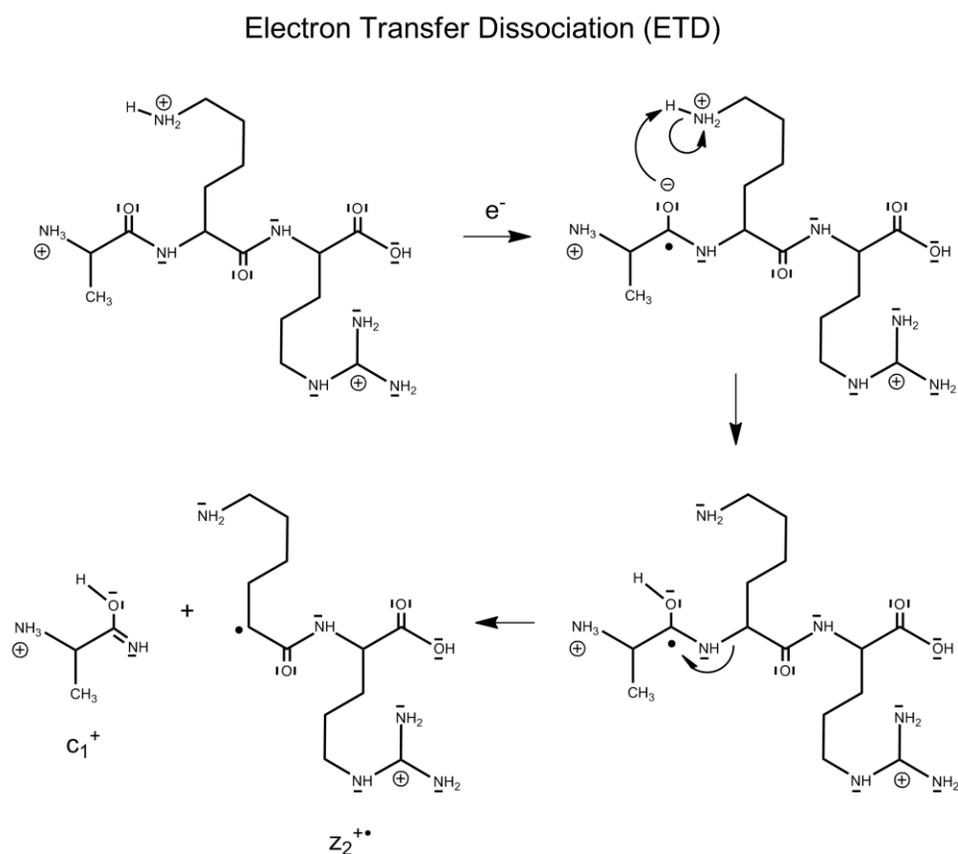


Figure 1.8: Mechanism for ETD fragmentation

This kind of fragmentation scheme possesses many advantages relative to alternative fragmentation techniques. Broadly speaking, ETD fragmentation specificity tends to be significantly lower than that of methods like collisional dissociation (38, 53).

Unlike CAD, which tends to strongly favor cleavages at particular amino acids, ETD fragmentation is largely random along the peptide backbone. This becomes particularly advantageous as molecule size increases; while collisional activation is often well suited for small peptides, cleavage patterns often tend to prefer a few highly preferred cleavage sites even once the total number of bond cleavages has increased by an order of magnitude. As a result, ETD sequence coverage is typically significantly better, particularly when analyzing larger molecules (57, 58). Additionally, ETD has been shown to reliably preserve many of the post-translation modifications often lost (47) or translocated (59) when using collisional activation types, for example phosphates and glycans (56, 57, 60). As such, it is particularly reliable at performing post translational modification site localization for these ubiquitous modifications. This feature equivalently scales reliably with size, as the potential for modification site scrambling generally increases with size as the number of potential modification sites goes up.

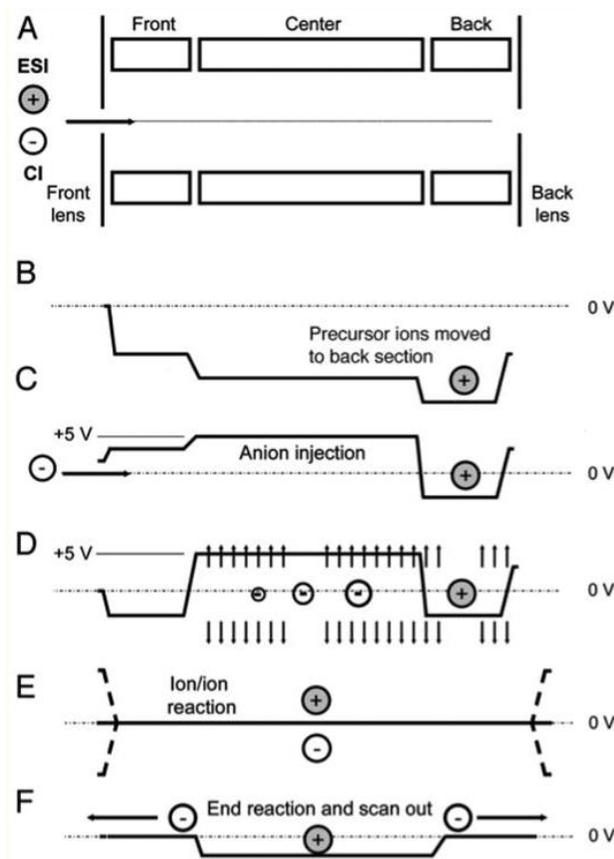
Unfortunately, ETD possesses its own limitations as well. Perhaps the most significant is ETD's strong dependence on charge density to generate fragmentation rich spectra. (57, 61, 62) Peptides and proteins with insufficient charge density have been shown to consistently generate higher proportions of non-dissociative electron transfer dissociation (ETnoD), wherein the molecule captures an electron but does not dissociate into informative *c/z* fragment ions. This is thought to occur in part as a result of a more compacted structure in the gas phase, resulting in cleavage of the peptide backbone, but causing the two fragment ions to remain noncovalently associated with one another (61); increased charge density not only unfolds the protein, but also increases the Coulombic repulsion between the resulting fragment ions, aiding in their dissociation. Further, ETD

tends to be ill suited for smaller peptides, as charge neutralization from the ensuing electron capture will neutralize the charge from either all or half of the resulting fragment ions if the charge state is as low as +1 or +2, respectively; fortunately this becomes less of an issue when working with larger peptides and proteins (63). Nevertheless, ETD represents an incredibly powerful tool for analyzing proteins which fall within its applicable scope, due to its strong propensity for random cleavages and resulting high sequence coverage.

Performing an ETD reaction is far less trivial than a collisional dissociation experiment, and as such requires somewhat more complex ion manipulation in order to utilize the technique. Appropriate ions must simply be manipulated such that both positive analyte and negative reagent ions are co-trapped in the same region of space, enabling the reaction to proceed. This is actually quite challenging given how both ion manipulation as well as axial trapping in a linear ion trap are both achieved with static DC offsets under normal circumstances. Application of a negative potential, for example, to confine the positively charged precursor ions would repel the negatively charged reagent ions, making spatial confinement of the two species challenging. Thankfully, the segmentation of the linear ion trap can be utilized to accomplish this feat.

The DC offset for each section of the linear ion trap can be manipulated independently, allowing different populations of ions to be accumulated within each (38). Using an ion source where all ions are generated from the same region of the instrument (front-end ETD), precursor ions are first injected into the trap, isolated (if not isolated in a previous step), and then stored within the back section of the ion trap by applying a strong negative potential to this trap sub-section (64). The center section of the trap is

then held at a positive potential, and the reagent ions are stored there. To enable the mixing of these two populations, the DC offset potentials are then moved to ground, resulting in no net trapping in the axial dimension and allowing the ion populations to mix. In order to prevent ions from falling out of the ends of the trap, the lenses just outside of the ion trap are given a high frequency RF potential (similar to that of the main trapping RF) rather than their DC potential, allowing them to serve as pseudo-endcap electrodes for the duration of the reaction. Finally, the trap center section is returned to a negative DC potential, selectively trapping the resulting fragment ions, but causing the remaining reagent ions to fall out of the trap, thus halting the reaction. This procedure is illustrated in Figure 9.



**Figure 1.9: Ion injection scheme during front-end ETD.** Adapted from (64).

This process comes with some slight limitations. The largest is the fact that this procedure significantly limits ion targets compared to what is usable in other fragmentation techniques. Using this scheme, precursor ions are held in the back section of the trap which possesses ~20% of the storage capacity of the center section; consequently, the total analyzed ion population is only a fifth of what it would be in an equivalent collisional experiment (64, 65). Naturally, the smaller analysis pool results in reduced signal intensity and can thus negatively impact MS/MS results. Thankfully, signal amplification strategies, for example multiple c-trap fills which sums iterative ion-ion reactions prior to analysis, can be used to compensate for this storage deficit that occurs in a single ion-ion reaction, (although these strategies can come with their own complications, which will be discussed later) (64).

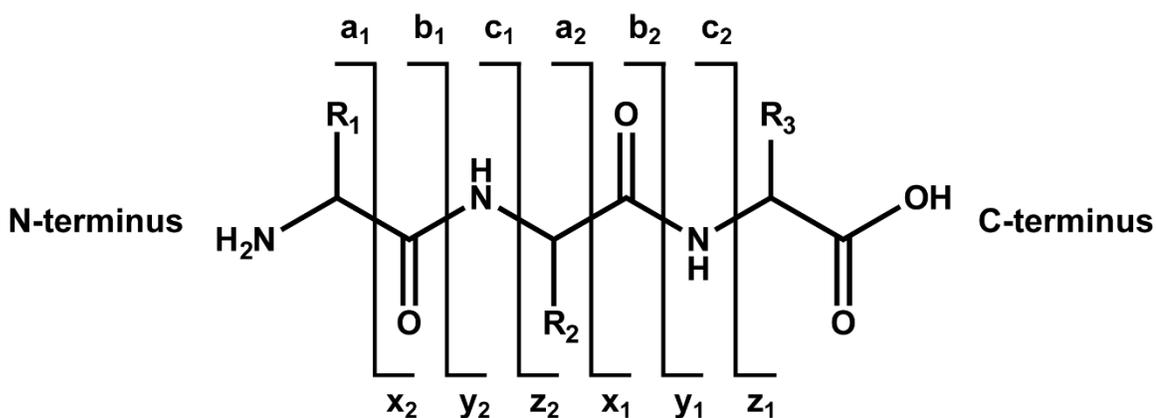
Despite these drawbacks, the advantages provided by ETD significantly outweigh any weaknesses associated with its use. ETD consistently shows among the highest performance in terms of sequence coverage on large biomolecules and is therefore invaluable for proceeding with the analysis of such systems.

## **1.5 Protein Sequencing**

Having established the tools necessary for protein identification, it becomes possible to use them in combination to deduce the structure of a particular analyte. Naturally, the first step toward identifying the peptides in the sample is going to be to perform an MS<sup>1</sup> scan, identifying the masses of all peptides eluting at a given point in the analysis. While the determination of protein or peptide's intact mass certainly limits its scope of possible sequences, under most circumstances it is insufficient to definitively

identify the peptide or protein in question. It is quite common for several sets of peptide sequences to be isobaric within a particular mass tolerance, either due to identical amino acid compositions or just the general chemical similarity among the amino acids. This is where the aforementioned dissociation techniques become necessary to extract further information.

On the basis of the peptide masses observed in the MS<sup>1</sup> spectrum, a particular ion of interest may then be isolated from the other ions in the population and dissociated by a variety of techniques, which will be discussed in detail later. Fundamentally, the goal of these dissociation methods is to generate consistent fragmentation at some site along the backbone, such that the mass differences between the sequential amino acids can be determined. Upon dissociation, the peptide backbone can segment at a variety of locations where n-terminal fragment ions are denoted a, b, and c ions and c-terminal ions are denoted x, y, and z ions based on the precise cleavage site, as denoted in Figure 1.10. A mass spectrum is then produced for all of these fragment ions in order to determine their masses. This technique is referred to as tandem mass spectrometry, MS/MS, or MS (66).



**Figure 1.11: Nomenclature for denoting peptide fragment ions.** Fragment ions are numbered sequentially from each terminus and fragments are denoted as a/x, b/y, and c/z pairs based upon the backbone cleavage position.

An important feature to note is that regardless of the actual location of the bond cleavage along the backbone, the mass difference between two sets of repeating amino acids remains the same, as the repeating subunits along the backbone should remain constant, and only the side-chain mass should change. This allows for the determination of the amino acid structure on the basis of the mass differences between each of these repeating cleavage sites. The basic process of sequencing a peptide is illustrated in Figure 1.11, using an example peptide segmented into b and y type fragment ions. Briefly, terminal fragment ions may be assigned as a given type of fragment ion on the basis of the static mass difference associated with the N- or C-terminus. This mass difference will change as a result of the backbone cleavage site in question but remains consistent for a given ions series. The next amino acid in the series may then be identified by looking for a fragment of the previous mass plus the residue mass of an additional amino acid, thus identifying the next amino acid in the sequence. Iterating this process over the length of the molecule thus allows for the order of amino acids within the molecule to be deduced from the MS<sup>2</sup> spectrum.

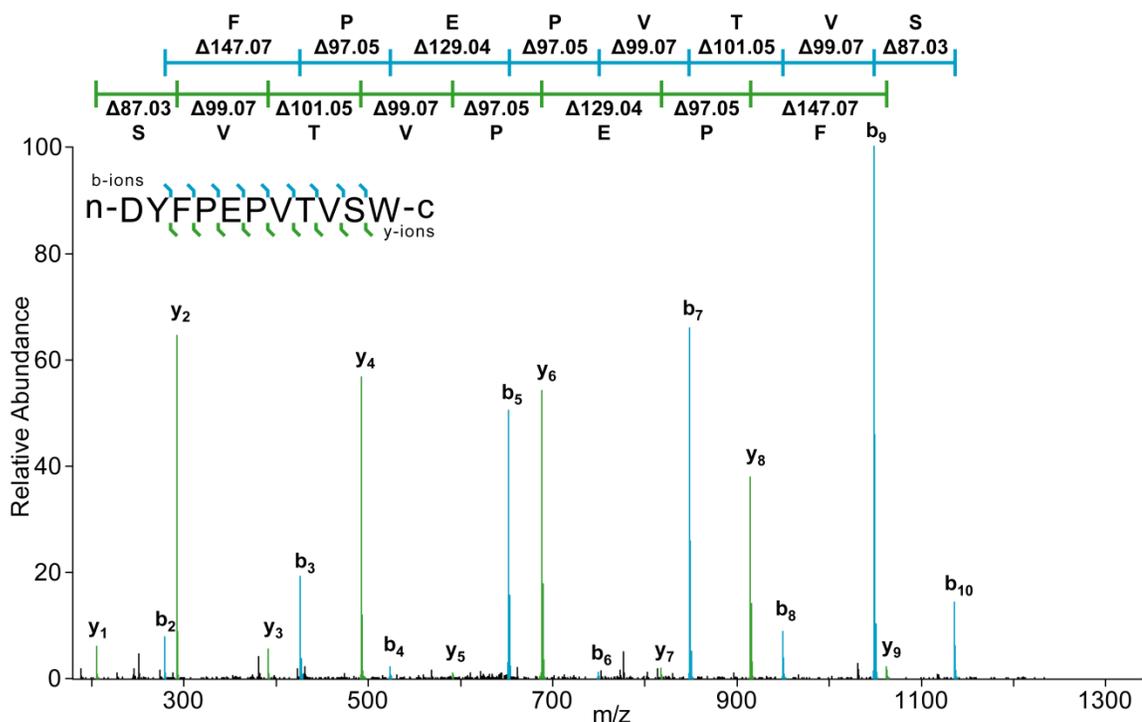


Figure 1.11: Strategy for protein sequencing on a CAD spectrum. Mass differences between both n-terminal (b-type) and c-terminal (y-type) ions can be related to the masses of amino acids, sequentially determining the sequence from both pairs of fragment ions.

In principle, sequence redundant or overlapping peptides can provide additional utility for protein sequencing beyond that of clearly associating two different peptides with one another. While this is indeed an advantage, in principle, related but dissimilar peptides can generate complimentary sets of sequence coverage. Although the precise explanation for this phenomenon is not definitively established, plausible explanations include altered chemical properties (owing to the disparate portions of the sequence) as well as possible differences in gas phase tertiary structure.

Once two overlapping peptides have been identified, their individual sequence coverages can be cross-referenced and superimposed in order to fill in the sequence information missing in one another's fragmentation spectra. An example of this utility is illustrated in Figure 1.12. Two different fragmentation spectra for the peptide in question

are displayed, one with ETD and one with CAD. Notably, neither fragmentation spectrum achieves observation of all bonds in the molecule. However, they provide different information which can be combined to generate a better composite picture. For example, the V-C bond remains intact during ETD, meaning that these two residues cannot be clearly placed in order. However, as that cleavage is present in the CAD spectrum, the order can be clearly delineated on that basis. Similarly, the R-E cleavage is unidentified in the CAD spectrum, yet is present in the ETD spectrum, enabling the discrimination of that amino acid order as well. Thus, the superimposed fragment ions generate a level of characterization unattainable using only a one of the two techniques.

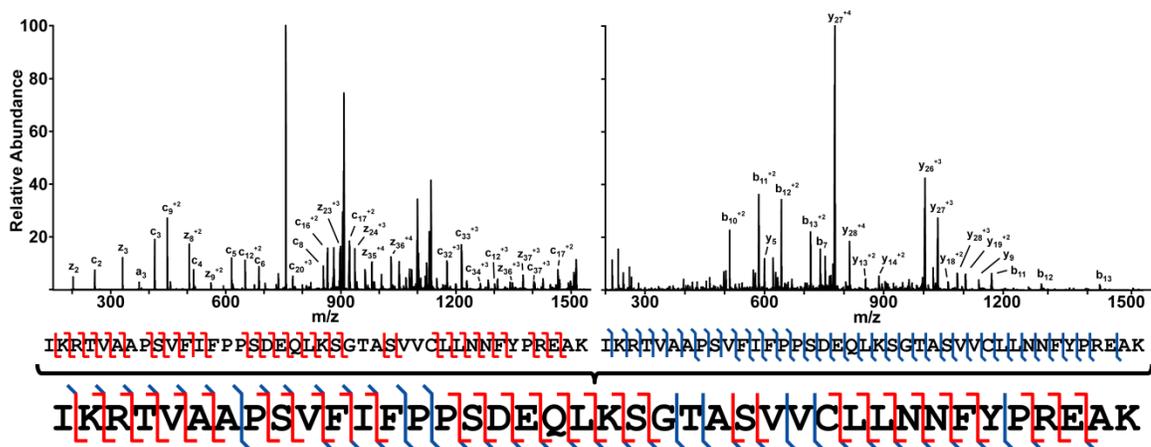


Figure 1.12: Superimposition of multiple fragmentations spectra to improve sequence coverage. Notably, each fragmentation spectrum contains bond cleavages which cannot be distinguished by fragment ions, but which can be distinguished in the alternate fragmentation type. Thus, combining the two spectra improves sequence information.

## 1.6 Modes of Sample Preparation

While these sequencing strategies are generally applicable for any proteinaceous material, the difficulty of sequencing any given species increases substantially as they go up in size. Several features are at play here, for example depleted signal-to-noise due to

separation into multiple charge states and isotopes, decreased copy number of trapped ions, increased number of competing fragmentation pathways, and significantly increased spectral complexity. As a result, the degree of characterization (under identical operating conditions) tends to go down as analyte size increases, often requiring strategies to enable more robust characterization.

Prior to analyzing protein material, a variety of preparation strategies are used in order to maximally extract information from the sample. The most conventional methodology, commonly known as “bottom-up” analysis, involves the digestion of analyte proteins with an endoprotease to convert all proteins to much smaller peptides, typically around ~1-3 kDa. This approach has a number of advantages. For instance, working with smaller peptides tends to significantly increase analysis sensitivity. This results, in part, from the better solubility, lower nonspecific binding, and superior chromatographic separation typically seen when working with small peptides. Smaller peptides are also typically divided into fewer detection pathways when analyzed in the mass spectrometer; the signal is divided up between fewer charge states as well as isotopes, reducing signal dilution for each individual species. Fragmentation spectra also tend to be significantly less complex, easing spectral interpretation, because they undergo fragmentation through a comparatively small pathways owing to the smaller number of dissociation sites between the limited number of amino acids in a given peptide.

However, digesting proteins into very small peptides also comes with a number of problems. Most straightforwardly, breaking the protein down into smaller peptides fundamentally removes the information about how the peptides were connected together to begin with. While this is partially remedied by comparing proposed spectra to

databases derived from genomic data, unambiguous assignment of a peptide to the exact gene product is often unachievable as a result of sequence similarity between protein isoforms or homologous protein sequences. Further, co-assignment of multiple highly sequence diverse protein subsections, for example antibody CDRs, is often impossible unless a peptide is generated which contains two or more of these sections, clearing connecting them to the same protein. Complete enzymatic digestion also loses information about the co-localization of post translational modifications which may be highly -significant biologically. As such, while complete digestion remains widely used, it faces many problems which make it inapplicable to many biologically interesting problems.

Intact protein analysis is quickly becoming a popular alternative to digestion analysis, however (67). Analysis of intact proteins circumvents many of the limitations inherent to analyzing digested proteins. The protein inference problem associated with peptide sequences to a larger product is largely eliminated, as the protein's structure should ideally remain unaltered prior to analysis within the mass spectrometer. For the same reason, the various combinations of post-translational modifications and sequence variable subunits can also be observed as independent proteoforms, allowing discrimination of and PTM cross talk or unambiguous co-localization of highly sequence variable regions of a given protein (68). These advantages make intact analysis a very attractive goal to pursue, as they may enable a depth of analysis unachievable from more conventional techniques.

Unfortunately, intact analysis comes with many technical hurdles which have yet to be resolved definitively. For instance, sample preparation for intact proteins has yet to

be effectively worked out to the same level as peptides; selective purification, minimal sample loss, and complete separation are all far more challenging on intact proteins than their peptide counterparts (69). Further, particularly large intact proteins are far more challenging to analyze in a mass spectrometer. Signal sensitivity drops precipitously as the signal from a given molecule gets divided up between both charge states and isotopes, quickly diluting the signal into a plethora of pathways and reducing signal intensity (70). Further, as mass spectrometers fundamentally store a given number of charges rather than a number of molecules, fewer copies of a protein are collected in the instrument as well (70). As a direct result of these two problems, when a protein is fragmented, not only is its signal divided between far more pathways, but there are also fewer copies to be divided up. This tends to limit the number of amino acids in the sequence which can be unambiguously matched back to a theoretical protein sequence as well as complicating the placement of a post-translational modification.

## **1.7 Antibodies**

The class of protein under investigation in this work is that of antibodies, in particular IgG-like monoclonal antibodies. Interest in exhaustively analyzing antibodies is rapidly increasing. Antibodies are large proteins which function, under natural circumstances, as a part of the adaptive immune system. They are produced and secreted by B cells, primarily by mature, differentiated B cells known as plasma cells (71). Within their role in the immune system, antibodies primarily serve to bind antigens present on foreign pathogens (to which they are trained to bind with a high affinity) and subsequently block and deactivate host virulence factors, aggregate foreign cells and

antigens, promote phagocytosis/cell killing, and activate the immune complement system (71). All of these functions help to either neutralize the pathogen or mediate the immune system's ability to eliminate them.

However, in addition to their important biological role, antibodies have been found to possess outstanding potential to be administered as therapeutics for a wide range of medical conditions. In this context, the immunological effects of antibodies are co-opted in order to target specific disease manifestations or certain cell types which are not being effectively regulated by the immune system. For instance, antibodies have been engineered to agonize (72) and antagonize specific receptors (73), precipitate out signaling molecules to reduce their effect (74), trigger cell killing of cancerous cells (75), or deliver attached drug payloads selectively against a specific cell type (76). One difficulty in using antibodies in this regard is maintaining a highly consistent product, but the exhaustive characterization of substantial molecules like antibodies is far more difficult than typical small molecule drugs.

Structurally, antibodies are quite complex which significantly complicates their analysis. Individual antibody "monomers" generally weight approximately 150 kDa and are formed from a heterotetrametric complex between a pair of sequence identical heavy chains (~50 kDa) and light chains (~25 kDa) which roughly forms a Y-shaped structure (77). A schematic representation of an antibody is pictured in Figure 1.13. An antibody monomer's structure can be broadly segmented beyond the paired chains – the fragment antigen binding (Fab) region and the fragment crystallizable (Fc) region.

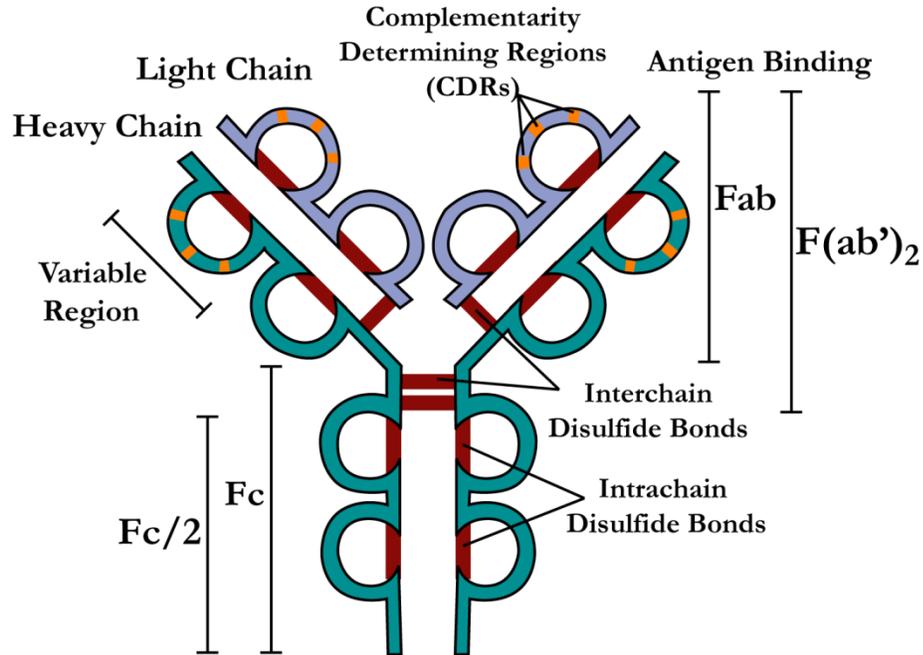


Figure 1.13: Antibody Structure Schematic

The Fab which forms the arms of the Y-shape is so named because it is the portion of the molecule which possesses an affinity to the target antigen. This portion of the molecule tends to exhibit the highest degree of sequence variation, particularly with its so-called variable region. This region is highly sequence polymorphous as a result of two processes, namely V(D)J recombination and somatic hypermutation, which insert a significant degree of random sequence variation into this region of the molecule (71). The variable region represents the N-terminal most portion of the molecule and contains the 6 complementarity determining regions (CDRs), 3 on each chain, which are both the most highly sequence variable portions of the molecule as well as the amino acids residues which predominantly participate in actual antigenic contact (78). Unsurprisingly, mutations or modifications to this region of the molecule can have a dramatic impact on an antibody's affinity to its target antigen, and may entirely

compromise its efficacy (79, 80). Further, the Fab is the most challenging to analyze because of the high potential for variation between different molecules.

The remaining C-terminal portion of the molecule is known as the fragment crystallizable (Fc) region. This portion of the molecule is generally less sequence variable, but still mediates a large portion of the biologically relevant activity. It is responsible for the so-called Fc effector functions. Many immune cells will possess a variety of Fc receptors which, upon interacting with the Fc of antibodies bound to pathogens, will promote immunological responses like phagocytosis, cytokine release, and antimicrobial cytotoxin release (71). Further, antibody hexamers will activate the complement cascade, typically in either pentamers or hexamers, using their Fc region (77). This complement system will itself promote inflammation, phagocytosis, and membrane damage to foreign cells (71). Unsurprisingly, these functions, as well as antibody half-life (81), are significantly impacted by both the Fc sequence as well as protein modifications like glycans, so effectively utilizing these properties requires consistency in this portion of the molecule as well (82).

Given the degree of characterization required, powerful instrumentation and unique sample preparation strategies become necessary to appropriately characterize these molecules. To this end, the work discussed in this dissertation illustrates the potential of novel ion-ion reaction manipulations and limited digestion strategies which, when used together, can significantly improve both the speed and degree of characterization achievable on these molecules.

## References

1. Voet D, Voet JG (2011) *Biochemistry* (John Wiley & Sons) Available at: <https://www.wiley.com/en-us/Biochemistry%2C+4th+Edition-p-9780470570951> [Accessed April 22, 2019].
2. Aebersold R, Nature MM-, 2003 U (2003) Mass spectrometry-based proteomics. *nature.com* 422:198–207.
3. Savaryn JP, Toby TK, Kelleher NL (2016) A researcher's guide to mass spectrometry-based proteomics. *Proteomics* 16(18):2435–2443.
4. Coon JJ, Syka JEP, Shabanowitz J, Hunt DF (2005) Tandem mass spectrometry for peptide and protein sequence analysis. *Biotechniques* 38(4):519–523.
5. Glish GL, Vachet RW (2003) The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov* 2(2):140–150.
6. Ardrey RE (2003) *Liquid Chromatography – Mass Spectrometry: An Introduction* doi:10.1002/0470867299.
7. Snyder L, Kirkland J, Dolan J (2011) *Introduction to modern liquid chromatography* (John Wiley & Sons, Inc, Hoboken, New Jersey). 3rd Editio Available at: <https://books.google.com/books?hl=en&lr=&id=taurNJU0u2AC&oi=fnd&pg=PT11&dq=L.+R.+Snyder,+J.J.+Kirkland,+and+J.+W.+Dolan,+Introduction+to+Modern+Liquid+Chromatography,+John+Wiley+%26+Sons,+New+York,+2009.&ots=ybFJ1o20vA&sig=bZx6nV1bDRo8-ErVzRXYd0CBYBE> [Accessed April 21, 2019].
8. Dawkins J V., L. Lloyd L, P. Warner F (1986) Chromatographic Characteristics of polymer-based high-performance liquid chromatography packings. *J Chromatogr A* 352(C):157–167.
9. Hamilton RJ, Sewell PA (1982) Introduction to high performance liquid chromatography. *Introduction to High Performance Liquid Chromatography* (Springer Netherlands, Dordrecht), pp 1–12.
10. Snyder LR, Dolan JW, Gant JR (1979) Gradient elution in high-performance liquid chromatography. I. Theoretical basis for reversed-phase systems. *J Chromatogr A* 165(1):3–30.
11. 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.
12. Hebert AS, et al. (2013) The One Hour Yeast Proteome. *Mol Cell Proteomics* 13(1):339–347.
13. Bath TS, Francavilla C, Olsen J V. (2014) Off-line high-pH reversed-phase

- fractionation for in-depth phosphoproteomics. *J Proteome Res* 13(12):6176–6186.
14. Washburn MP, Wolters D, Yates JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19(3):242–247.
  15. Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB (1985) Electrospray Interface for Liquid Chromatographs and Mass Spectrometers. *Anal Chem* 57(3):675–679.
  16. Cech NB, Enke CG (2001) Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* 20(6):362–387.
  17. Konermann L, Ahadi E, Rodriguez AD, Vahidi S (2013) Unraveling the mechanism of electrospray ionization. *Anal Chem* 85(1):2–9.
  18. Disintegration of water drops in an electric field (2006) *Proc R Soc London Ser A Math Phys Sci* 280(1382):383–397.
  19. Tafflin DC, Ward TL, Davis EJ (1989) Electrified Droplet Fission and the Rayleigh Limit. *Langmuir* 5(2):376–384.
  20. Iribarne J V. (1976) On the evaporation of small ions from charged droplets. *J Chem Phys* 64(6):2287.
  21. Mack LL, Kralik P, Rheude A, Dole M (1970) Molecular beams of macroions. II. *J Chem Phys* 52(10):4977–4986.
  22. Kebarle P, Verkcerk UH (2009) Electrospray: From Ions in solution to Ions in the gas phase, what we know now. *Mass Spectrom Rev* 28(6):898–917.
  23. Donor MT, Ewing SA, Zenaidee MA, Donald WA, Prell JS (2017) Extended Protein Ions Are Formed by the Chain Ejection Model in Chemical Supercharging Electrospray Ionization. *Anal Chem* 89(9):5107–5114.
  24. March R, Todd J (2005) *Quadrupole ion trap mass spectrometry* ed Winefordner JD (John Wiley & Sons, Ltd, Hoboken, New Jersey). 2nd Ed. Available at: <https://books.google.com/books?hl=en&lr=&id=JDKCoIshgSIC&oi=fnd&pg=PR5&dq=March+Todd+ion+trap&ots=GPfulPzMkQ&sig=7jjegYFkS8dmaaq3b4qbraw8sFo> [Accessed April 19, 2019].
  25. Douglas DJ (2009) Linear quadrupoles in mass spectrometry. *Mass Spectrom Rev* 28(6):937–960.
  26. Douglas DJ, Frank AJ, Mao D (2005) Linear ion traps in mass spectrometry. *Mass Spectrom Rev* 24(1):1–29.
  27. Douglas DJ, Kononkov N V. (2002) Influence of the 6th and 10th spatial harmonics on the peak shape of a quadrupole mass filter with round rods. *Rapid Commun Mass Spectrom* 16(15):1425–1431.
  28. Henchman M, Steel C (2009) Understanding the Quadrupole Mass Filter through

- Computer Simulation. *J Chem Educ* 75(8):1049.
29. Yost RA, Enke CG (2007) Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal Chem* 51(12):1251–1264.
  30. Dillen L, et al. (2012) Comparison of triple quadrupole and high-resolution TOF-MS for quantification of peptides. *Bioanalysis* 4(5):565–579.
  31. Picotti P, Aebersold R (2012) Selected reaction monitoring-based proteomics: Workflows, potential, pitfalls and future directions. *Nat Methods* 9(6):555–566.
  32. March RE (1997) An introduction to quadrupole ion trap mass spectrometry. *J Mass Spectrom* 32(4):351–369.
  33. Schwartz JC, Senko MW, Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 13(6):659–669.
  34. Stafford GC, Kelley PE, Syka JEP, Reynolds WE, Todd JFJ (1984) Recent improvements in and analytical applications of advanced ion trap technology. *Int J Mass Spectrom Ion Process* 60(1):85–98.
  35. Goeringer DE, Whitten WB, Ramsey JM, McLuckey SA, Glish GL (1992) Theory of High-Resolution Mass Spectrometry Achieved via Resonance Ejection in the Quadrupole Ion Trap. *Anal Chem* 64(13):1434–1439.
  36. Snyder DT, Cooks RG (2017) Single Analyzer Precursor Ion Scans in a Linear Quadrupole Ion Trap Using Orthogonal Double Resonance Excitation. *J Am Soc Mass Spectrom* 28(9):1929–1938.
  37. Mitchell Wells J, McLuckey SA (2005) Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol* 402(1993):148–185.
  38. 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* 101(26):9528–9533.
  39. McLuckey SA, Reid GE, Wells JM (2002) Ion parking during ion/ion reactions in electrodynamic ion traps. *Anal Chem* 74(2):336–346.
  40. Brodbelt JS (2014) Photodissociation mass spectrometry: New tools for characterization of biological molecules. *Chem Soc Rev* 43(8):2757–2783.
  41. Hu Q, et al. (2005) The Orbitrap: A new mass spectrometer. *J Mass Spectrom* 40(4):430–443.
  42. Kingdon KH (1923) A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys Rev* 21(4):408–418.
  43. Hardman M, Makarov AA (2003) Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal Chem* 75(7):1699–705.

44. 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):2010–2021.
45. Pekar Second T, et al. (2008) Dual-Pressure Linear Ion Trap Mass Spectrometer Improving the Analysis of Complex Protein Mixtures. *J R Proc Natl Acad Sci USA* 105(4):7757–7765.
46. Michalski A, et al. (2011) 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.
47. Palzs B, Suhal S (2005) Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* 24(4):508–548.
48. Xia Y, Liang X, McLuckey SA (2006) Ion trap versus low-energy beam-type collision-induced dissociation of protonated ubiquitin ions. *Anal Chem* 78(4):1218–1227.
49. Wysocki VH, Tsaprailis G, Smith LL, Brezi LA (2000) Mobile and localized protons: A framework for understanding peptide dissociation. *J Mass Spectrom* 35(12):1399–1406.
50. Olsen J V, et al. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 4(9):709–712.
51. Senko MW, Beu SC, McLafferty FW (1994) High-Resolution Tandem Mass Spectrometry of Carbonic Anhydrase. *Anal Chem* 66(3):415–417.
52. Michalski A, Neuhauser N, Cox J, Mann M (2012) A systematic investigation into the nature of tryptic HCD spectra. *J Proteome Res* 11(11):5479–5491.
53. Zubarev R, Kelleher NL, McLafferty FW (1998) Electron capture dissociation of multiply charged protein cations. A .... *J Am Chem Soc* 120(16):3265–3266.
54. Tureček F (2003) N-C $\alpha$  bond dissociation energies and kinetics in amide and peptide radicals. Is the dissociation a non-ergodic process? *J Am Chem Soc* 125(19):5954–5963.
55. Chen X, Tureček F (2006) The Arginine Anomaly: Arginine Radicals Are Poor Hydrogen Atom Donors in Electron Transfer Induced Dissociations. doi:10.1021/ja063676o.
56. Tureček F, Julian RR (2013) Peptide Radicals and Cation Radicals in the Gas Phase. *Chem Rev* 113(8):6691–6733.
57. Riley NM, Coon JJ (2018) The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal Chem* 90(1):40–64.
58. Anderson LC, et al. (2015) Protein derivatization and sequential ion/ion reactions

- to enhance sequence coverage produced by electron transfer dissociation mass spectrometry. *Int J Mass Spectrom* 377(1):617–624.
59. Palumbo AM, Reid GE (2008) Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. *Anal Chem* 80(24):9735–9747.
  60. Mikesh LM, et al. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta - Proteins Proteomics* 1764(12):1811–1822.
  61. Riley NM, Westphall MS, Coon JJ (2015) Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal Chem* 87(14):7109–7116.
  62. 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* 330–332:174–181.
  63. Swaney DL, et al. (2007) Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* 79(2):477–485.
  64. Earley L, et al. (2013) Front-end electron transfer dissociation: A new ionization source. *Anal Chem* 85(17):8385–8390.
  65. 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.
  66. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR (2006) Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci* 83(17):6233–6237.
  67. Catherman A, Skinner O, Biophysical NK-B and, 2014 U Top down proteomics: facts and perspectives. *Elsevier*. Available at: <https://www.sciencedirect.com/science/article/pii/S0006291X14002939> [Accessed April 21, 2019].
  68. Siuti N, Kelleher NL (2007) Decoding protein modifications using top-down mass spectrometry. *Nat Methods* 4(10):817–821.
  69. Lee JE, et al. (2009) A Robust Two-Dimensional Separation for Top-Down Tandem Mass Spectrometry of the Low-Mass Proteome. *J Am Soc Mass Spectrom* 20(12):2183–2191.
  70. 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.
  71. Murphy K, Weaver C, Weaver C (2016) *Janeway's Immunobiology* (Garland Science, 9th edition. | New York, NY : Garland Science/Taylor & Francis)

doi:10.1201/9781315533247.

72. Nahta R, Esteva FJ (2006) Herceptin: Mechanisms of action and resistance. *Cancer Lett* 232(2):123–138.
73. Mary C, et al. (2013) Antagonist properties of monoclonal antibodies targeting human CD28. *MAbs* 5(1):47–55.
74. Weinblatt ME, et al. (2003) Adalimumab, a fully human anti-tumor necrosis factor  $\alpha$  monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: The ARMADA trial. *Arthritis Rheum* 48(1):35–45.
75. Scott AM, Wolchok JD, Old LJ (2012) Antibody therapy of cancer. *Nat Rev Cancer* 12(4):278–287.
76. Alley SC, Okeley NM, Senter PD (2010) Antibody-drug conjugates: Targeted drug delivery for cancer. *Curr Opin Chem Biol* 14(4):529–537.
77. Diebolder CA, et al. (2014) Complement is activated by IgG hexamers assembled at the cell surface. *Science* (80- ) 343(6176):1260–1263.
78. Sela-Culang I, Kunik V, Ofran Y (2013) The structural basis of antibody-antigen recognition. *Front Immunol* 4(OCT):1–13.
79. Vlasak J, et al. (2009) Identification and characterization of asparagine deamidation in the light chain CDR1 of a humanized IgG1 antibody. *Anal Biochem* 392(2):145–154.
80. Habberger M, et al. (2014) Assessment of chemical modifications of sites in the CDRs of recombinant antibodies: Susceptibility vs. functionality of critical quality attributes. *MAbs* 6(2):327–339.
81. Higel F, Seidl A, Sörgel F, Friess W (2016) N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur J Pharm Biopharm* 100:94–100.
82. Borrok MJ, Jung ST, Kang TH, Monzingo AF, Georgiou G (2012) Revisiting the Role of Glycosylation in the Structure of Human IgG Fc. *ACS Chem Biol* 7(9):1596–1602.

## **ETD Parallel Ion Parking of Antibody Subunits**

### **2.1 Introduction**

The optimal platform for the complete characterization of intact proteins or protein subunits would be capable of fragmenting an ion such that the spectrum contains both possible fragment ions generated by every bond cleavage in the molecule. This would enable not only the unambiguous mapping of all amino acids in order while also clearly distinguishing all different variant proteoforms (1, 2). Although this is currently an unachievable standard, electron transfer dissociation represents the most promising fragmentation technique, given its largely nonspecific fragmentation specificity as well as the preservation of labile post-translational modifications (3, 4). However, the overall success of ETD is limited by problems such as inefficient precursor consumption (5) and fragment ion destruction(6), which in turn impedes its ability to achieve optimal fragment ion generation and therefore sequence coverage. Solving these problems is quite challenging, as they are in part a manifestation of gas phase ion-ion reaction kinetics in general rather than some feature unique to ETD in particular (6, 7).

Fortunately, the selective ion manipulations achievable on a linear ion trap provide a unique opportunity to control ions in a manner that allows the manipulation of this process rather than allowing the reaction to proceed in an uncontrolled manner, as is currently the most typical approach (8). To this end, this chapter aims to illustrate how parallel ion parking, a novel technique for selectively modulating ion-ion reaction kinetics, can be used to selectively manipulate ETD gas phase kinetics, thus maximizing the generation of useful fragment ions (8). When applied to “intact” antibody subunits,

generated by a highly specific single-cleavage event followed by disulfide reduction, comprehensive sequence coverage can be achieved on a chromatographic timescale.

## 2.2 Background

### *Quadrupolar Fields*

In order to better understand the theoretical underpinnings of the ion manipulations discussed in this chapter, it becomes necessary to employ a more detailed discussion of the operation of ion traps. For the sake of scope, this discussion will cover the behavior of ions within a 2 dimensional linear ion trap (LIT) (9), but the same principles apply to x-y confinement in a quadrupole mass filter. The behavior is also similar to that of a three dimensional (3D) ion trap (10), with the predominant distinction regarding trapping being related to the radial symmetry of the 3D ion trap in order to confine ions in all three dimensions rather than simply the x-y plane as in an LIT (11). Relevant differences between 2D and 3D traps will be noted where appropriate, but specific characteristics will be aimed at LITs. For a more complete explanation of quadrupolar fields refer Quadrupole Ion Trap Mass Spectrometry by March and Todd (12) as well as Linear Ion Trap Mass Spectrometry by Douglas et al. (13).

In order to effectively confine ions within the device, ion traps make use of oscillating quadrupolar fields in which the field potential increases with the square of displacement from the origin (12). In particular, the potential for the two-dimensional quadrupolar field in the x-y axis of a typical linear ion trap can be modeled by the equation (Equation 2.1):

$$\phi_{x,y} = A(\lambda x^2 + \sigma y^2) + C \quad (2.1)$$

Where  $\phi$  represents the electrical potential with respect to coordinates  $x$  and  $y$ ,  $\lambda$  and  $\sigma$  represent weightings for these respective coordinates,  $A$  represents a scalar to describe the magnitude of the potential independent of  $x$ - $y$  coordinates, and  $C$  represents a fixed DC potential applied to the entire field (and can be omitted if the device is held at ground). It is perhaps worth noting that additional terms for the  $z$  axis would be present in the case of 3 dimensional trapping in a 3D ion trap (12), but the field used in linear ions traps do not make use of a quadrupolar field to confine ions in that dimension. Further, note that the potential in each respective axis is independent of the other, a feature which becomes very useful when investigating ion behavior in the trap.

In general, a chargeless electric field must satisfy the Laplace equation (12) such that the second differential of the equation is equal to 0 or (Equation 2.2):

$$\nabla^2 \phi = \frac{\partial^2 x}{\partial t^2} + \frac{\partial^2 y}{\partial t^2} = 0 \quad (2.2)$$

as a result of charge free regions of space lacking any minima or maxima. Fortunately, the partial derivatives can be derived from Equation 2.1 as (Equation 2.3 & 2.4):

$$\frac{\partial^2 x}{\partial t^2} = 2\lambda A \text{ and } (2.4) \frac{\partial^2 y}{\partial t^2} = 2\sigma A \quad (2.3 \& 2.4)$$

for the  $x$  and  $y$  dimensions, respectively. These values can then be substituted into equation 2.2 to give (Equation 2.5):

$$\nabla^2 \phi = 2\lambda A + 2\sigma A = 0 \quad (2.5)$$

which can be simplified to the relationship (Equation 2.6):

$$\lambda + \sigma = 0 \text{ or } \lambda = -\sigma \quad (2.6)$$

Technically speaking, an infinite number of terms values could be used for  $\lambda$  and  $\sigma$  (12). However, as the magnitude of these two values will ultimately cancel out, the relative

polarity is really all that matters for a two dimensional field. Note that this is not necessarily the case for a 3D ion trap, as an additional parameter for the z axis is included. Given the relationship between these terms, we can use the values 1 and -1 for  $\lambda$  and  $\sigma$ , respectively, as is the case for nearly all commercial LITs. These numbers can be replaced into equation 2.1 to give the relationship (Equation 2.7):

$$\phi_{x,y} = A(x^2 - y^2) + C \quad (2.7)$$

In order to generate this field, a typical LIT consists of two pairs of opposing hyperbolic rods are spaced equidistant from their center, with  $r_0$  representing field radius or the distance from the center of the field to the nearest point on any of the rods (9). Technically speaking, one pair of rods could be moved further away than the other set of rods and generate a quadrupolar field, but both hyperbolic surface of the rods must still share an asymptote with adjacent rods in order to generate the field effectively (12). Given this design, the actual potential experienced by an ion,  $\phi_0$ , needs to be determined. As the potential felt by an ion is the potential difference between the x rods and y rods, it can be represented as (Equation 2.8):

$$\phi_0 = \phi_{x \text{ rods}} - \phi_{y \text{ rods}} \quad (2.8)$$

Given that we know the potential is constant along the electrode surfaces and we also know that we are at the x rod surface when  $x=r_0$  and  $y=0$  and at the y rod surface when  $y=r_0$  and  $x=0$ , we can use these facts in combination with equation 2.7 to determine that the potential applied to the surface of the rods has the relationship (Equations 2.9 & 2.10):

$$\phi_{x \text{ rods}} = A(r_0^2) + C \text{ and } \phi_{y \text{ rods}} = A(-r_0^2) + C \quad (2.9 \& 2.10)$$

When substituted back into equation 2.8, we can determine that (Equation 2.11):

$$\phi_0 = 2Ar_0^2 \quad (2.11)$$

which can be rearranged to (Equation 2.12):

$$A = \frac{\phi_0}{2r_0^2} \quad (2.12)$$

and subsequently substituted back into equation 2.7 to complete the relationship that (Equation 2.13):

$$\phi_{x,y} = \frac{\phi_0}{2r_0^2} (x^2 - y^2) + C \quad (2.13)$$

### ***Ion Stability***

Given the now established potential of an ion at a given coordinate, the force produced by an ion can effectively be determined. Because the ion motion in a given axis is decoupled from the others, the potential gradient in a given axis can be determined by taking the partial derivative with respect to that axis such that (Equation 2.14 & 2.15):

$$\frac{\partial \phi}{\partial x} = \frac{\phi_0 x}{r_0^2} \text{ and } \frac{\partial \phi}{\partial y} = -\frac{\phi_0 y}{r_0^2} \quad (2.14 \& 2.15)$$

Given that we know the potential gradient in a given axis, we can then determine the force on an ion in each axis such that (Equation 2.16 & 2.17):

$$F_x = -ez \frac{\phi_0 x}{r_0^2} \text{ and } F_y = ez \frac{\phi_0 y}{r_0^2} \quad (2.16 \& 2.17)$$

Where e is the magnitude of an elementary charge and z is the integer charge state of an ion within the field. Given that we know that force = mass x acceleration, the acceleration of an ion over time can be derived from the relationships (Equations 2.18 & 2.19):

$$-\frac{\phi_0 e z x}{r_0^2} = m \frac{d^2 x}{dt^2} \text{ and } ez \frac{\phi_0 y}{r_0^2} = m \frac{d^2 y}{dt^2} \quad (2.18 \& 2.19)$$

Which when rearranged give the acceleration in each axis (Equations 2.20 & 2.21):

$$\frac{d^2x}{dt^2} = -\frac{\phi_0 e z x}{m r_0^2} \text{ and } \frac{d^2y}{dt^2} = \frac{\phi_0 e z y}{m r_0^2} \quad (2.20 \text{ \& } 2.21)$$

Finally, having determined the acceleration of an ion in each axis, the potential of a real quadrupolar system can be incorporated into the equation. The potentials applied to the trap are given by (Equation 2.22):

$$\phi_0 = 2(U + V \cos \Omega t) \quad (2.22)$$

Where  $V$  is the peak-to-ground voltage amplitude applied to the rods at angular frequency  $\Omega$  and  $U$  represents the amplitude of a fixed DC potential applied as a positive voltage to the x-rods and a negative voltage to the y-rods. When substituted into equations 2.20 and 2.21, the final equations for the acceleration in the x and y axes can be given by (Equations 2.23 & 2.24):

$$\frac{d^2x}{dt^2} = -\frac{2(U + V \cos \Omega t) e z x}{m r_0^2} \text{ and } \frac{d^2y}{dt^2} = \frac{2(U + V \cos \Omega t) e z y}{m r_0^2} \quad (2.23 \text{ \& } 2.24)$$

As such, the trajectory of an ion within a quadrupolar field can be effectively modeled as a parametric oscillator governed by the periodic force of the RF trapping potential and the fixed force from the static DC offset. An effective tool for analyzing parametric oscillators is the Mathieu equation (13), the canonical version of which takes the form (Equation 2.25):

$$\frac{d^2u}{d\xi^2} + (a_u - 2q_u \cos 2\xi)u = 0 \quad (2.25)$$

where  $\xi$  represents the nondimensionalized time parameter, corresponding to  $\Omega t/2$  where  $\Omega$  is the oscillation frequency of the driving force (in this case the RF potential) and  $t$  is time. The terms  $q_u$  and  $a_u$  represent nondimensionalized amplitude terms for the oscillating driving force and any additional fixed force, respectively. These terms are

often referred to as “stability” or “trapping” parameters in the context of mass spectrometry as the stability – and therefore effective trapping – of a solution ultimately depends on the value of these terms.

The terms  $a$  and  $q$  can be given experimental significance by solving for their value in terms of the experimental variables in Equation 2.25. This can be done by replacing the term  $\xi$  with  $\Omega t/2$  to give (Equation 2.26):

$$\frac{d^2u}{dt^2} = -\left(\frac{\Omega^2}{4}a_u - \frac{\Omega^2}{2}q_u \cos t\right)u \quad (2.26)$$

Setting equation 2.26 equal to equations 2.23 and 2.24, we can determine the relationship (Equation 2.27):

$$-\frac{2(U + V\cos\Omega t)ezx}{mr_0^2} = -\left(\frac{\Omega^2}{4}a_x - \frac{\Omega^2}{2}q_x \cos \Omega t\right)x \quad (2.27)$$

for the x dimension and the relationship (Equation 2.28):

$$\frac{2(U + V\cos\Omega t)ezy}{mr_0^2} = -\left(\frac{\Omega^2}{4}a_y - \frac{\Omega^2}{2}q_y \cos \Omega t\right)y \quad (2.28)$$

for the y dimension. These sets of equations can be used to set  $a$  and  $q$  in terms of instrumentally relevant values such that (Equation 2.29):

$$a_x = -a_y = \frac{8ezU}{mr_0^2\Omega^2} \quad (2.29)$$

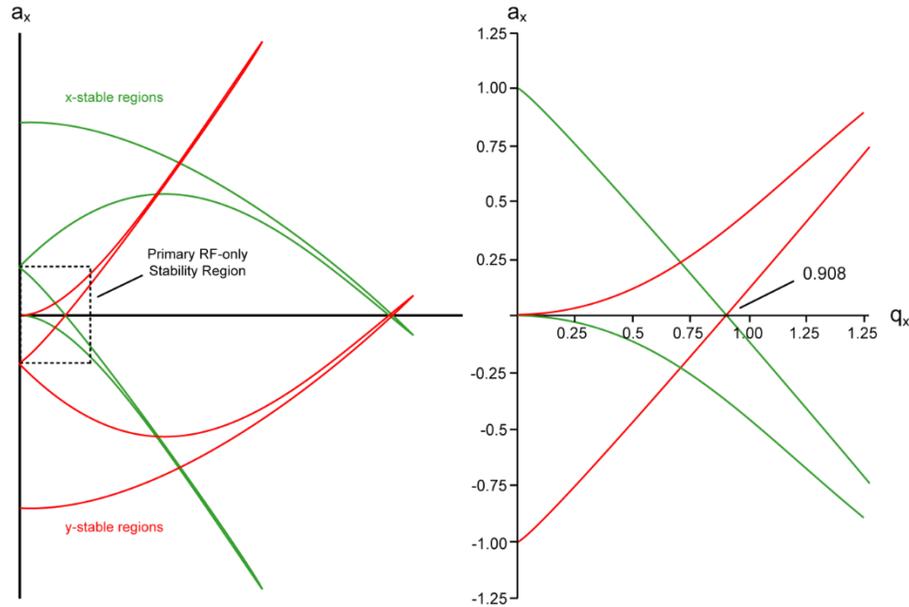
And (Equation 2.30):

$$q_x = -q_y = -\frac{4ezV}{mr_0^2\Omega^2} \quad (2.30)$$

Using these equations, an ion’s behavior can be easily evaluated with respect to the Mathieu  $q$  and  $a$  terms given known instrumental values – namely the RF and DC

amplitudes, the trapping RF frequency, the  $r_0$  of the ion trap, and the  $m/z$  of a particular ion.

Having determined the equivalence between the terms in the Mathieu equation and the instrumental variables, one needs to ultimately determine the stability or instability of ion trajectories with respect to these terms. The stability of an ion within the field can be derived by solving the Mathieu equation, solutions to which will generally fall into one of two categories – periodic and stable or periodic but unstable (13). Naturally, this translates experimentally to stable solutions describing ions which remain confined within the ion trap and unstable solutions describing ions which are not confined. These sets of solutions can be mapped with respect to the Mathieu parameters  $q$  and  $a$  to generate a diagram mapping the stability of an ion in a given dimension with respect to these parameters, as shown in Figure 2.1. As an ion must be stably trapped in both axes, the stability plots of both the  $x$  and  $y$  dimensions must be superimposed on one another, to give the set of  $a$  and  $q$  values in which the ion remains confined in both dimensions.



**Figure 2.1: Stability diagram for x-stable and y-stable ion trajectories in  $Q_x$  space.** Ions must remain stable in both x and y dimensions to be stably confined within the trap. Regions of x-stability and y-stability (left) are largely non-overlapping and overlap in a few select areas. The primary stability region close to the origin is shown (right) along with the stability limit at  $Q = 0.908$  of this region. Adapted from (12).

### *Secular Frequency and Resonant Excitation*

Having established the stability of an ion in the trap, it also becomes important to understand the behavior of particular ions in order to perform manipulations that are selective based on an ion's properties. Ion motion inside the trap can be roughly divided into two components – the ion's low-frequency secular motion from oscillating in the parabolic quadrupolar potential and a much higher frequency oscillation derived from rapid oscillating RF used to generate the field. For the purposes of ion manipulation, an ion's fundamental secular frequency is of principal concern.

An ion's secular frequency in dimension  $u$  can be determined with respect to an additional parameter,  $\beta_u$ , which is itself relatable to the parameters  $q_u$  and  $a_u$ . The precise

equation for  $\beta_u$  is given by an extended fraction, but it can be reasonably approximated (12) by Equation 2.31:

$$\beta_u = \sqrt{a_2 + \frac{q_u^2}{2}} \quad (2.31)$$

which simply models the ion as oscillating in a fixed pseudopotential rather than the time-variant RF actually used on the trap. This is a reasonable approximation when the trapping frequency is sufficiently fast when compared to an ion's natural oscillatory frequency, and thus works best at lower  $q$  values ( $<0.4$ ) and values where  $a < q$ . Having determined the  $\beta$  value for a particular ion, the parameter may then be related to the ion's secular frequency (14),  $\omega_{u,n}$ , by Equation 2.32:

$$\omega_{u,n} = \left( n + \frac{1}{2} \beta_u \right) \quad (2.32)$$

where  $\Omega$  is once again the frequency of the trapping RF, and  $n$  is the order of the secular frequency. Although higher order secular frequencies exist, as may be inferred by the equation, only the fundamental secular frequency possesses utility for typical operation of the instrument.

This frequency of oscillation can be advantageously coopted for a variety of purposes through the use of a process known as resonant excitation (15). Resonance is a property of periodic systems which occurs when either the drive frequency or an auxiliary frequency is applied at the oscillator's natural frequency under a given set of conditions – in the case of trapped ions, the ion's fundamental secular frequency (16). When these frequencies match, the oscillating ion will absorb significantly more energy due to constructive interference with the applied frequency, dramatically increasing its amplitude of oscillation. Doing so has a number of implications on the particular ion's

behavior. Primarily, increased amplitude of oscillation at a greater frequency is naturally going to increase the ion's total displacement as well as its velocity(13). This increased displacement will result in additional buffer gas collisions (which may result in collisional dissociation) and may be sufficient to eject the ion from the trap (17).

Performing resonant excitation in a linear ion trap (as is the case in this work) is most commonly applied via dipolar resonance on one pair of opposing rods within the trap(15, 18, 19). That is, additional, opposing potentials are applied to the rods opposite one another, and these potentials oscillate at the frequency of the ion being excited. Technically, the choice of rods is arbitrary, but the most common occurrence is the x-rods as those rods are also most typically used for ion ejection and detection (9).

### *Gas phase reaction kinetics*

The ability to appropriately perform gas phase ion reactions is inextricably linked to a suitable understanding of the reaction kinetics when performing one of these reactions. Without the ability to accurately predict reaction behavior in a given experiment, often sub-optimal results are achieved due to under- or over-reaction of precursor ions.

Ion-ion molecules can be effectively modeled using a simple capture collision model as outlined by Stephenson et al (7). Briefly, this theoretical framework models the formation of the ion-ion collision complex, the rate limiting step in both ion-ion and ion-molecule reactions, by predominantly considering the impact of the Coulombic attraction for forming collision complexes at long distances. The effective attractive potential can be related to the relative energy of the collision pair  $\left(\frac{1}{2}uv^2\right)$  to ultimately derive the

effective collision radius under which the ions will be sufficiently attracted to form a collision complex. This radius of separation can be expressed as Equation 2.33:

$$r_c = \frac{Z_1 Z_2 e^2}{2uv^2} \quad (2.33)$$

which when treated as the collisional radius can be used to derive the collisional cross section of an ion-ion complex gives Equation 2.34:

$$\sigma_c = \pi(2r_c)^2 = \left(\frac{\pi}{2}\right) \left[\frac{Z_1 Z_2 e^2}{uv^2}\right]^2 \quad (2.34)$$

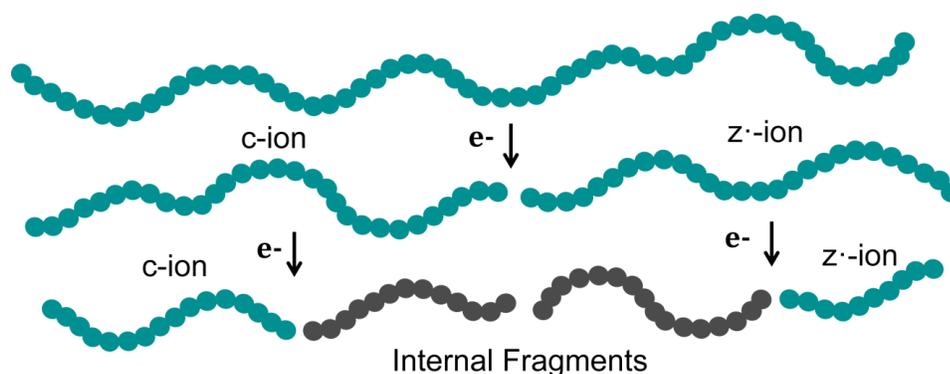
and ultimately the rate constant for ion-ion reaction complex formation (Equation 2.35):

$$k_c = v\sigma_c = v \left(\frac{\pi}{2}\right) \left[\frac{Z_1 Z_2 e^2}{uv^2}\right]^2 \quad (2.35)$$

where  $v$  is the relative velocity of the reactions in cm/s,  $u$  is the reduced mass of the collision pair in grams,  $Z_1$  and  $Z_2$  are the integer charge states of the cation and anion, respectively,  $e$  is the charge of an electron in electrostatic units. Notably, this expression of reaction kinetics is ultimately predictive of the rate at which ions will form collision pairs. However, it does not incorporate the probability of a reaction occurring in a given ion-ion collision complex, making the model underdetermine the exact quantitative rate of a given ion-ion reaction. Nevertheless, this theoretical framework allows for incredibly useful predictions, both in terms of approximating appropriate reaction times (20) and manipulating ion kinetics (8).

First, note that in equations 2 and 3, the collision cross section and therefore rate constant possess a charge-squared dependence with the charge of both rate cation and the anion. This relationship has significant implications for the progression of a typical ETD fragmentation event. Although typical ETD reagent anions only possess a single charge,

the use of electrospray ionization generally produces multiply charged precursor ions (21). As a result, the requisite reaction times to generate equivalent fragmentations can vary dramatically between precursors, potentially leading to orders of magnitude difference in reaction rate when comparing species of significantly different charge states. However, this feature of the reaction kinetics has a further implication. When performing ETD on a large peptide or protein, protons are likely to be distributed roughly evenly across the molecule. Particularly for cleavages closer to the termini, the protein may fragment and generate a small ion with only one or two protons as well as a large, highly charged fragment ion containing the majority of the protons from the intact molecule, as depicted in Figure 2.2. However, as a result of their high charge state, the relative reaction rate remains fairly close to the precursor molecule (7, 20, 22). For example, a +29 charge state fragment ion derived from the fragmentation of a +30 charge state parent ion will react at 93.4% the speed of the precursor molecule. Ultimately, this results in large fragment ions being very difficult to preserve due to capturing a second electron and fragmenting, destroying useful sequence information.



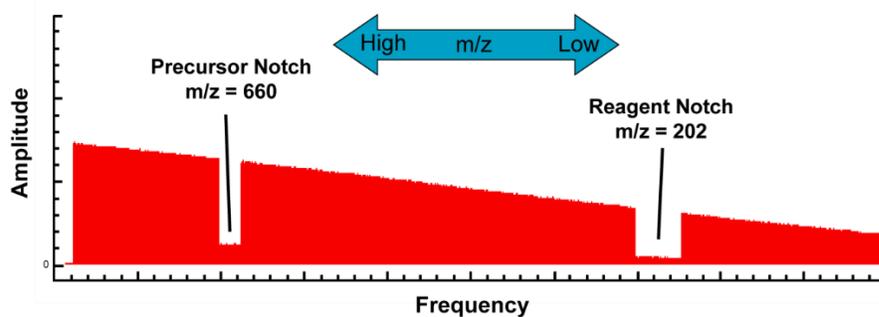
**Figure 2.2: Internal fragment ion generated from secondary electron transfers.** Additional ETD events cause second fragmentation events, destroying large fragment ions and generating smaller c/z ion and a sequence-useless internal fragment ion.

A more promising aspect of the ion kinetics, however, is inverse cubed relationship with velocity(7, 23). Such a strong relationship suggests that small changes in reactant velocity can lead to dramatic changes in the reaction rate of the various ion partners. In particular, slightly increasing the velocity of any of the reaction species will cause its reaction rate to plummet. Capitalizing on this feature has significant potential given the capacity to selectively inhibit the reaction rate of informative precursors while allowing uninformative ions to remain unperturbed and allowed to continue reacting.

### ***Parallel Ion Parking***

Given the aforementioned limitations associated with ETD reaction kinetics, it is imperative to pursue strategies in order to minimize the number of secondary fragmentation events to maximize the potential to identify target analytes. A strategy to counteract this problem was first demonstrated by McLuckey et al. in 2002 (8). They demonstrated that the application of resonant potentials to the end-caps of a 3D ion trap provided sufficient kinetic energy to significantly reduce the reaction rate of a given molecule without leading to ejection or fragmentation. This is largely the result of the strong velocity-cubed dependence of the ion-ion reaction kinetics, allowing small changes in velocity to generate large changes in reaction rate (23). Further, applying selective resonant excitation, as is enabled by an ion trap, allows one to reduce the reaction rate of some species within the trap while enabling others to remain largely unperturbed. The use of such excitation to slow ion-ion reaction rates is known as “ion parking.”

Limiting the reaction rate of a single species is insufficient for controlling ETD kinetics, however (8). Fragmentation of any given species is going to generate many different fragment ions which will each reside at a unique mass-to-charge ratio (24). They therefore all possess different frequencies for resonant excitation. Further, as the identity of a given species is unknown, the appropriate parked  $m/z$  cannot be determined prior to conducting the actual fragmentation. It is therefore necessary to expand the scope of excitation beyond that of a single frequency. In order to excite multiple frequencies, a broadband waveform is applied to the ion trap. A wide variety of frequencies are included in this ensemble with two major omissions – the frequencies corresponding to the precursor and the reagent ion are generally omitted to allow the reaction rate of these two species to proceed largely unperturbed (6). The Fourier transform of an ETD parking waveform is shown in Figure 2.3.

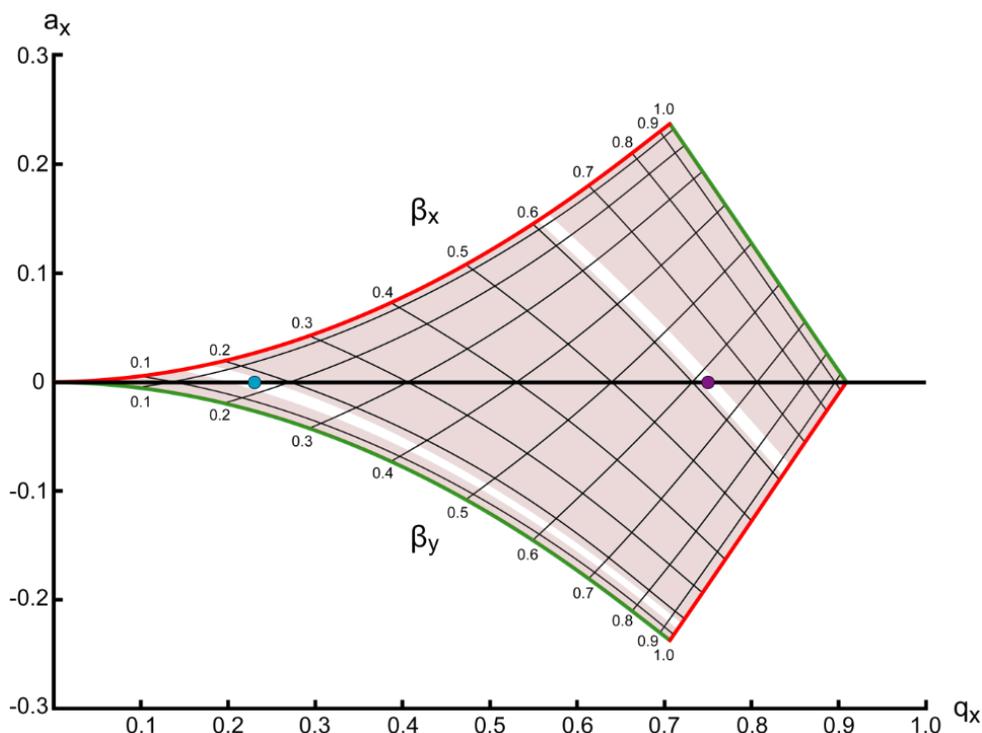


**Figure 2.3: Schematic Fourier transform of an ETD parking waveform.** The +13 charge state of ubiquitin ( $m/z = 660$ ) was used as the precursor; the location of the precursor notch will vary depending on the precursor  $m/z$ .

The application of the aforementioned waveform resonantly excites and consequently slows the reaction rate of all ions which are sufficiently close to the frequencies included within the waveform(8). A visualization of this excitation with respect to Mathieu stability parameters is pictured in Figure 2.4. The lines denoted on the

axes of the plot are the iso-beta lines, at which ions will share a secular frequency within a particular axis. Shaded regions denote  $q$  and  $a$  parameters at which ions will experience reduced ion-ion reaction kinetics, notably following the  $\beta x$  lines as excitation in a linear ion trap is typically performed on the x-rods of the trap.

Rather than using the end-cap electrodes of a 3D ion trap as in McLuckey's experiments, ion parking may also be reasonably implemented using either pair of rods of a linear ion trap. This enables a much greater ion storage capacity, improving fragment ion detection in the resulting MS2 experiment and increasing sequence coverage (9, 12). Further, such devices can be coupled to higher resolution mass analyzers to further improve data relative to what is achievable on an ion trap alone (25, 26). To this end, this chapter illustrates how the application of pipETD on such an instrument can be used to achieve unprecedented sequence coverage on "intact" antibody subunits generated by a highly specific enzymatic digestion (27, 28) and characterized on a chromatographic timescale.



**Figure 2.4: Depiction of slowed ion kinetics via ion parking with respect to ion stability in a linear ion trap.** The theoretical stability parameters of fluoranthene ( $m/z = 202$ ) and ubiquitin+13 ( $m/z = 660$ ) when the reagent is held at  $Q = 0.75$  are denoted on the figure. Shaded regions denote regions of the stability plot at which ions would be excited and parked.

## 2.3 Materials and Instrumentation

### Agilent Technologies (Palo Alto, CA)

1100 Series high performance liquid chromatograph

1100 Series vacuum degasser

### Eppendorf (Hauppauge, NY)

5414R Benchtop centrifuge

### GTS-Welco (Allentown, PA)

10 ppm SF<sub>6</sub> in nitrogen gas

### 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

**Sigma Aldrich (St. Louis, MO)**

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

Glacial acetic acid,  $\geq 99.99\%$  trace metal basis

Fluoranthene,  $>99\%$  purity

2-propanol, LC-MS grade

Tris(2-carboxyethyl)phosphine,  $>98.0\%$  purity

**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  $\mu\text{m}$

Formic Acid, LC-MS Grade

Orbitrap Fusion™ Tribrid™ Mass Spectrometer

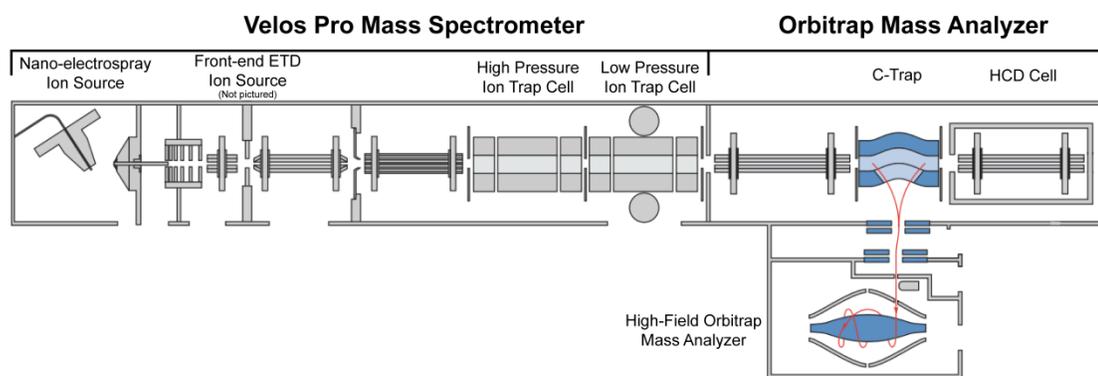
Pierce® Water, LC-MS Grade

Urea,  $>99.0\%$  purity

## 2.4 Methods

### *Apomyoglobin Standard Preparation and Analysis*

Prior to characterization of antibody samples, an analysis of a standard protein was first used to optimize the fragmentation parameters to be used for the antibody subunits. A 1 nmol sample of apomyoglobin was reconstituted at a concentration of 1 pmol/ $\mu$ L 1 mL of 40% water, 60% acetonitrile, and 0.5% acetic acid. During infusion, the +25 charge state of apomyoglobin ( $\sim$ 679  $m/z$ ) was isolated and subjected to pipETD. ETD parking waveform parameters were tuned to simultaneously achieve near-complete precursor consumption, reaction arrest predominantly in the first charge-reduced precursor, and minimal collisional fragmentation in the resulting spectrum. ETD ion-ion reactions were performed such that the Mathieu Q of the fluoranthene reagent ion ( $m/z = 202$ ) was held at 0.75. All frequencies were parked except those corresponding to  $m/z$  40 below and 10 above the precursor  $m/z$  as well as  $\pm 10$  around fluoranthene ( $m/z = 202$ ).



**Figure 2.5: Instrument Schematic of a front-end ETD enabled Orbitrap Elite.** The location of the front-end ETD source is noticed but not pictures; normal Orbitrap Elites contain a back-end ETD source instead. Adapted from (29).

In order to evaluate the overall performance of fragmentation conditions, apomyoglobin was also evaluated for total sequence coverage. The +25 charge state of apomyoglobin (679 m/z) was fragmented using the optimized pipETD conditions and subsequently charge-reduced with IIPIT prior to mass analysis. Ion targets of  $1e5$  precursor ions and  $7e4$  reagent ions were used for the experiments. 15 multiple c-trap fills were collected before analyzing the ion population in the Orbitrap at 240,000 resolution in full profile mode, and 5 microscans were averaged to generate each analytical scan. A total of 37 analytical scans were then averaged and the resulting composite spectrum was evaluated for sequence coverage.

### ***NIST mAb Preparation and Analysis***

Stock samples of NISTmAb (30) were acquired from NIST at 10 mg/mL in 12.5 mM L-histidine HCl at pH 6. A 10  $\mu$ L aliquot of the stock sample was diluted in 70  $\mu$ L of 50 mM ammonium bicarbonate and combined with 100  $\mu$ L of immunoglobulin degrading enzyme (IdeS, Genovis) which has been reconstituted in 50 mM ammonium bicarbonate as well. The mixture was allowed to react at 37° C for 30 minutes and before being aliquoted into 20  $\mu$ g aliquots and dried to completeness in a centrivap. One aliquot was then dried in a centrivap to completeness and reconstituted in 10  $\mu$ L of 10 mM tris(2-carboxyethyl)phosphine and 8M urea and reacted for 12 minutes at 50° for 12 minutes to reduce all disulfide bonds present in the antibody. Following reduction, the sample was analyzed by LC-MS without further sample preparation.

The reduced antibody subunits were analyzed on an in-house modified Orbitrap Elite mass spectrometer. Six hundred fmol of antibody were loaded onto a 10 cm PLRP-s

reverse phase column (75  $\mu\text{m}$  inner diameter, 10 cm bed, 3  $\mu\text{m}$  diameter particles) and desalted by flowing 0.3% formic acid through the column at 50 bar for ~40 minutes. Peptides were then gradient eluted using a gradient of 30%-45%-100% solvent B in 5-45-50 minutes where solvent A was 0.3% formic acid in water and solvent B was 72% acetonitrile, 18% isopropyl alcohol, 10% water, 0.3% formic acid. The flow rate through the column during MS analysis was ~100 nL/min and peptides were ionized from a column-integrated nano-electrospray ionization tip at 2.2 kV.

The sample was first screened using only 120,000 resolution FTMS<sup>1</sup> scans to determine peak elution profile and retention time. Following this determination, a follow-up analysis was conducted in order to specifically target each peak during its elution window. FTMS profile scans were taken and each antibody subunit was selectively fragmented over its given elution peak. The charge states selected for fragmentation were the +30 for the heavy chain Fc ( $m/z = \sim 842$ ), the +28 for the light chain ( $m/z = 827$ ) and the +30 for the heavy chain Fd ( $m/z = 857$ ). Five multiple c-trap fills were collected prior to analyzing the ion population in the Orbitrap at 240,000 resolution in full profile mode, and no additional averaging was performed prior to the generation of each analytical scan. 33, 37, and 37 analytical scans were averaged for the heavy chain Fc, light chain, and heavy chain Fd' composite spectra, respectively.

## 2.5 Results and Discussion

### *Parking Parameter Evaluation with Apomyoglobin*

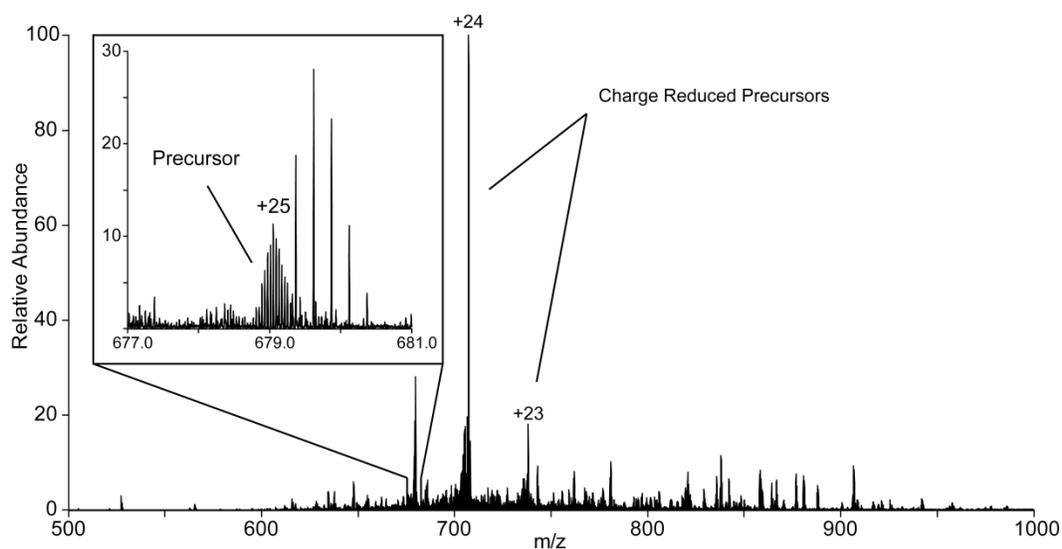
The key feature in performing a parallel ion parking experiment is to generate a rate difference between the desirable product ions and the undesirable parent ion to

maximize both precursor consumption and product ion retention. Multiple experimental variables may impact this result, however, and therefore most often be determined experimentally until they can be manipulated in a perfectly predictable manner (6, 8, 31). Unfortunately, in its current state, many of these experimental variables are inter-related, making changing one variable difficult as it may have unintended consequences on another. For example, increasing the amplitude of the entire waveform can result in a slight reduction in reaction rate, making the original reaction rate too slow even if an ideal rate difference is achieved (8, 23).

Given that a full-fledged calibration routine has not yet been developed for on-the-fly determination of parking parameters, performance must first be evaluated by tuning experimental parameters to achieve ideal performance on a standard protein prior to implementation in a chromatographic method. To this end, apomyoglobin was evaluated to optimize the performance of the parking waveform due to its size (~17 kDa) being roughly similar to the antibody subunits and requiring minimal sample preparation prior to direct infusion.

A variety of parameters were adjusted in order to optimize the reaction performance on apomyoglobin. Preliminarily, the amplitude of the parking waveform was evaluated to find the threshold at which it led to fragmentation of generated product ions, and reduced by about 10-20% to avoid these secondary fragmentation events. The boundaries of the frequency notch (containing the precursor ion and ideally remaining largely un-excited) were expanded until they ceased to have an impact on the precursor reaction rate. This is because, as originally characterized by Elizabeth Duselis, space-charge from the reagent ion cloud can lead to an unexpectedly large shift in the precursor

frequency, potentially shifting the precursor ions back into the parked region of the spectrum and halting their reaction. A lack of change implied that none of the frequencies removed would impact the experiment, making them safe to include. Finally, having eliminated the excessive reduction in reaction rate, the ETD reaction time was adjusted to maintain near-complete precursor ion consumption and verifying that the majority of the charge-reduced precursor ions continued to remain unreacted despite complete reaction of the precursor. The resulting spectrum following optimization is shown in Figure 2.6, illustrating near-complete precursor consumption and primary arrest of the reaction in the first charge-reduced precursor (+24,  $m/z = 707$ ).



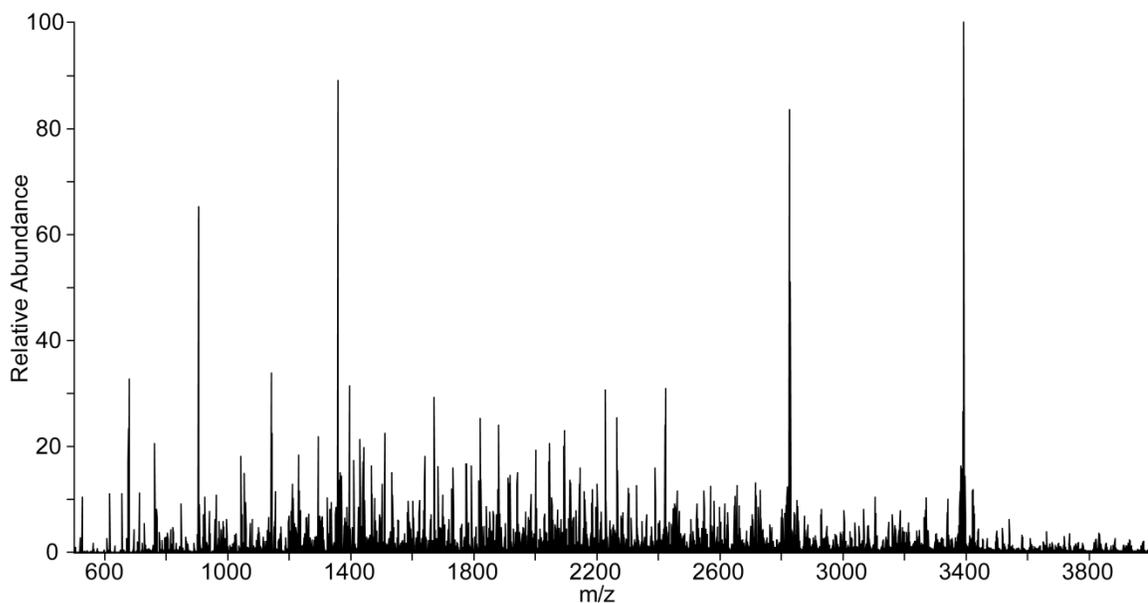
**Figure 2.6: ETD parking only Spectrum of Apomyoglobin<sup>+25</sup>.** The majority of charge-reduced precursor remains in the singly-reduced species (+24) despite near-complete precursor ion consumption, suggesting successful cessation of the reaction.

Following optimization of ETD parking parameters, the reaction conditions were evaluated in combination with other instrumental techniques to validate the generation of high sequence coverage and preservation of large fragment ions. Apomyoglobin was fragmented using the optimized pipETD reaction conditions and subsequently reacted

with 40 ms of ion-ion proton transfer to maximally resolve fragment ions. Multiple c-trap fills and signal averaging were also used for signal amplification to further improve spectral quality and generate maximal product ion information.

The resulting spectrum from this procedure is depicted in Figure 2.7 and was evaluated for sequence coverage. Cleavages observed in the spectrum are shown in Figure 2.8. The approach provided exceptional results, generating observable product ions corresponding to 95.9% of all ETD cleavable bonds within the molecule or 93.4% of all cleavages total. Further, fragment ions demonstrated significant overlap, often maintaining both fragment ions for a given bond cleavage even for fragment ions nearly identical in size to the intact molecule. In total, 81.2% of all possible fragment ions were observed in the spectrum, and fragment ions corresponding to the loss of small numbers of amino acids, for example z152, were readily apparent in the spectrum. This demonstrates a clear cessation of ETD fragmentation on product ions despite enabling the reaction of nearly the entire precursor ion population.

More work is still required to refine this method and make it more broadly applicable for on-the-fly analysis of complex mixtures. However, these experiments still clearly illustrate the efficacy of pipETD for generating highly sequence informative fragmentation spectra, and supports its utility when used on fairly large, subunit-sized peptides.



**Figure 2.7: Spectrum of pipETD/IPT Apomyoglobin.** Sequential reactions with 75 ms pipETD and 40ms IPT were used to fragment the +25 charge state of apomyoglobin ( $m/z = 659$ ) and charge-disperse the product ions.

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

**Figure 2.8: Simplified Apomyoglobin Coverage.**

### *NISTmAb Intact Subunit Analysis*

Having verified that the optimized pipETD reaction conditions can indeed maintain large fragment ions in the spectrum, these conditions were applied to analyze NISTmAb. However, sequence characterization of a native, intact antibody is quite

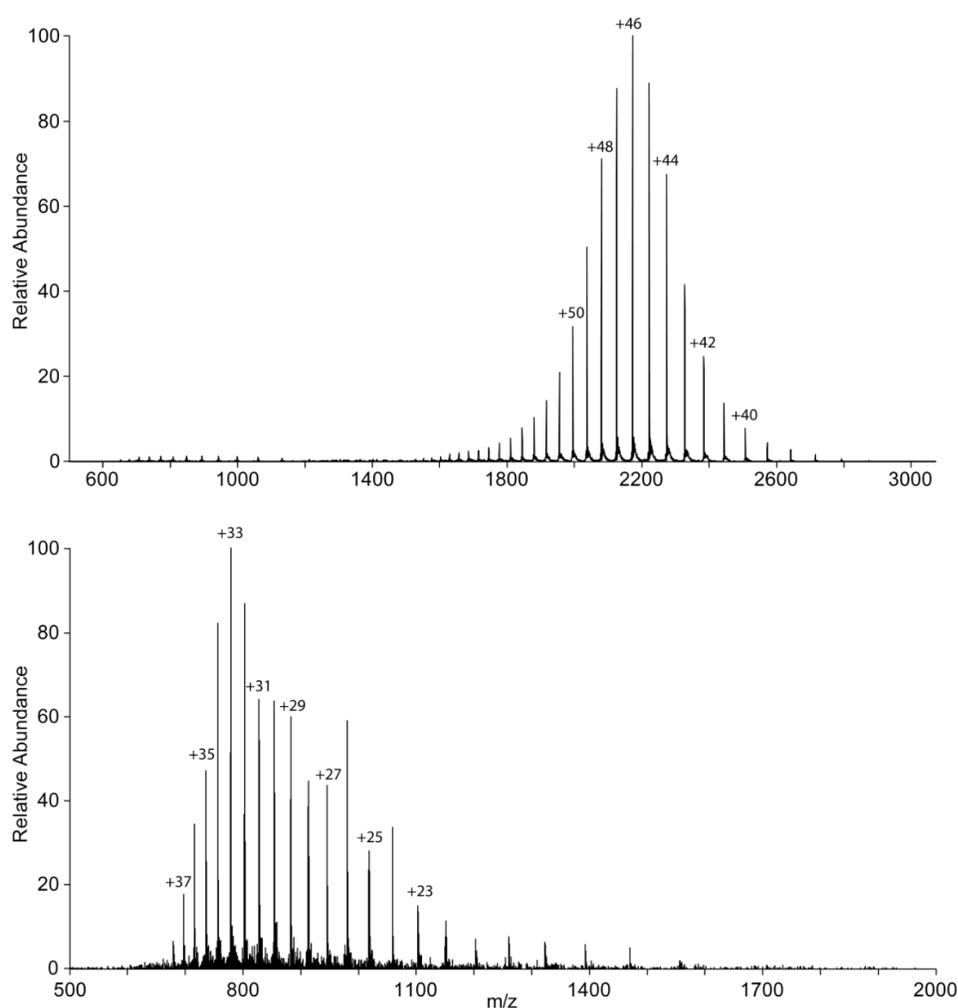
difficult for a variety of reasons. Primarily, it is a result of their large size. The average human protein weighs only ~46 kDa (32, 33), making an antibody's molecular weight of ~150 kDa significantly above the average. Further, antibodies tend to be fairly hydrophobic molecules and will remain somewhat folded even under extreme conditions due to their extensive disulfide networks (34, 35). When ionized using electrospray ionization, this has a tendency to produce much lower charge-state distribution, as illustrated on the F(ab')<sub>2</sub> in Figure 2.9 (35). This further limits the ability to characterize the molecule using electron-based fragmentation techniques like ETD, which often show the highest performance on intact molecules when not charge-depleted (36).

To combat this issue, a highly specific, limited digestion was used to segment the antibody into smaller subunits for analysis. Immunoglobulin degrading enzyme, or IdeS, is a highly specific cysteine protease (27). Normally aimed at inhibiting antibody-directed cell killing (27), the enzyme cleaves at the Fc portion of an antibody (below the inter-chain hinge region) in a highly selective manner (37). It produces a single, highly specific cleavage at the conserved CPAPPELLG|GGPSVFLFP sequence motif found in this region of the molecule (37). In the context of antibody sequencing, this cleavage segments the heavy chain roughly in half, generating the ~100 kDa F(ab')<sub>2</sub> and ~25 kDa Fc/2 subunits in an intact antibody. The f(ab')<sub>2</sub> piece, being composed of two light chains and two n-terminal heavy chain regions (Fd') disulfide bound together, can then be reduced to separate the ~25 kDa intact light chain and heavy chain Fd' subunits as well as eliminate inter-chain disulfide bonds (28, 38). In doing so, the molecule is separated into subunits that are significantly easier to analyze by virtue of both their lower mass(39) as well as much higher charge ESI charge state distributions.

Although these digestion and reduction procedures succeeded in simplifying the molecule in terms of size, they introduce a new dimension of complexity; rather than being composed of a singular protein species as in the case of apomyoglobin, the processed antibody sample contains a mixture of various protein subunits. This mixture makes analysis by infusion impractical as the charge state distributions of the various components are very like to fall close to one another, making isolation in terms of mass-to-charge ratio significantly more difficult. As a result, it becomes necessary to separate these pieces chromatographically so that the subunits may be analyzed individually, but this significantly constrains the simplicity of the analysis as each species can only be characterized over its limited elution window.

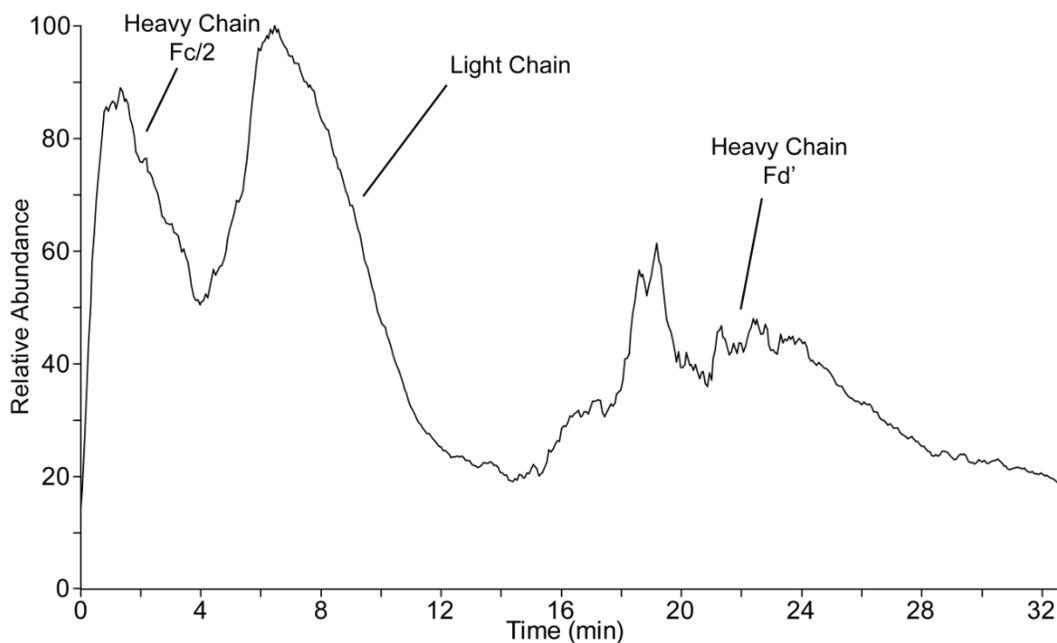
A further complication added by chromatographic separation is variable ion flux over time. In an infusion experiment, the intensity of a given ion tends not to fluctuate significantly, but the rising and falling edges of a chromatographic peak will have very different ion intensities than the center of the peak. This becomes a significant issue because of the way the instrument regulates ion injection into the ion trap. A rapid automatic gain control (AGC) pre-scan is typically used to estimate ion flux prior to injecting the full ion population for the normal analytical scan (40, 41). However, in their current implementation, signal amplifying techniques like multiple c-trap fills and microscans will occur without taking additional AGC scans over the course of the process (42). This means that the signal can change over the length of the peak, but the instrument will not appropriately compensate, potentially over or under injecting ions compared to the expected amount. Due to the space-charge dependent frequency shift noted previously, the inability to appropriately regulate these ion injections can lead to

inconsistencies in ion secular frequencies, potentially compromising the precursor-fragment ion rate difference at certain points in the peak due to space-charging the precursor frequency back to one which is being excited. To adjust for this, these averaging parameters were decreased for the chromatographic analysis, reducing multiple c-trap fills to 10 and eliminating additional microscans entirely. This is likely to negatively impact the final results relative to an infusion experiment but is necessary to successfully perform the ETD parking procedure.

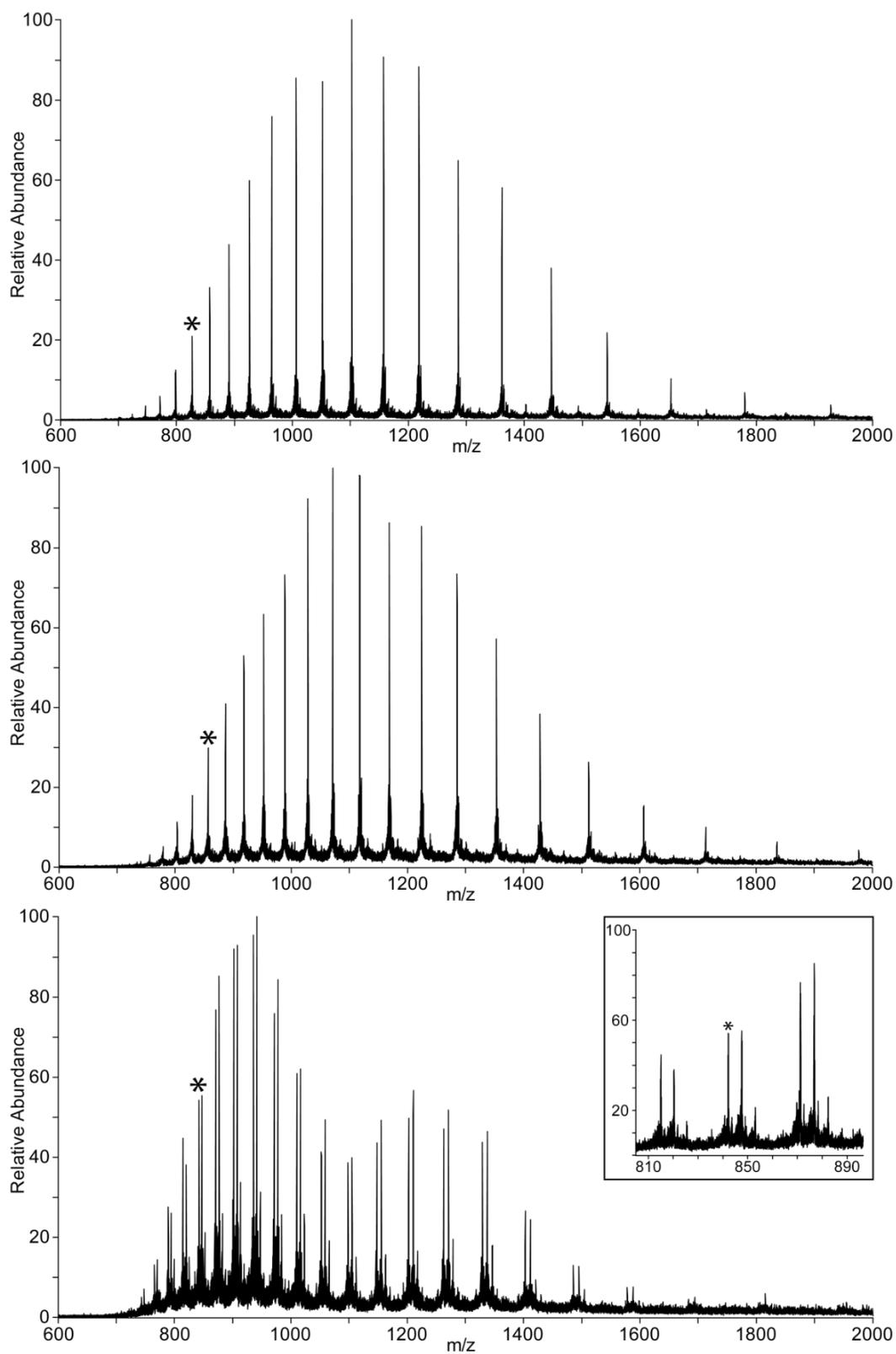


**Figure 2.9: Positive ESI charge-state distributions of the un-reduced F(ab')<sub>2</sub> (Top) and Reduced/NAEM alkylated heavy chain (Bottom).** Disulfide bond reduction generates a substantial increase in charge density by improving protein unfolding during ESI.

Following these modifications, the NISTmAb subunits were separated and analyzed in-line and selectively fragmented under their respective chromatographic peaks using both retention times and charge states determined by a prior screen of the sample. The analysis chromatogram, shown in Figure 2.10, demonstrates good separation between each of the three subunits, enabling relatively clean isolation and fragmentation of each species. Charge-state distributions for each of the three subunits are illustrated in Figure 2.11. Fortunately, each of the charge-state distributions contained relatively charged peaks which were suitable for ETD fragmentation. Starred species were isolated and fragmented by pipETD/IIPT over their respective chromatographic peaks. These correspond to the +28, +30, and +30 charge states of the light chain, heavy chain Fd', and heavy chain Fc/2, respectively. Multiple glycan variants were observed on the Fc subunit, and the presumed G0F glycan modified species was selected for fragmentation.

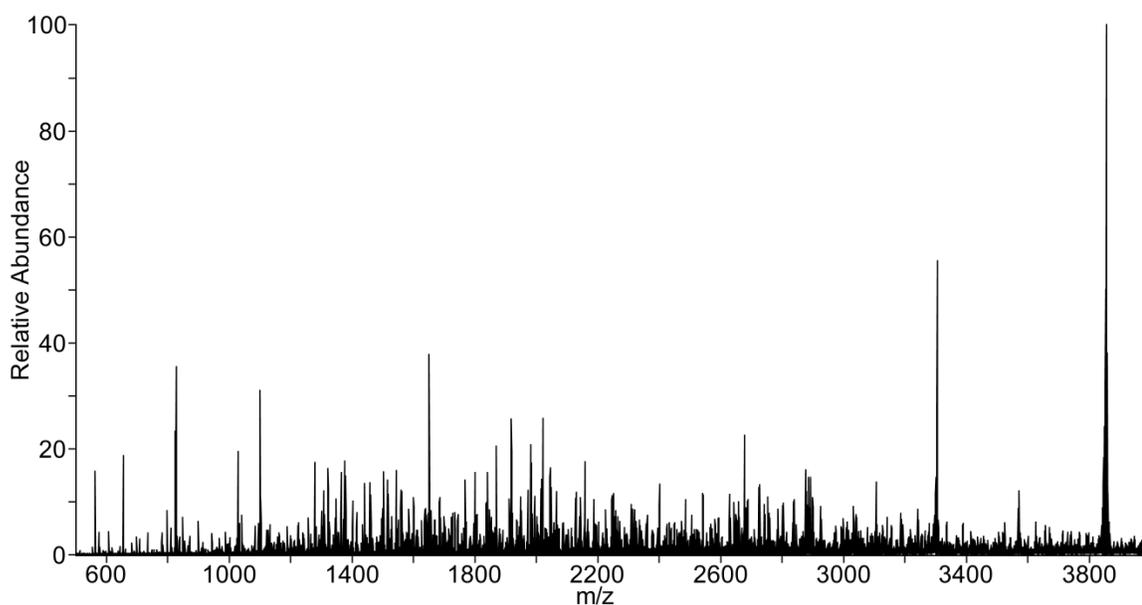


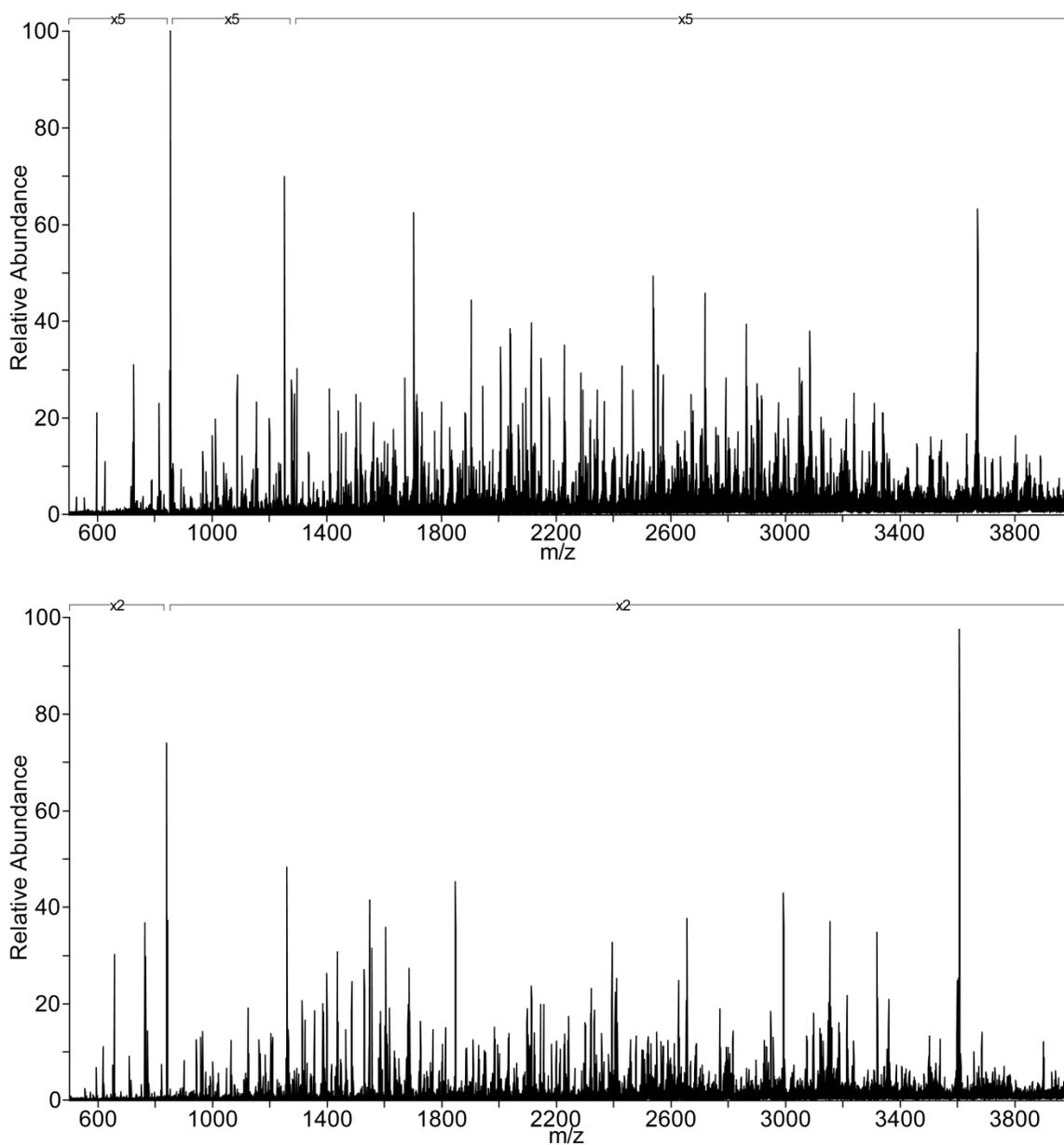
**Figure 2.10: Total Ion Chromatogram of from NIST Subunit Analysis.** Each major subunit is denoted on the chromatogram. Pre-elution column volume is omitted from the chromatogram in order to limit the resulting file size.



**Figure 2.11: MS1 spectra of NIST light chain (top), heavy chain Fd' (middle), and heavy chain Fc/2 (bottom). Isolated species are denoted with an asterisk.**

Averaged pipETD/IIPT spectra for each of the three proteins are shown in Figure 2.12. The spectra all show significant ion dispersion across the mass range rather than the peak clustering around the precursor, as seen in ETD-only spectra. Notably, each of the spectra only show a few charge-reduced precursors rather than a wider variety, as typically seen in IIPT spectra. This is largely the result of the extended IIPT time chosen, and a result of limiting the reaction of these species as in pipETD. Indeed, around half of the charge-reduced precursors were likely charge-reduced beyond the mass limit of 4000 m/z. While this risks losing some amount of useful fragment ion current as large fragment ions could be charge-reduced off of the mass range as well, it is likely to have a net benefit because it maximizes fragment ion dispersion (thereby improving the effective resolution of the scan), and the most abundance charge-reduced precursor likely remains within the trap. Therefore, the most abundant fragment ion charge states should as well.





**Figure 2.12: pipETD/IPT Spectra of NIST light chain (top), heavy chain Fd' (middle), and heavy chain Fc/2 (bottom)**

The sequence coverage maps derived from each of the fragmentation spectra are pictured in Figure 2.13. Sequence coverage derived from all three subunits was extensive, achieving 79.1%, 73.5%, and 88.5% of ETD cleavable bonds (or 75.0%, 67.6%, and 81.0% of total bonds) for the light chain, heavy chain Fd', and heavy chain Fc/2, respectively. Fragmentation achieved respectable but incomplete characterization of the

CDRs of both chains. Light chain CDRs generated sequence coverages of 73%, 100%, and 50% for CDR1, CDR2, and CDR3, respectively, while the heavy chain CDRs generated sequence coverages of 63%, 82%, and 91% for CDR1, CDR2, and CDR3, respectively. The Fc glycan was also successfully site localized to G63 of the Fc/2 (or G301 of the intact heavy chain) with clear fragmentation across the entire flanking region.

Surprisingly, fragmentation of the fab portion of the molecule resulted in substantially lower sequence coverage than that of the Fc/2. This may be partially explained by the significantly lower number of charged residues present on the light chain and heavy chain Fd'. These subunits contain 23 and 25 basic residues (counting lysine, arginine, and histidine) while the Fc/2 contains 31 basic residues while also being the smallest subunit of the three. Notably, most of the lost coverage actually occurs near the termini rather than the middle of the molecule. Typically, analysis of the termini of proteins is actually easier with ETD owing to the better S/N of the smallest fragment ions as well the tendency to degrade larger fragment ions into these smaller ions. Although the former reason should remain constant, the latter explanation suggests that parking spectra will actually generate lower abundances of these most terminal ions. Normally, this would not be a significant imposition, given the maximized sensitivity when analyzing these ions, but the amino acid sequences near the n-terminus of the light chain and c-terminus of the F(ab')<sub>2</sub> are largely devoid of charged residues, significantly reducing ETD fragmentation efficiency within this region of the protein. As such, the same kind of experiment would likely perform notably better on a much less charge-depleted molecule.

DIQMTQSPSTLSASVGDRTITCSASSRVGY  
**MHWYQQKPKAPKLLIYDTSKLASGVPSTRFS**  
 GSGSGTEFTLTISSLQPDDFATYYCFQGSY  
**PFTFGGGTKVEIKRTVAAPSVFIFPPSDEQL**  
 KSGTASVVCLLNNFYPREAKVQWKVDNALQS  
 GNSQESVTEQDSKDSSTYSLSSTLTLSKADYE  
 KHKVVYACEVTHQGLSSPVTKSFNRGEC

QVTLRESGPALVKPTQTLTLTCTFSGFSLS**T**  
**AGMSVGVIRQPPGKALEWLAD IWWDDKKHYN**  
**PSLKDRLTISKDTSKNQVVLKVTNMDPADTA**  
 TYYCAR**DMIFNFYFDV**WGQGTTVTVSSASTK  
 GPSVFPPLAPSSKSTSGGTAALGCLVKDYFPE  
 PVTVSWNSGALTSGVHTFPAVLQSSGLYSLS  
 SVVTVPSSSLGTQTYICNVNHKPSNTKVKDKR  
 VEPKSCDKTHTCPPCPAPELLG

GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSD  
 HEDPEVKFNWYVDGVEVHNAKTKPREEQYNS  
 TYRVVSVLTVLHQDWLNGKEYKCKVSNKALP  
 APIEKTIISKAKGQPREPQVYTLPPSRREEMTK  
 NQVSLTCLVKGFYPSDIAVEWESNGQPENNY  
 KTTTPVLDSDGSFFFLYSKLTVDKSRWQQGNV  
 FSCSVMHEALHNHYTQKSLSLSPGK

**Figure 2.13: Sequence coverage from pipETD/IPT spectra of the NIST light chain (top), heavy chain Fd' (middle), and heavy chain Fc/2 (bottom). CDRs on the light chain and Fd' as well as the glycan modification site on the Fc/2 are denoted in bold.**

In comparison to the apomyoglobin benchmark experiments, significantly fewer of the largest subset of fragment ions (e.g. ions losing only one or two amino acids) were observed. This is likely a product of the sequences in the most terminal portions of the molecule. Not only do these portions of the molecule generally exhibit lower sequence coverage, as noted earlier, but the small corresponding *c/z* fragment ions are often very low in abundance, unlike in the more highly charged apomyoglobin. Therefore, the most likely scenario is that, while these fragment ions were likely formed and largely preserved, their fragment ion yield was low enough that the signal was sufficiently diluted across charge states and isotopes such that they were no longer above the limit of detection (unlike their smaller counterparts which are consolidated into a comparatively small number of peaks).

Nonetheless, the antibody sequence coverage achieved through these experiments is substantially better than equivalent experiments using alternative fragmentation techniques or standalone ETD (without ion parking) (43–45). While imperfect, these results suggest a significant improvement in the ability to analyze antibody subunits following a very limited digestion, strongly suggesting the utility of this fragmentation approach on large protein subunits. Moreover, advancements to both pipETD as well as mass spectrometric instrumentation in general are likely to improve these results further, extending the analyze size range that can be comprehensively mapped by pipETD fragmentation.

## References

1. Wu C, et al. (2012) A protease for “middle-down” proteomics. *Nat Methods* 9(8):822–824.
2. Smith LM, Kelleher NL (2013) Proteoform: A single term describing protein complexity. *Nat Methods* 10(3):186–187.
3. Mikesh LM, et al. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta - Proteins Proteomics* 1764(12):1811–1822.
4. Riley NM, Coon JJ (2018) The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal Chem* 90(1):40–64.
5. Riley NM, Westphall MS, Coon JJ (2015) Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal Chem* 87(14):7109–7116.
6. 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.
7. 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.
8. McLuckey SA, Reid GE, Wells JM (2002) Ion parking during ion/ion reactions in electrodynamic ion traps. *Anal Chem* 74(2):336–346.
9. Schwartz JC, Senko MW, Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 13(6):659–669.
10. March RE (1997) An introduction to quadrupole ion trap mass spectrometry. *J Mass Spectrom* 32(4):351–369.
11. Wong PSH, Cooks GR (1994) Ion Trap Mass Spectrometry. *Curr Sep Drug Dev* 16/3(1):237–76.
12. March R, Todd J (2005) *Quadrupole ion trap mass spectrometry* ed Winefordner JD (John Wiley & Sons, Ltd, Hoboken, New Jersey). 2nd Ed. Available at: <https://books.google.com/books?hl=en&lr=&id=JDKCoIshgSIC&oi=fnd&pg=PR5&dq=March+Todd+ion+trap&ots=GPfulPzMkQ&sig=7jjegYFkS8dmaa3b4qbraw8sFo> [Accessed April 19, 2019].
13. Douglas DJ, Frank AJ, Mao D (2005) Linear ion traps in mass spectrometry. *Mass Spectrom Rev* 24(1):1–29.
14. Douglas DJ (2009) Linear quadrupoles in mass spectrometry. *Mass Spectrom Rev* 28(6):937–960.

15. Stafford GC, Kelley PE, Syka JEP, Reynolds WE, Todd JFJ (1984) Recent improvements in and analytical applications of advanced ion trap technology. *Int J Mass Spectrom Ion Process* 60(1):85–98.
16. Glish GL, Vachet RW (2003) The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov* 2(2):140–150.
17. Charles MJ, McLuckey SA, Glish GL (1994) Competition between resonance ejection and ion dissociation during resonant excitation in a quadrupole ion trap. *J Am Soc Mass Spectrom* 5(12):1031–1041.
18. Snyder DT, Cooks RG (2017) Single Analyzer Precursor Ion Scans in a Linear Quadrupole Ion Trap Using Orthogonal Double Resonance Excitation. *J Am Soc Mass Spectrom* 28(9):1929–1938.
19. Zhang X, et al. (2016) Reducing Space Charge Effects in a Linear Ion Trap by Rhombic Ion Excitation and Ejection. *J Am Soc Mass Spectrom* 27(7):1256–1262.
20. 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.
21. FENN JB, MANN M, MENG CK, WONG SF, WHITEHOUSE CM (2016) ChemInform Abstract: Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *ChemInform* 21(5). doi:10.1002/chin.199005359.
22. McLuckey SA, Stephenson JL (1998) Ion/ion chemistry of high-mass multiply charged ions. *Mass Spectrom Rev* 17(6):369–407.
23. 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.
24. 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* 101(26):9528–9533.
25. Hunt D, Shabanowitz J, Yates J, ... RM-A, 1985 U Tandem-quadrupole Fourier transform mass spectrometry of oligopeptides. *ACS Publ*. Available at: <https://pubs.acs.org/doi/pdf/10.1021/ac00290a065> [Accessed April 22, 2019].
26. McAlister GC, et al. (2008) A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J Proteome Res* 7(8):3127–3136.
27. 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.
28. Chevreux G, Tilly N, Bihoreau N (2011) Fast analysis of recombinant monoclonal

- antibodies using IdeS proteolytic digestion and electrospray mass spectrometry. *Anal Biochem* 415(2):212–214.
29. Michalski A, et al. (2011) 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.
  30. Schiel JE, et al. (2018) The NISTmAb Reference Material 8671 value assignment, homogeneity, and stability. *Anal Bioanal Chem* 410(8):2127–2139.
  31. Chrisman PA, Pitteri SJ, McLuckey SA (2006) Parallel ion parking of protein mixtures. *Anal Chem* 78(1):310–316.
  32. Brocchieri L, Karlin S (2005) Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res* 33(10):3390–3400.
  33. Senko MW, Beu SC, Mclafferty FW Determination of Monoisotopic Masses and Ion Populations for Large Bio... (Averagine algorithm).pdf. Available at: <https://link.springer.com/content/pdf/10.1016%2F1044-0305%2895%2900017-8.pdf> [Accessed April 22, 2019].
  34. Betz SF (1993) Disulfide bonds and the stability of globular proteins. *Protein Sci* 2(10):1551–1558.
  35. Loo JA, Edmonds CG, Udseth HR, Smith RD (1990) Effect of Reducing Disulfide-Containing Proteins on Electrospray Ionization Mass Spectra. *Anal Chem* 62(7):693–698.
  36. 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* 330–332:174–181.
  37. An Y, Zhang Y, Mueller HM, Shameem M, Chen X (2014) A new tool for monoclonal antibody analysis Application of IdeS proteolysis in IgG domain-specific characterization. *MAbs* 6(4):879–893.
  38. Srzentić K, et al. (2014) Advantages of Extended Bottom-Up Proteomics Using Sap9 for Analysis of Monoclonal Antibodies. *Anal Chem* 86(19):9945–9953.
  39. 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.
  40. Makarov A, et al. (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* 78(7):2113–2120.
  41. Syka JEP, et al. (2004) Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J Proteome Res* 3(3):621–626.

42. Earley L, et al. (2013) Front-end electron transfer dissociation: A new ionization source. *Anal Chem* 85(17):8385–8390.
43. 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. *Journal of the American Society for Mass Spectrometry*, pp 827–838.
44. 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.
45. Fornelli L, et al. (2018) Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal Chem* 90(14):8421–8429.

## **ETD Parallel Ion Parking Coupled to Size-Controlled Proteolysis for Antibody CDR Characterization**

### **3.1 Introduction**

Although the aforementioned subunit analysis of antibodies is indeed a powerful approach, the total sequence coverage remains insufficient to unambiguously identify all residues in the CDRs. As a result, it becomes necessary to generate a smaller peptide containing only the variable region (rather than the entire Fab) in order to comprehensively analyze and co-localize all of these CDRs to a particular antibody chain. However, as no highly specific enzyme like IdeS can perform that kind of cleavages, it becomes necessary to pursue alternative strategies in order to generate the peptide of interest.

Typical proteomics pipelines make use of complete digestion with a specific protease in order to digest the protein into more manageable pieces. The most commonly used enzyme for this task is trypsin (1). Trypsin is a serine protease which hydrolyzes peptide bonds c-terminal to lysine or arginine (2). Given the relative abundance of these amino acids across the proteome, tryptic digests normally create quite small peptides, approximately 1 kDa on average (3).

Unfortunately, complete digestion with such an enzyme has a number of limitations. For one, co-localizing multiple sequence variant portions of a protein like antibody CDRs becomes highly implausible, as a peptide is very unlikely to be generated which spans a long enough portion of the protein to contain both relevant regions (1, 4). Further, sufficiently small peptides generally are not retained on reverse phase LC columns, causing them to be lost and remain undetected (1, 3). However, these problems

would be mitigated if the peptides generated by a digest were larger on average. As a result, increasing efforts are aimed at using modified digestion schemes to generate significantly larger peptides during an enzymatic digestion.

## 3.2 Background

### *Conventional Strategies for Generating Larger Peptides*

Given the attractive qualities of using enzymatic digestions to generate longer peptides, a variety of techniques have been employed in order to accomplish this task. One of the most common techniques for doing this is by performing a complete digestion with a more specific protease. As previously noted, the cleavage motif of trypsin is actually quite common across the proteome, resulting in small peptides. Naturally, a way to extend the length of these peptides is to use a protease that cleaves at far fewer sites. The most straightforward comparison is a protease like Lys-C, one of the more common alternatives to trypsin (1, 5). Lys-C cleaves c-terminally to lysine only instead of both lysine and arginine (5). Simply by eliminating the possibility for arginine cleavages, the average peptide length increases to ~1.6 kDa. Enzymes with even further limited specificity can increase this size even further. For example OmpT, an enzyme which cleaves selectively between two adjacent basic residues (lysine/arginine), generates proteins that are >6 kDa on average (6).

Increased cleavage specificity comes at a cost, however. As the enzyme cleavage sites become less frequent, the peptide size distribution tends to broaden, generating a plethora of differently sized peptides(3, 6); the decreased frequency in the particular cleavage motif and random variation in amino acid sequence ultimately result in a much

large variation in peptide size. For example, when used to analyze the entire proteome, trypsin typically generates peptides such that the overwhelming majority will still fall below 3 kDa in mass (3). Conversely, OmpT digestion generates peptides ranging from 1 kDa to nearly 14 kDa (6). A necessary clarification here is that these widely size variable peptides are also not overlapping, as the digestion is still progressing to completion, the variation in cleavage site distance is just incredibly wide. As a result, particularly large or small peptides may evade detection by exceeding the scope of LC column retention (3) or mass spectrometer performance capabilities (7, 8), respectively. They therefore still risk failing to detect biologically relevant portions of a protein simply on the basis of its amino acid sequence.

An alternative to using a more specific enzyme is to instead perform the enzymatic digestion for a shorter period of time (9). This ensures that many of the potential cleavages sites instead remain un-cleaved, extending the average length of the peptide. This can be done in principle by using very low enzyme concentrations to dramatically reduce the rate of proteolysis, but the more common strategy is to quench the reaction after a short reaction time (9). Both of these strategies tend to face problems in reproducibility (1), either because of the difficulty of accurately measuring small increments of enzyme, or adequately quenching the enzymatic digestion within a very specific timeframe. Nevertheless, the results from limited digestion are quite promising in that, when successfully employed, they generate a series of overlapping peptides which tends to extend protein coverage (9, 10).

Employing this strategy with standard proteins like trypsin doesn't entirely resolve the aforementioned sequence dependence, however. While trypsin cleavage sites

are indeed common throughout the proteome, significant variability in the distribution of basic residues can still vary wildly, particularly in very hydrophobic proteins (11). As a result, although these approaches offer greater sequence overlap than a highly specific digestion, they are still ultimately an imperfect solution. The only way to truly generate consistent cleavages across a molecule in a sequence-independent manner is to precisely employ an enzyme which cleavages peptides largely independently of their sequence (12). Such a technique is employed here to systematically digest proteins regardless of their sequence.

### ***Aspergillopepsin***

The enzyme integrated into this work possesses a variety of unique properties that make it an ideal candidate for primary structure analysis. The enzyme, known as Aspergillopepsin I, is an aspartic protease isolated from *Aspergillus phoenicis*, a mold commonly used to ferment a variety of Japanese liquors (13). It is composed of 325 amino acids and weights ~34,302 Da and is primarily functional in an acidic environment. It exhibits the highest activity when held at a pH approximately from 2.5-3, but it remains stable and maintaining good activity ranging up until ~pH 6 (13). This is beneficial, as slightly acidic environments eliminate many of the artifacts, for example asparagine and glutamine deamidation, that are commonly induced in digestions performed under more neutral or basic conditions (14). Additionally, the enzyme remains stable in 8M urea under these conditions for up to 1 hour (13, 15). This kind of stability allows proteins to be digested in highly denaturing conditions, unfolding the protein and

ensuring consistent cleavages along the backbone and reducing the extent to which dominant cleavages are the result of a protein's tertiary structure (16, 17).



Figure 3.1: Crystal Structure of Aspergillopepsin. Taken from (18)

Aspergillopepsin's cleavage specificity is largely nonspecific, although it does appear to show a slight preference for certain polar or charged residues like R, K or N in the P1 position as well as aliphatic amino acids in the P1' position (13, 19). This provides a significant advantage as it allows the methodology to remain extensible for a wide variety of systems rather than relying on a protein to possess a particular amino acid distribution. Despite this capability to cleave nearly all peptides bonds in the molecule, digestions produced with Aspergillopepsin tend to be quite reproducible in their cleavage sites (13, 19). These features are incredibly valuable because they mean that the digestion

should be applicable to proteins of any size (given that they are being digested in a largely sequence-independent manner), yet a given set of digestion conditions should still produce the same results if replicated in a similar manner.

The biggest difficulty of working with an enzyme like Aspergillopepsin is managing to reproducibly end the reaction before the enzyme digests the protein of interest into incredibly small pieces. This becomes a significant issue for two main reasons. Preliminarily, as noted earlier, Aspergillopepsin is an incredibly stable enzyme, remaining active in 8M urea and acidic conditions (13). More common digestions like trypsin are often quenched by adding a small portion of acid, reducing the pH below its active range. However, naturally, this strategy is ineffective for aspergillopepsin. Further, aspergillopepsin has a very high enzymatic activity, capable of chewing proteins into incredibly small pieces in mere minutes, unlike other enzymes which often react for hours (20). Efficiently quenching the reaction on that kind of timescale is impractical. While significantly dropping the enzyme:protein ratio can be massively reduced to something like 1:1000 and maintain large pieces, in principle. However, the amount of enzyme used for the experiment becomes impractically small to handle effectively, increasingly the variability in enzyme added to the experiment and hampering reproducibility.

Fortunately, both of these limitations can be largely mitigated by surface conjugating the enzyme to a solid scaffold within an enzyme reactor (15). In doing so, protein solutions can be passed through the column for the appropriate length of time to both reproducibly generate the desirable number of cleavages and effectively “quenching” the reaction as the proteins elute out of reactor (15). This has two further

advantages. The enzymes are generally sterically limited when linked to the surface of the beads, which dramatically reduces the amount of enzyme autolysis peptides observed within a final digestion (21). It also tends to rapidly accelerate the digestion process because the local surface concentration of the enzyme is significantly higher than its concentration would be when working in a solution (21–23).

A proof of principle design of this kind was previously developed in our lab by Weihan Wang and Lichao Zhang (15). Enzyme particles were conjugated to highly porous 20  $\mu\text{m}$  particles, which were pressure loaded into a reaction vessel. Passing a protein solution through this reactor demonstrated nonspecific, reproducible digestions of analyte proteins, and importantly, the degree of digestion and therefore peptide size could be effectively modulated by manipulating the flow rate through the reactor, as depicted in Figure 3.2. Faster flow rates resulted in shorter residence times within the reactor and therefore larger peptides on average (15). The combination of multiple LC-MS analyses of these digestions was shown to be capable of reconstructing the majority of the sequence of a model antibody. Both the instrument methodology and digestion size distribution on this system leave significant room for improvement and as such, improvements and extensions of this methodology are pursued here.

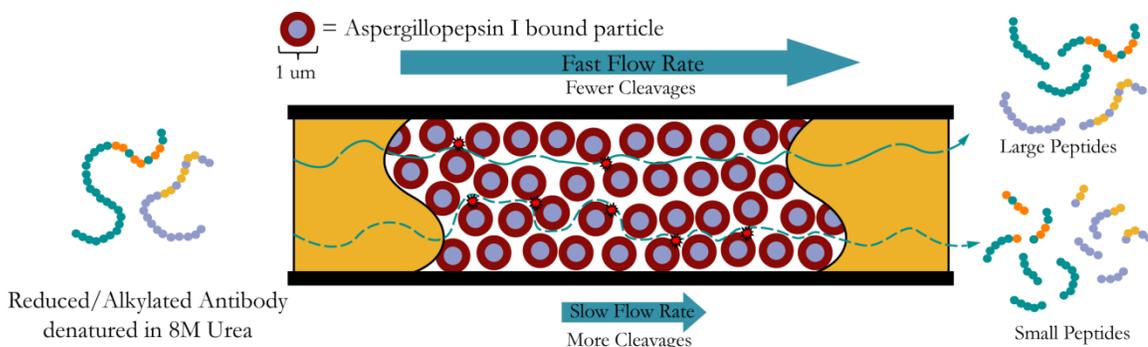


Figure 3.2: Schematic representation of controlled digestion by Aspergillopepsin I Enzyme Reactor

### ***Immobilized Reactor Digestion Complexity***

One of the more technical limitations of the initial enzyme reactor construction is the ability to generate a peptide size distribution that selectively generates the desired size profile. While the reactor can indeed generate the desired size distribution in principle, the range of peptide sizes in the analysis is still quite wide, ranging from 1-2 kDa “tryptic-like” peptides all the way to peptides consisting of >70% of the intact molecular weight of the protein (15). Such a broad distribution of analyze sizes complicates the analysis of the resulting peptides, particularly the larger peptides, due to segmenting protein material into so many different pathways. This reduces the intensity of any given peptide, both reducing the MS/MS quality on lower level precursors as well as the sensitivity of the technique more broadly.

Ultimately, this kind of size variation results from some analyte protein copies residing in digestion-active portions of the reactor for longer than others (24). As such, they tend to undergo multiple cleavages while other copies of the same protein may undergo one or no cleavages in the same length digestion. This degree of variation results in a much lower intensity of the peptides within the size range of interest, as it both destroys some of the peptides formed during the digestion (due to receiving additional cleavages prior to leaving the reactor) and reduces the number generated in the first place (due to larger peptides receiving too few cleavages). Steps must therefore be taken accordingly in order to try to minimize this degree of digestion variation in order to both produce a simpler mixture of peptides and to increase peptide abundance within the size range of interest.

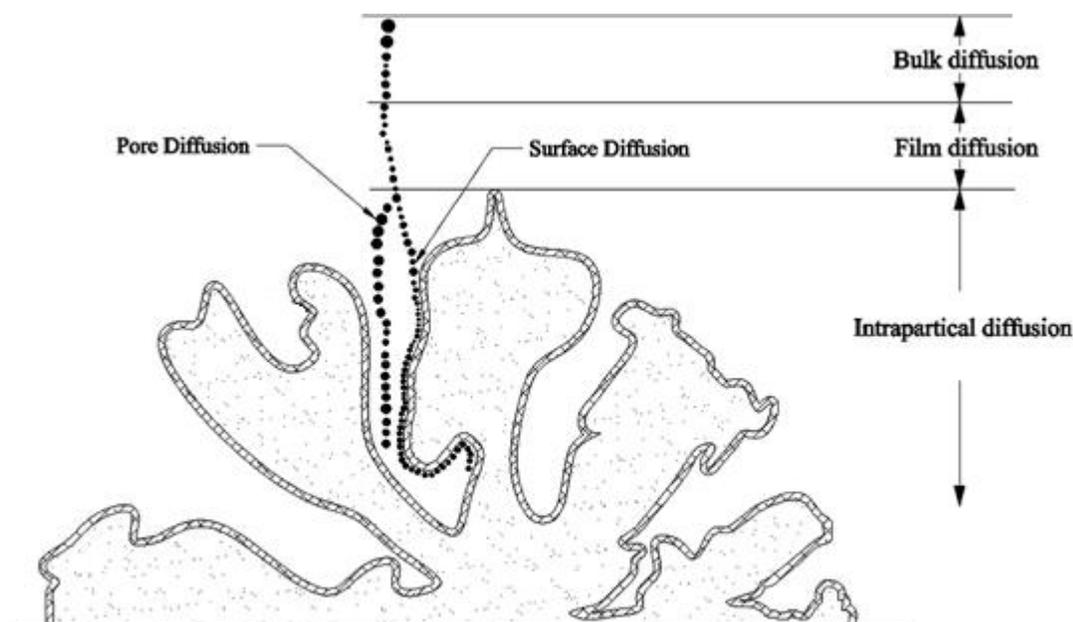
Ensuring that proteins flowing through the reactor are exposed to nearly the same proteolytic environment becomes quite complicated when considering the number of potential steps that a given protein must go through in order to be digested and subsequently elute from the column. Unlike an in-solution digestion where the catalytically active enzyme is evenly distributed through solution, enzyme reactors instead act as catalytic surfaces (25, 26), wherein only substrate present at the scaffold surface will be digested while any protein in the bulk solution will remain unreacted. As such, in order to be digested, the protein must ultimately go through the steps of 1) diffusing to the surface of the conjugated particle, 2) diffuse into pores of the particle (assuming a porous medium), 3) adsorb to the surface of the particle, 4) undergo the proteolytic reaction, 5) desorb from the surface of the particle, 6) diffuse out of the pore, and then finally 7) diffuse back into the bulk solution (26). Modifications to any of these steps can ultimately generate differences in the catalytic profile of a reactor, which in turn modifies the end result.

### ***Diffusion Limitations***

The diffusive effects in the aforementioned description are indeed non-trivial when evaluating the behavior of a catalytic reactor, particularly when using a porous substrate. One must consider that for the reactor in question, the enzyme is localized to the catalytic surface, and when using a porous particle, the overwhelming majority of catalytic surface area is located within the pore rather than at the particle's outer surface. As such, in order for a molecule to reach the porous surface they must overcome two

types of diffusion, namely, external (film) diffusion and internal (pore) diffusion (26).

These diffusive processes are pictured in Figure 3.3.



**Figure 3.3: Depiction of intra- and inter- particle diffusion schemes.** Adapted from (27).

Analytes at various stages in these diffusional steps will exhibit different velocities through the reactor. Under normal circumstances, the external fluid layer surrounding a particle in a packed bed is typically immobile (owing to the viscous shear forces on the sides of the particle) such that analyte movement within this film is dominated by diffusion (26, 28). Further, under typical flow conditions through a packed bed, the vast majority of flow occurs in the bulks solution in the inter-particle space rather than the intra-particle pore matrix (29). As a result, once a molecule enters one of these regions, it no longer progressed through the reactor at the expected flow velocity.

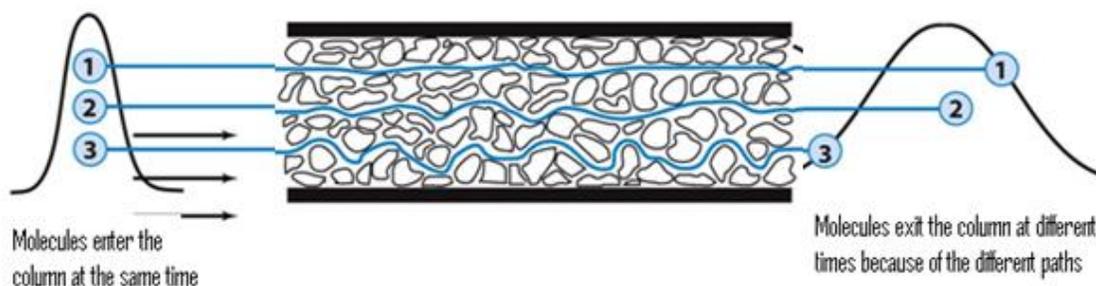
Thus, the longer a molecule lingers within either of these spaces, the more its actual residence time within the reactor varies from the average residence time.

This becomes a particular problem when achieving highly consistent particle residence times is necessary to achieve a uniform distribution of catalytic reactions. Fortunately, this issue can be mitigated by a variety of changes. For instance, both diffusion components can be reduced by dropping the packed particle diameter (30). This is because the decreased particle size both inherently limits pore depth as well as decreases the depth of the film layer by reducing the shear force on the particle (31, 32). However, transitioning to solid particle may be used to eliminate the pore diffusion component altogether (25, 33, 34). This has the trade-off of significantly reducing the reactive surface area, decreasing the reaction rate in a catalytically limited regime.

### ***Multipath Diffusion***

An additional dispersion component also contributes to alterations in residence time within a packed bed reactor. In addition to variations within the stagnant solvent in or near the surface of the packed particles, particles will tend to adopt different trajectories through the packed bed, even when driven at the same flow rate (30, 35, 36), as depicted in Figure 3.4. This is the result of a variety of factors. It can in part be associated with irregularities in either the packed bed or its pore network. As the pore network through each individual particle is largely variable, the path analytes travel through this pore network will be inherently variable as a result of this non-uniformity. Additionally, under experimental circumstances, sections of the particle bed may not pack perfectly ideally, leading to areas of the bed with lower resistance to flow and

causing channeling. Any analyte which passes through this region of the bed will flow significantly faster than its counterparts elsewhere. Even in an ideally packed bed of solid spheres, a level of stochastic variation in analyte path at any flow channel junction will inevitably lead some analytes to proceed down paths that are inherently longer than others (37). Thus, contributions from multipath diffusion are difficult to eliminate entirely.



**Figure 3.4: Multipath Diffusion (Eddy Diffusion).** Random variation in analyte path through the packed bed leads in actual residence times that vary from the average expected residence time based on solvent flow rate. Adapted from (38).

Fortunately, a straightforward modification may be made to reduce the multipath diffusion through the reactor. As particle size decreases, the variation in analyte path length correspondingly decreases. An intuitive way to imagine this is that the time difference between taking a long path and a short path around a smaller particle is going to decrease. While one might imagine that this would get evened out by a larger number of small particles, the variation decreases as a result of more averaging between taking the short path vs the long path over the larger number of iterations.

To address some of these limitations, here we illustrate how making necessary scaffold modifications to increase large peptide yield enables the generation of large antibody pieces which encompass significant swathes of the molecule, in particular the

entire variable region of the antibody. Further, we show how robust characterization of these large pieces enables near-complete characterization of the CDRs of the molecule at a level of depth that neither intact nor complete digestion analyses are capable.

### **3.3 Materials and Instrumentation**

#### **Agilent Technologies (Palo Alto, CA)**

1100 Series high performance liquid chromatograph

1100 Series vacuum degasser

#### **Applied Biosystems (Foster City, CA)**

20  $\mu\text{m}$  diameter POROS AL Beads

#### **Eppendorf (Hauppauge, NY)**

5414R Benchtop centrifuge

#### **GTS-Welco (Allentown, PA)**

10 ppm SF<sub>6</sub> in nitrogen gas

#### **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

Sizes: (360  $\mu\text{m}$  o.d. x 75 & 150  $\mu\text{m}$  i.d.)

#### **Sigma Aldrich (St. Louis, MO)**

N-(2-Aminoethyl)maleimide trifluoroacetate salt,  $\geq 95\%$  (HPLC),  $\geq 98\%$  (T)

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

Glacial acetic acid,  $\geq 99.99\%$  trace metal basis

Fluoranthene,  $>99\%$  purity

2-propanol, LC-MS grade

Tris(2-carboxyethyl)phosphine, ( $>98.0\%$ )

**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  $\mu\text{m}$

Formic Acid, LC-MS Grade

Orbitrap Elute<sup>TM</sup> Mass Spectrometer (Modified)

Pierce<sup>®</sup> Water, LC-MS Grade

Urea ( $>99.0\%$ )

### **3.4. Methods**

#### ***Aspergillopepsin I Enzyme Reactor Fabrication***

Preliminarily, the enzyme reactor scaffold must be surface conjugated with Aspergillopepsin I prior to any kind of reactor digestion. This procedure was performed for both 20  $\mu\text{m}$  Poros beads and 1  $\mu\text{m}$  solid beads in order to compare the performance of the two beads. Aspergillopepsin is surface conjugated to the reactor particles via reductive amination, the mechanism for which is shown in Figure 3.5. For 20  $\mu\text{m}$  bead preparation,  $\sim 7$  mg of dried POROS 20 AL aldehyde-coated particles are weighed out for

conjugation. A solution of 0.28g/mL (saturated) sodium sulfate is prepared, and dried Aspergillopepsin I enzyme is dissolved in the solution to generate a 10 mg/mL enzyme solution. The beads are reconstituted in the sodium sulfate/Aspergillopepsin I solution. When in the presence of this solution, the enzyme will have the tendency to nucleate on the surface of the beads and reversibly form imines with the aldehyde groups on the bead surface. Subsequently, an 80 mg/mL solution of sodium cyanoborohydride is prepared in water. In order to reduce the imine double bond (39), 1  $\mu$ L of the solution is added to the reaction mixture per 200  $\mu$ L of solution, to achieve a final concentration of 0.4 mg/mL. The reaction is allowed to proceed with shaking for approximately 20 hours at room temperature.

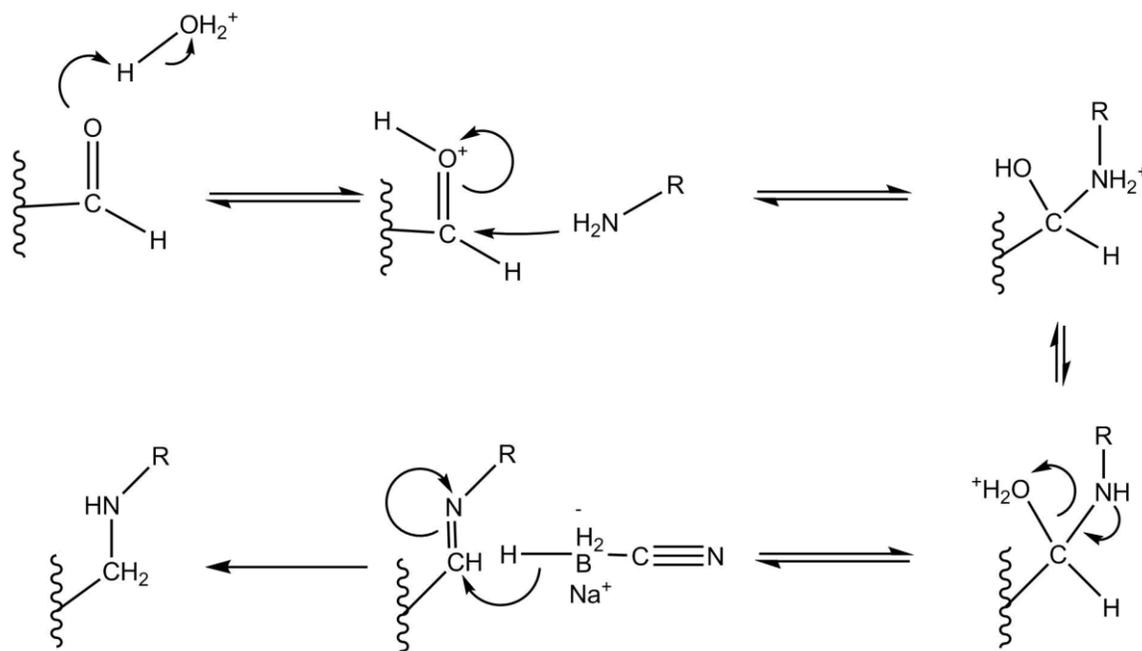


Figure 3.5: Mechanism for Surface Conjugation by Reductive Amination. The conjugated enzyme is represented as the R- group on the primary amine.

Following enzyme conjugation, the enzyme reaction mixture is removed by use of a spin filter, leaving the dried, enzyme conjugated particles. The particles are then reconstituted in 0.2M, pH 6.5 Tris buffer and 1  $\mu$ L of sodium cyanoborohydride is added

to the bead slurry. This reaction is allowed to proceed for 3 hours at room temperature with shaking in order to block the remaining unreacted aldehyde sites on the bead surface; the amine moiety on Tris similarly reacts with the aldehyde functional group to form an imine, resulting in free hydroxyl groups on the bead surface rather than reactive aldehydes. Following aldehyde blocking, the beads are once again dried down on a spin filter and stored at 4° centigrade until used for sample digestion.

Conjugation of 1  $\mu\text{m}$  beads proceed similarly, albeit with minor modifications. Unlike the POROS beads, the 1  $\mu\text{m}$  particles used for the reactor are 1  $\mu\text{m}$  Invitrogen aldehyde/sulfate latex beads. These beads are received in a solution rather than a solid, so ~10 mg of beads is measured out by removing 250  $\mu\text{L}$  of the stock solution. The stock solution is removed, and the beads are washed twice with water to remove as much as possible. Sodium sulfate/Aspergillopepsin and sodium cyanoborohydride are added in the same manner as 20  $\mu\text{m}$  beads and reacted for 20 hours followed by the 3-hour tris blocking procedure. The beads are then washed again, pelleted, and the supernatant is pulled off with a pipette. The beads are then stored mostly dry at 4° centigrade, although some residual water is likely unavoidable. As a result, while 20  $\mu\text{m}$  beads appear to be stable while dry for as long as 2 years, 1  $\mu\text{m}$  beads generally only maintain appropriate digestion behavior for ~3 months when being used consistently.

### ***Apomyoglobin Standard Digestions***

In order to optimize the performance of the enzyme reactor, reactor scaffold optimization was tested on Apomyoglobin as a test protein. An aliquot of 17.0  $\mu\text{g}$  (1 nmol) apomyoglobin was reconstituted in 85.8  $\mu\text{L}$  of digestion buffer – 0.5% acetic acid,

50 mM ammonium acetate, 8M urea (digestion buffer). A series of digestions was performed on each bed with different digestion times in order to find comparable protein consumption between the two reactor constructs.

For 20  $\mu\text{m}$  particle digestions, the reactor was prepared by packing 2.2 cm of aspergillopepsin I conjugated particles into a piece 360 OD (outer diameter) x 150 ID (inner diameter) fused silica. The reactor was allowed to compress for ~30 minutes by passing LC-MS grade water through the reactor at 500 psi. The water was then exchanged for digestion buffer and the apomyoglobin sample was passed through the reactor for digestion. Flow rates were specified such that digestion times would be generated according to the equation:

$$\text{Digestion Time} = \frac{\text{Trapped Volume}}{\text{Flow Rate}} = \frac{3.60 * \text{Bed Length (cm)}}{\text{Flow Rate (\mu L/min)}} \quad (3.1)$$

where 3.60 represents an adjustment for the column inner diameter and packed bed porosity such that:

$$3.60 = 60 * 1000 * \pi * \left(\frac{d_i^2}{2}\right) * \rho = 6000 * \pi * \left(\frac{0.015^2}{2}\right) * 0.3 \quad (3.2)$$

where  $d_i$  represents the column i.d. in centimeters,  $\rho$  represents the porosity of the packed bed, and the multiplication by 60 and 1000 convert the units from minutes to seconds and  $\text{cm}^3$  to  $\mu\text{L}$ , respectively. Digestion times of ~0.53 s and ~0.70 s were collected.

1  $\mu\text{m}$  scaffold digestions were performed similarly, albeit on a smaller bed to compensate for the increased back pressure. 1.5 mm of enzyme conjugated 1  $\mu\text{m}$  particles were packed into a 360 x 150 piece of fused silica. The reactor was similarly compressed in water and equilibrated to digestion buffer prior to digestion. Digestion times of ~0.20

s, ~0.47 s, and ~0.71 s were collected, with digestion times being estimated by the slightly modified equation:

$$\text{Digestion Time} = \frac{\text{Trapped Volume}}{\text{Flow Rate}} = \frac{3.03 * \text{Bed Length (cm)}}{\text{Flow Rate (}\mu\text{L/min)}} \quad (3.3)$$

where the adjustment parameter is modified as a result of the 1  $\mu\text{m}$  solid bed having a scaffold porosity of 28.5% rather than 34%.

All digests were individually analyzed by LC-MS on an Orbitrap Classic Mass Spectrometer. 1 pmol of each digest was pressure loaded on a reverse phase HPLC column containing a 10 cm, 360 x 75 analytical column butt-connected to a 10 cm. 360x100 preparatory column, both packed with 3  $\mu\text{m}$  Poroshell C-18 packing material. The column was desalted by flowing 0.3% formic acid at 50 bar for ~40 minutes to eliminate urea and salts from the column prior to MS analysis. Digests were then gradient eluted in-line with the instrument using 0.3% formic acid as solvent A and 72% ACN, 18% IPA, 10% water, 0.3% formic acid as solvent B and a gradient of 25%-55%-100% solvent B in 5-40-80 min. Eluting peptides were fragmented with ETD to facilitate identification.

Following digestion analysis, the ~0.70 s 20  $\mu\text{m}$  digestion and the ~0.53 s 1  $\mu\text{m}$  digestions were selected for comparison on the basis that they both contained significant fragment ion intensity and similar levels of intact protein normalize precursor consumption. Base-peak level peptides of varying sizes were selected for quantification, and the mass area for each peptide was summed over their respective peaks. The ratio of peak abundances was taken to compare the difference in peptide size profile between the two digestions.

### ***Antibody Sample Preparation***

Stock samples of originator adalimumab were received from Pfizer at -80° centigrade at a concentration of 50 µg/µL in 6.16 mg/mL sodium chloride, 0.76 mg/mL monobasic sodium phosphate, 2.30 mg/mL dibasic sodium phosphate, 0.3 mg/mL sodium citrate dihydrate, 1.3 mg/mL citric acid monohydrate, 12 mg/mL mannitol, 1 mg/mL polysorbate 80 and adjusted to a pH of 5.2 using sodium hydroxide. These were consistently stored at -80° centigrade until prepared for analysis.

Prior to antibody analysis, the stock was thawed and 100 µg (2 µL) was removed and diluted in 50 mM ammonium bicarbonate. 100 units of IdeS were reconstituted in 20 µL ammonium bicarbonate and combines with the adalimumab, giving a final antibody concentration of 1 µg/µL. This sample was reacted for 30 minutes at 37° centigrade to cleave the middle of the antibody heavy chain below the inter-heavy chain disulfide bonds, creating the ~100 kDa F(ab')<sub>2</sub> (containing the entire antibody variable region) and the ~25 kDa Fc/2 (containing the antibody glycosylation site).

For antibody variable region parking experiments, the Fc/2 was removed from the antibody mixture in order to simplify the mixture and reduce the probability of co-isolation when fragmenting the relevant species. While this reduces the total amount of protein information in the run, it better facilitates maximum characterization in the sub-region of interest. To do this, the IdeS digested adalimumab was loaded on a Protein A spin column (NAb Protein A Plus Spin Columns, Thermo Scientific) and shaken at room temperature for 12 min. Protein A is a bacterial protein with a high and specific affinity for antibodies (40). For human IgG<sub>κ</sub> antibodies like adalimumab, the binding site for

protein A is present on the Fc of the molecule, thus removing the Fc/2 from the reaction mixture and allowing the F(ab')<sub>2</sub> to be eluted out of the column (40).

Following protein A purification, 15 µg of the sample was sub-aliquoted and reconstituted in 10 µL of 10 mM tris(2-carboxyethyl)phosphine in 8M urea and 0.5% acetic acid. The sample was incubated for 12 minutes at 50° centigrade to fully reduce any disulfide bonds present in the molecule. This not only better unfolds the molecule by reducing inter-chain disulfide bonds, but also segments the 100 kDa F(ab')<sub>2</sub> into two ~25 kDa heavy chains and two ~25 kDa light chains. The two copies of each of these chains are sequence identical to one another, effectively reducing the sample complexity down to two ~25 kDa proteins. After reduction, the solution was neutralized to pH 6.5 with X µL of 0.2 M ammonium hydroxide and diluted in 10 µL of 20 mM N-(2-Aminoethyl)maleimide (NAEM) in 8M urea and 0.5M ammonium acetate and reacted at room temperature for 10 minutes in order to alkylate the free sulfhydryl groups of the reduced cysteines. This procedure not only prevents the reformation of the reduced disulfide bonds, but also adds an extra basic moiety as cysteine residues in order to enhance the charge density of the protein and improve ETD fragmentation (15). The reaction was allowed to proceed for 10 minutes before being acidified with 3 µL 25% formic acid. After alkylation, the sample was diluted to a final concentration of 0.2 µg/µL in 0.5% acetic acid, and 8M urea, resulting in a final buffer composition of 0.5% acetic acid, 50 mM ammonium acetate, and 8M urea at pH 4.

Following antibody sample preparation, adalimumab sample was digested within the Aspergillopepsin I immobilized enzyme reactor. A 1µm particle enzyme reactor was packed to a bed length of 2.0 mm and packed down for 30 minutes. The 0.2 µg/µL

adalimumab solution was passed through the reactor at a flow rate of 0.66  $\mu\text{L}/\text{min}$ , equating to a digestion time of  $\sim 0.91$  seconds within the reactor. 10  $\mu\text{L}$  were collected and the sample was diluted to 1 pmol/ $\mu\text{L}$  in digestion buffer prior to analysis.

### *Antibody Digestion Analysis*

The resulting antibody digestion analyzed on an in-house modified Orbitrap Elite. The digestion products were separated chromatographically on a PLRP-s reverse phase analytical column (75  $\mu\text{m}$  inner diameter, 10 cm bed, 3  $\mu\text{m}$  diameter particles). 1 pmol of the digestion products were loaded onto the column and rinsed with solvent A to remove excess salts in the sample. Peptides were then gradient eluted in-line with the instrument using a gradient of 25%-50%-100% solvent B in 5-80-85 minutes. Flow through the column was approximately 100 nL/min and eluting peptides were ionized from a column-integrated nano-electrospray ionization tip at 2200 V.

Masses of eluting peptides were determined by a 60,000 resolution MS1 profile scan, and targeted pipETD/IIPT fragmentation was directed at the +14 charge state of the light chain VL peptide D<sub>1</sub>-E<sub>105</sub> ( $m/z = 837$ ) and the +20 charge state of heavy chain VH peptide E<sub>1</sub>-K<sub>151</sub> ( $m/z = 824$ ). Reaction times of 85 and 75 ms were used for pipETD fragmentation for the light chain and heavy chain peptides, respectively, and the resulting fragment ions were charge-reduce with 15 ms IIPT in both cases. Parallel ion parking during the ETD phase of the reaction used the same base waveform as described in Chapter 2. All frequencies along the mass range were parked except for frequencies  $\pm 10$   $m/z$  around fluoranthene ( $m/z = 202$ ), and frequencies 70  $m/z$  below and 10  $m/z$  above the precursor's theoretical frequency. The waveform's amplitude scalar was set to 0.1

(arb.). Ion-ion reactions were performed such that the Mathieu Q for the reagent ion was held at 0.75 for pipETD reactions and 0.55 for IIPT reactions. 5 multiple C-trap fills were performed prior to each 120,000 resolution MS2 transients, and 2 transient microscans were averaged to generate each analytical scan.

Analytical MS/MS scans performed over each of the respective chromatographic peaks were averaged, resulting in 6 averaged scans for light chain D<sub>1</sub>-E<sub>105</sub> and 10 averaged scans for heavy chain E<sub>1</sub>-K<sub>151</sub>. The resulting spectra were then manually annotated to assign the total sequence coverage for each peptide.

### **3.5 Results and Discussion**

#### ***Enzyme Reactor Scaffold Modifications***

Perhaps the biggest concern at the outset of specific subunit analysis by nonspecific digestion is the ability to generate any given subunit at appreciable enough abundances to readily analyze. Unlike a more typical complete digestion, one difficulty with using a limited digestion is that peptide products can continue to be digested by the enzyme, destroying some of the peptides within the ideal size range of the particular application. While modifying the flow rate through the reactor can effectively tailor the average peptide size distribution, variability in actual peptide residence time within the reactor can result in a fairly broad distribution of peptide sizes. This is further compounded by the immense number of potential cleavage sites when using a nonspecific protease, dividing up the ion current between an even larger number of pathways and increasing the likelihood that a second cleavage will occur. This risks significantly diluting individual peptide intensity into a variety of fragmentation

pathways. This division of ion current detrimentally impacts the sequence information extracted from fragmentation spectra as a result of MS/MS spectral quality being related, in part, to peptide abundance (both through achieving sufficient ion flux to achieve ion targets as well as maximizing the proportion of isolated ion current corresponding to the desired precursor). As such, maximizing the abundance of large peptides of interest becomes imperative to achieving maximum sequence information from the region in question.

In order to improve the efficiency with which the digestion generates large pieces, scaffold modifications were pursued in order to optimize the generation of larger fragments. To accomplish this, the scaffold was changed from 20  $\mu\text{m}$  highly porous particles to 1  $\mu\text{m}$  solid spheres. These changes were aimed at mitigating the components which drive the digestion complexity discussed in Section 3.2. For instance, the change to particle size positively impacts both multipath diffusion as well as resistance to mass transfer. As the size of particles within the reactor decrease, the variation in path length between them tends to go down, narrowing the digestion profile (30, 31). Ordinarily, decreased particle size would also positively impact the pore diffusion component as well, as a smaller particle naturally contains more shallow pores owing to its reduced diameter (32, 33, 41). However, in this instance, pore diffusion was removed entirely by transitioning instead to non-porous particles, leaving the film diffusion as the only relevant mass transfer component remaining in the reactor. As such, normalization of both the protein path-length through the bed as well as a significant reduction in the mass transfer component of the reactor is much more likely to narrow the resulting peptide

distribution, increasing the yield of the target peptide size at a given reactor digestion time.

The advantages of these modifications were first evaluated using standard digestions of apomyoglobin for differences in peptide abundances. As the modifications to the reactor scaffold resulted in variations in digestion time, a number of digestion time points were required to find equivalent reaction conditions for the two reactors; the 1  $\mu\text{m}$  particles appears to possess a higher digestion rate, achieving equivalent precursor protein consumption at equivalent digestion times. The best digestion times for comparison were found using an  $\sim 0.70$  s digestion on the 20  $\mu\text{m}$  scaffold and an  $\sim 0.53$  s digestion on the 1  $\mu\text{m}$  scaffold, as pictured in Figure 3.6, on the basis that these digestion times contained a significant abundance of large peptides as well as contained some remaining intact apomyoglobin such that the digestion time could be normalized on the basis of precursor protein consumption.

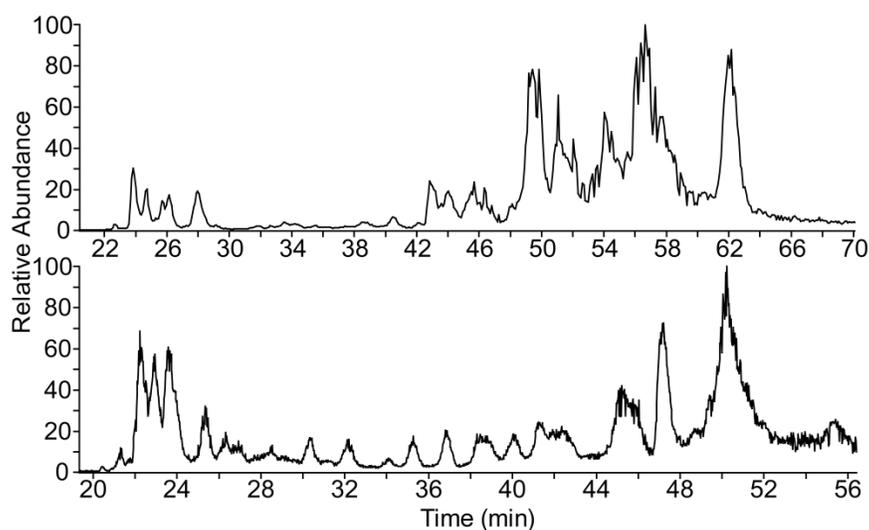
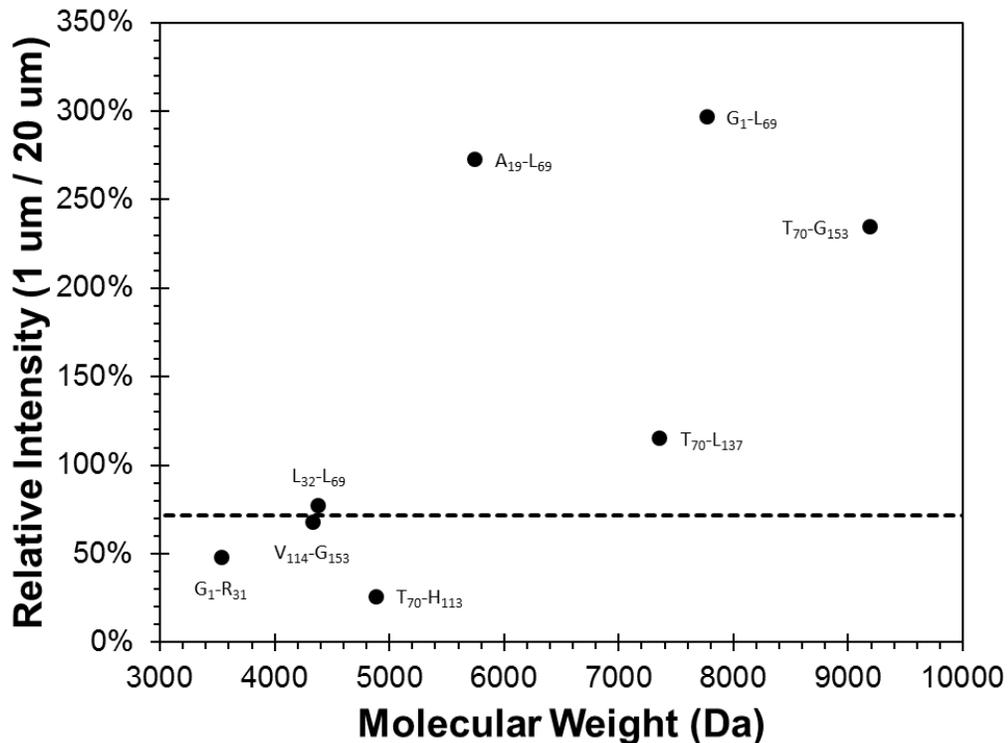


Figure 3.6: **Total ion chromatogram comparison of 1  $\mu\text{m}$  (top) vs 20  $\mu\text{m}$  (bottom) digestions.** Note that the smaller, earlier eluting species are significantly smaller in the 1  $\mu\text{m}$  digestion, and ion current is significantly more concentrated in a few, larger species.

Quantification of several different peptides of various sizes showed significant increase in large peptide abundances present in the 1  $\mu\text{m}$  digestion. Figure 3.7 plots the relative differences in abundances of these peptide with respect to peptide size, as well as depicting the differences in intact apomyoglobin abundances. Despite a relatively small difference in intact protein abundance, peptide abundances varied dramatically. The largest quantified peptides, ranging from ~6-10 kDa showed as much as a 3-fold increase in abundance, while the smaller (~3-5 kDa) peptides in fact showed a distinct decrease in abundance. This decreased trend in smaller peptides as well as comparable intact protein indicates that the digestion profile was indeed narrowed rather than simply a result of increased digestion on the 1 $\mu\text{m}$  scaffold. Therefore, the modified scaffold is significantly more efficient at generating a particular sub-region of a protein at an optimized digestion time, making it more ideal for studying particular subunits of a protein.

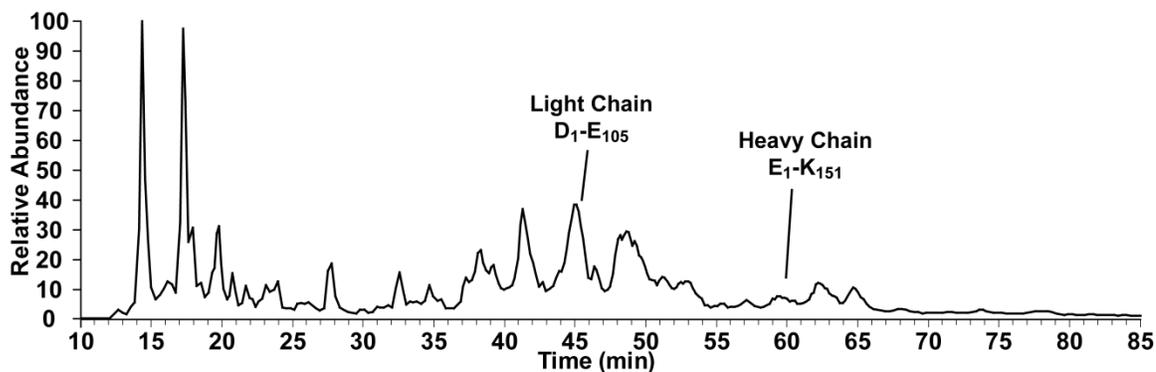


**Figure 3.7: Comparison of peptide abundances by size in 20  $\mu\text{m}$  vs 1  $\mu\text{m}$  scaffold apomyoglobin digestions.** Abundances were determined based off of peptide mass area over each chromatographic peak. The dashed line represents the abundance ratio of the remaining intact protein.

### *Antibody Variable Region Analysis*

The aforementioned modifications provide a useful experimental toolkit to analyze large swathes of a target protein without the need for sequence reconstruction. To illustrate the capabilities of the improved reactor, this platform was applied to the characterization of an antibody's intact variable region, namely that of adalimumab. Using this kind of strategy to generate peptides spanning the full variable region such that all of the CDRs could be co-localized on a given peptide allows their unambiguous assignment to a given protein product. In principle, this enables the discrimination of highly homologous antibodies that only vary on the basis of their CDRs.

The isolated, reduced, and alkylated F(ab')<sub>2</sub> region of adalimumab was subjected to ~0.91s digestion within the 1  $\mu\text{m}$  particle enzyme reactor in order to generate the appropriately sized subunit proteins. A ~0.91s digestion time was selected on the basis that it appeared to generate a significant degree of the largest peptides, while also minimizing the abundance of the intact molecule, thus maximizing the amount of useful peptide products present within the digestion. The resulting total ion chromatogram of this digestion is depicted in Figure 3.8. Analysis of the peptides generated by the digestion revealed a wide variety of peptide product ions which collectively spanned the full length of the molecule. Significantly, many peptides generated by the digestion were large enough to cover over half of their respective parent molecules.

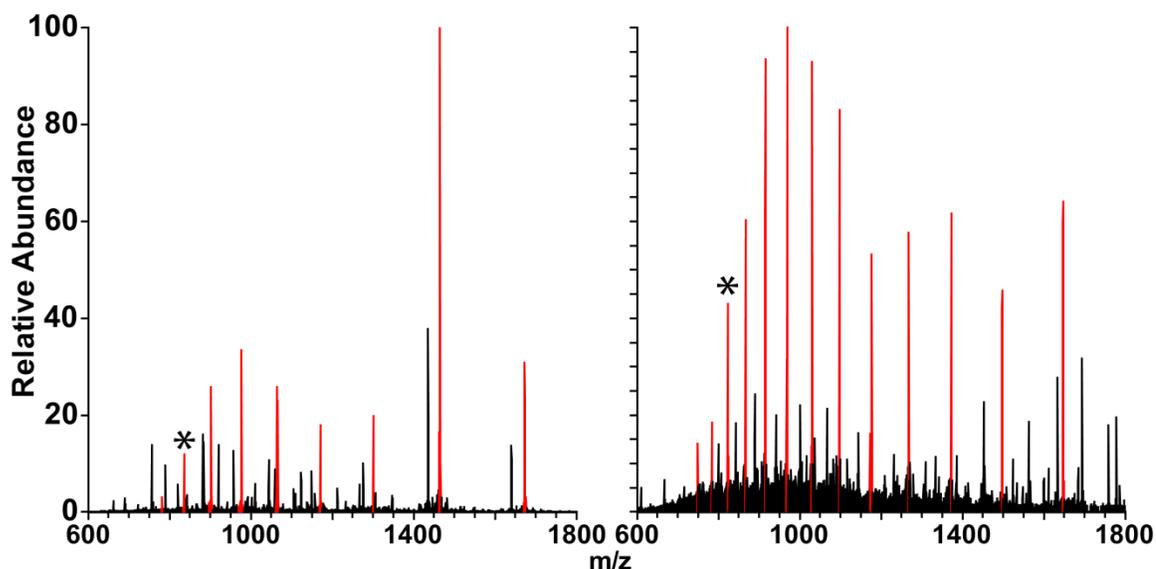


**Figure 3.8: Total ion chromatogram of ~0.91s adalimumab digestion.** The variable region peptide of each chain is denoted on the chromatogram.

Two particular promising peptides were identified within the analysis, namely light chain peptide D<sub>1</sub>-E<sub>105</sub> and heavy chain peptide E<sub>1</sub>-K<sub>151</sub>. Each of these peptides covered the N-terminal portion of their respective antibody chain and contained nearly the entire variable region, and most importantly contained all of the CDRs of their respective chain. Both of the peptides were generated from a single proteolytic cleavage on their respective chain and were generated with a high enough abundance to be clearly observable in the chromatographic profile, making them ideal candidates for in-depth sequencing in order to comprehensively characterize the variable regions of the antibody.

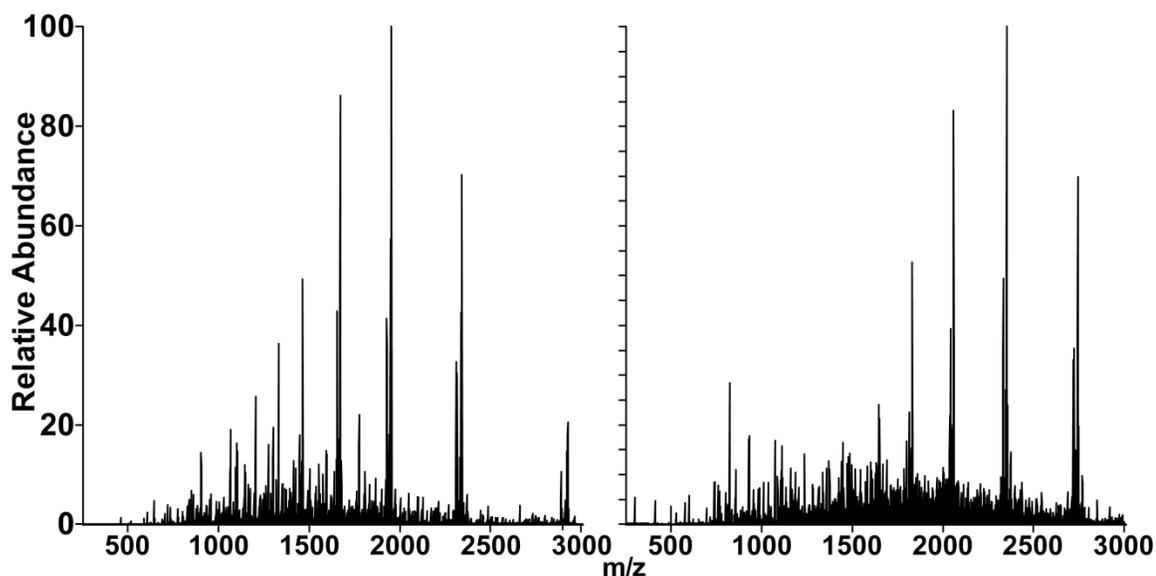
In order to most effectively extract sequence information from these peptides, the pipETD/IPT fragmentation methodology utilized on the NISTmAb in Chapter 2 was applied to the variable region fragments generated by the enzyme reactor digestion. Fortunately, both peptides generated charge states largely amenable to ETD fragmentation, in large part due to the extra charge added by the NAEM alkylation used during the sample preparation step. The MS<sup>1</sup> spectra illustrating the resulting charge-state distribution are pictured in Figure 3.10. The +14 charge state of the light chain peptide

( $m/z = 837$ ) and the +20 charge state of the heavy chain peptide ( $m/z = 824$ ) were subjected to pipETD/IIPT fragmentation and are denoted within the figure.



**Figure 3.9: MS1 spectra containing the adalimumab light chain (left) and heavy chain (right) variable region peptides.** Red peaks denote the peptides of interest and black peaks are co-eluting peptides. The isolated charge-state is denoted with an asterisk.

Fragmentation spectra of each of the peptides are depicted in Figure 3.11. Both fragmentation spectra generated significant fragment ion yields, suggestive of high sequence information. The spectra contain a multitude of charge-reduced precursors, again predominantly from the IIPT step rather than the pipETD step. However, the spectrum also shows a strong mass loss next to each of the charge reduced precursors corresponding to the loss of the NAEM side chain during ETD. Unfortunately, this sequence uninformative side chain loss results in a significant loss of the mass range (covering otherwise sequence informative ions) relative to an equal charge state precursor without the modification.



**Figure 3.10: Parked ETD/IPT MS2 fragmentation spectra of the adalimumab light chain (left) and heavy chain (right) variable region peptides.**

Sequence coverage maps derived from annotation of the fragmentation spectra are shown in Figure 3.12. Fragmentation of the light chain and heavy chain achieved 90% and 85% of all possible ETD cleavages (or ~85% and 79% of all total bond cleavages), respectively. Despite the incomplete sequence coverage, all three of the light chain's CDRs were still unambiguously mapped, achieving bond cleavages between all amino acids in the CDRs. The heavy chain peptide underperformed slightly, but still achieved all bond cleavages with the exception of 4 in CDR2 and 1 in CDR3. Proportionally, this equates to CDR coverages of 78% and 92% for CDR2 and CDR3, respectively. Fragment ions also show considerable overlap, illustrated by the fact that ~68% of light chain bond cleavages and 64% of heavy chain bond cleavages generated both observable c and z fragment ions. Indeed, were it not for the obfuscation caused by the NAEM side chain loss during ETD fragmentation, it is likely that this fragmentation pattern would have continued the full length of the molecule, enabling the observation of fragment ions corresponding to single amino acid cleavages. Regardless, the charge-state advantage

offered by NAEM in this context certainly makes its application highly advantageous in this context, strongly suggesting its use in similar experiments.

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

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

**Figure 3.11: Sequence coverage from ETD Parking/IPT spectra of adalimumab light chain (top) and heavy chain (bottom) variable region peptides. Residues within the CDR of each antibody are denoted in bold.**

Interestingly, the majority of missing sequence coverage in the molecules occurs predominantly on the N-terminus, particularly on the heavy chain peptide. This behavior mirrors the behavior of the NISTmab subunit analysis (Chapter 2); namely, the majority of sequence coverage loss occurs near the termini of the molecule despite the fact that the most terminal fragment ions are commonly the easiest to generate. Given the similarities between the antibodies, the explanation is likely the same: pipETD experiments will generate small, terminal fragment ions in lower yields due to preventing secondary reactions of larger fragment ions, and this is sufficient to prohibit the observation of these fragment ions in sufficiently charge-depleted regions such as those seen in antibody variable regions. Nonetheless, pipETD fragmentation generates impressive sequence coverage overall, and lacking coverage in a region of the molecule easily supplemented by other fragmentation techniques represents a significant advantage should they be used complementarily.

Cleavage site sequence comparison to other antibodies suggests significant promise for the extensibility of this approach to other antibodies. Sequence comparison with several antibody therapeutics (sequences from IMGT (42)) suggests that the cleavage sites used to generate the variable region peptides remain present within a wide variety of commercially used antibodies. The alignment of these sequences denoting the cleavage sites is pictured in Figure 3.13. It remains possible that other antibodies sequences would contain more favorable cleavage sites elsewhere in their variable region, increasing the propensity to destroy these peptides under those circumstances. However, the presence of highly conserved cleavage points to generate peptides containing an

antibody's entire variable region represents a promising opportunity to extend this methodology to a wider variety of antibodies.

	<b>Light Chain</b>	<b>Heavy Chain</b>
<b>Adalimumab</b>	YTFGQGTKVEIKRTVAAPSV	GTAALGCLVKDYFPEPVTVS
<b>Trastuzumab</b>	PTFGQGTKVEIKRTVAAPSV	GTAALGCLVKDYFPEPVTVS
<b>Rituximab</b>	PTFGGGTKLEIKRTVAAPSV	GTAALGCLVKDYFPEPVTVS
<b>Bevacizumab</b>	WTFGQGTKVEIKRTVAAPSV	GTAALGCLVKDYFPEPVTVS
<b>Ranibizumab</b>	WTFGQGTKVEIKRTVAAPSV	GTAALGCLVKDYFPEPVTVS
<b>Palivizumab</b>	FTFGGGTKLEIKRTVAAPSV	GTAALGCLVKDYFPEPVTVS

**Figure 3.12: Aspergillopepsin I cleavage site sequence homology comparison of several therapeutic antibodies.** Antibody sequences were derived from IMGT (42).

Taken together, this data strongly recommends the application of targeted subunit analysis using limited proteolysis. Time-controlled digestion proves to be an effective tool for generating arbitrarily large pieces which can be tuned to encompass a particular region of interest. When paired with potent fragmentation techniques, more detailed structural characterization may be achieved than if the molecule were analyzed intact or following a complete digestion.

## References

1. Tsiatsiani L, Heck AJR (2015) Proteomics beyond trypsin. *FEBS J* 282(14):2612–2626.
2. Rodriguez J, Gupta N, Smith RD, Pevzner PA (2008) Does trypsin cut before proline? *J Proteome Res* 7(1):300–305.
3. 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.
4. Srzentić K, et al. (2014) Advantages of Extended Bottom-Up Proteomics Using Sap9 for Analysis of Monoclonal Antibodies. *Anal Chem* 86(19):9945–9953.
5. Jekel PA, Weijer WJ, Beintema JJ (1983) Use of endoproteinase Lys-C from *Lysobacter enzymogenes* in protein sequence analysis. *Anal Biochem* 134(2):347–354.
6. Wu C, et al. (2012) A protease for “middle-down” proteomics. *Nat Methods* 9(8):822–824.
7. 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(11):1709–1717.
8. Anderson LC, et al. (2015) Protein derivatization and sequential ion/ion reactions to enhance sequence coverage produced by electron transfer dissociation mass spectrometry. *Int J Mass Spectrom* 377(1):617–624.
9. Deng J, Julian MH, Lazar IM (2018) Partial enzymatic reactions: A missed opportunity in proteomics research. *Rapid Commun Mass Spectrom* 32(23):2065–2073.
10. She YM, et al. (2010) Mass spectrometry following mild enzymatic digestion reveals phosphorylation of recombinant proteins in *Escherichia coli* through mechanisms involving direct nucleotide binding. *J Proteome Res* 9(6):3311–3318.
11. Fischer F, Poetsch A (2006) Protein cleavage strategies for an improved analysis of the membrane proteome. *Proteome Sci* 4. doi:10.1186/1477-5956-4-2.
12. Pang Y, Wang WH, 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.
13. Templin MF, Höltje J-V (2013) *Handbook of Proteolytic Enzymes* doi:10.1016/B978-0-12-382219-2.00317-3.
14. Wakankar AA, Borchardt RT (2006) Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization. *J Pharm Sci* 95(11):2321–2336.

15. Zhang L, et al. (2015) Analysis of Monoclonal Antibody Sequence and Post-translational Modifications by Time-controlled Proteolysis and Tandem Mass Spectrometry. *Mol Cell Proteomics* 15(4):1479–1488.
16. León I, Schwämmle V, ... OJ-M& C, 2013 undefined Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *ASBMB*. Available at: <https://www.mcponline.org/content/12/10/2992.short> [Accessed April 21, 2019].
17. Chen EI, Cociorva D, Norris JL, Yates JR (2007) Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J Proteome Res* 6(7):2529–2538.
18. Sang Woo Cho, Kim NJ, Choi MU, 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.
19. Ahn J, Cao MJ, Yu YQ, Engen JR (2013) Accessing the reproducibility and specificity of pepsin and other aspartic proteases. *Biochim Biophys Acta - Proteins Proteomics* 1834(6):1222–1229.
20. Zhang H-M, et al. *Enhanced Digestion Efficiency, Peptide Ionization Efficiency, and Sequence Resolution for Protein Hydrogen/Deuterium Exchange Monitored by FT-ICR Mass Spectrometry* Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2784605/pdf/nihms86585.pdf> [Accessed April 21, 2019].
21. Gao J, Xu J, Locascio LE, Lee CS (2001) Integrated microfluidic system enabling protein digestion, peptide separation, and protein identification. *Anal Chem* 73(11):2648–2655.
22. Datta S, Christena LR, Rajaram YRS (2013) Enzyme immobilization: an overview on techniques and support materials. *3 Biotech* 3(1):1–9.
23. Tan YJ, et al. (2012) Limited proteolysis via millisecond digestions in protease-modified membranes. *Anal Chem* 84(19):8357–8363.
24. P.V D (1953) Continuous flow systems: Distribution of residence times. *Chem Eng Sci* 2(1):1–13.
25. Liese A, Hilterhaus L (2013) Evaluation of immobilized enzymes for industrial applications. *Chem Soc Rev* 42(15):6236–6249.
26. Scott Fogler H (2002) Elements of chemical reaction engineering. *Chem Eng Sci* 42(10):2493.
27. Yan D, Gang DD, Zhang N, Lin L (2012) Adsorptive Selenite Removal Using Iron-Coated GAC: Modeling Selenite Breakthrough with the Pore Surface Diffusion Model. *J Environ Eng* 139(2):213–219.

28. Ardrey RE (2003) *Liquid Chromatography – Mass Spectrometry: An Introduction* doi:10.1002/0470867299.
29. Tallarek U, Vergeldt FJ, As H Van (2002) Stagnant Mobile Phase Mass Transfer in Chromatographic Media: Intraparticle Diffusion and Exchange Kinetics. *J Phys Chem B* 103(36):7654–7664.
30. Gritti F, Farkas T, Heng J, Guiochon G (2011) On the relationship between band broadening and the particle-size distribution of the packing material in liquid chromatography: Theory and practice. *J Chromatogr A* 1218(45):8209–8221.
31. Nondek L, Brinkman UAT, Frei RW (1983) Band Broadening in Solid Phase Reactors Packed with Catalyst for Reactions in Continuous-Flow Systems. *Anal Chem* 55(9):1466–1470.
32. Van Den Berg JHM, Deelder RS, Egberrink HGM (1980) Dispersion phenomena in reactors for flow analysis. *Anal Chim Acta* 114(C):91–104.
33. Bartholomew C, Farrauto R (2011) *Fundamentals of industrial catalytic processes* Available at: [https://books.google.com/books?hl=en&lr=&id=ornmC\\_v1AZ8C&oi=fnd&pg=PT14&dq=Bartholomew,+C.+H.%3B+Farrauto,+R.+J.+Fundamentals+of+Industrial+Catalytic+Processes%3B+John+Wiley+%26+Sons,+2011.&ots=osmfZ70C0-&sig=CEKhIkWUZfKHKPs1uXnX97fW5Ro](https://books.google.com/books?hl=en&lr=&id=ornmC_v1AZ8C&oi=fnd&pg=PT14&dq=Bartholomew,+C.+H.%3B+Farrauto,+R.+J.+Fundamentals+of+Industrial+Catalytic+Processes%3B+John+Wiley+%26+Sons,+2011.&ots=osmfZ70C0-&sig=CEKhIkWUZfKHKPs1uXnX97fW5Ro) [Accessed April 22, 2019].
34. Wu N, Liu Y, Lee ML (2006) Sub-2  $\mu\text{m}$  porous and nonporous particles for fast separation in reversed-phase high performance liquid chromatography. *J Chromatogr A* 1131(1–2):142–150.
35. van Deemter JJ (1995) Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. J. J. van Deemter, F. J. Zuiderweg and A. Klinkenberg, *Chem. Engng Sci.* 5 271-289, 1956. *Chem Eng Sci* 50(24):3867.
36. Daneyko A, Hölzel 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.
37. Gritti F (2018) A stochastic view on column efficiency. *J Chromatogr A* 1540:55–67.
38. Wenzel T (2016) Eddy Diffusion (Multipath) Broadening in Chromatography. *Chem Libr* (Figure 21):1–4.
39. Wu H -L, Means GE (1981) Immobilization of proteins by reductive alkylation with hydrophobic aldehydes. *Biotechnol Bioeng* 23(4):855–861.
40. Hober S, Nord K, Linhult M (2007) Protein A chromatography for antibody purification. *J Chromatogr B Anal Technol Biomed Life Sci* 848(1):40–47.
41. Bacskay I, Sepsey A, Felinger A (2014) Determination of the pore size distribution

of high-performance liquid chromatography stationary phases via inverse size exclusion chromatography. *J Chromatogr A* 1339:110–117.

42. Lefranc MP (2003) IMGT, the international ImMunoGeneTics database®. *Nucleic Acids Res* 31(1):307–310.

## **Unambiguous Antibody Sequence Determination by Shotgun Decision Tree Methodology**

### **Introduction**

One of the limitations of the preceding methodologies is their reliance on attaining a number of optimal characteristics for sequence analysis. In fact, the majority of peptides in a given digestion may be unsuitable for this kind of analysis; for example, a slight decrease in the charge-state distribution of the variable region peptides discussed in Chapter 3 would have likely rendered them suboptimal for pipETD sequence analysis. As such, analytical techniques typically can be quite sample dependent, and a variety of techniques need to be used to completely analyze a particular sample.

One increasingly common strategy to optimize the mass spectrometric analytical capabilities is the use of a decision tree methodology (1). This technique makes use of sophisticated instrumental software to make on-the-fly decisions about how to analyze particular analytes based on both their chemical properties as well as their instrumental demands. Fortunately, the variety of dissociation techniques available for mass spectrometric analysis combined with recent advances in state-of-the-art instrumentation have resulted in a platform which facilitates more comprehensive and informative peptide identifications, significantly improving the performance of a decision tree-based analysis (2–4). To this end, this chapter pursues the application of a decision tree-based strategy to maximally identify the variety of peptides generated by an Aspergillopepsin I digestion. The combined information extracted from all of these different digestion products is sufficient to unambiguously reconstruct the sequence of an antibody within a single LC-MS analysis.

## Background

### *Complementarity of Fragmentation Techniques*

Despite being derived from the same basic structural units, the chemical variation present across the different amino acids is sufficient to generate proteins with wildly different chemical properties. When analyzing peptides by mass spectrometry, this diversity can become problematic because the effectiveness of different dissociation techniques tends to vary significantly based on their physical and chemical properties. To date, no one fragmentation strategy has been shown to be perfectly applicable for all analytes, so it often becomes necessary to employ multiple dissociation modalities in order to maximally characterize a particular sample.

The predominant advantage of using multiple fragmentation techniques comes in the dissimilarity of their fragmentation mechanisms, largely leading to different bond cleavages of the same analyte. The most obvious source of cleavage complementarity occurs at cleavages n-terminal to proline – a cleavage that does not occur in ETD but is highly favorable in collisional dissociation (5). However, their broader cleavage specificities have been demonstrated to be complementary as well (6–8). The combination of electron based and collisional fragmentation types have been demonstrated to not only generate broadly complementary fragment ions, but often result in the identification of entirely different subsets of peptides within a dataset (1). Therefore, applying them cooperatively in a single run is likely to improve not only the variety of bond cleavages observed, but possibly even the breadth of peptides identified as well.

In addition to their generally different cleavage preferences generally, electron based and collision based fragmentation techniques have an additional facet of complementarity in that optimal fragmentation is achieved at different charge densities dataset (3, 7). As discussed previously, ETD strongly depends on charge density to achieve adequate results. However, collisional fragmentation has its own charge dependence. As the most typical fragmentation pathway for collisional dissociations is driven by the presence of mobile protons, differences in the number of mobile protons can have a significant impact in the randomness of fragmentation (3, 9, 10). It has been found that collisional fragmentation of very high charge state precursors tends to generate dominant proline cleavages, likely as a result of low proton mobility along the backbone (3, 11), meanwhile sufficiently low charge state precursors will undergo a very limited number of charge-remote fragmentation events, both of which will limit the resulting sequence coverage (10–12). Consequently, ETD proves more useful for fragmenting highly charged peptides which CAD is ill suited for, yet collision based dissociation is proficient at fragmenting poorly charged peptides where ETD struggles.

The two techniques also show differing specificities for the analysis of post-translational modifications as well. Collisional modes of fragmentation have been well established to result in preferential side-chain fragmentation of a variety of modifications, for example phosphates (13) and glycans (6), particularly when performed using the slower heating of trap-type CAD. Further, these modifications have been demonstrated to potentially scramble in the context of CAD, eliminating the possibility of confident modification site localization. Conversely, ETD results in no side-chain loss or scrambling of these modification sites, enabling their confident identification in a manner

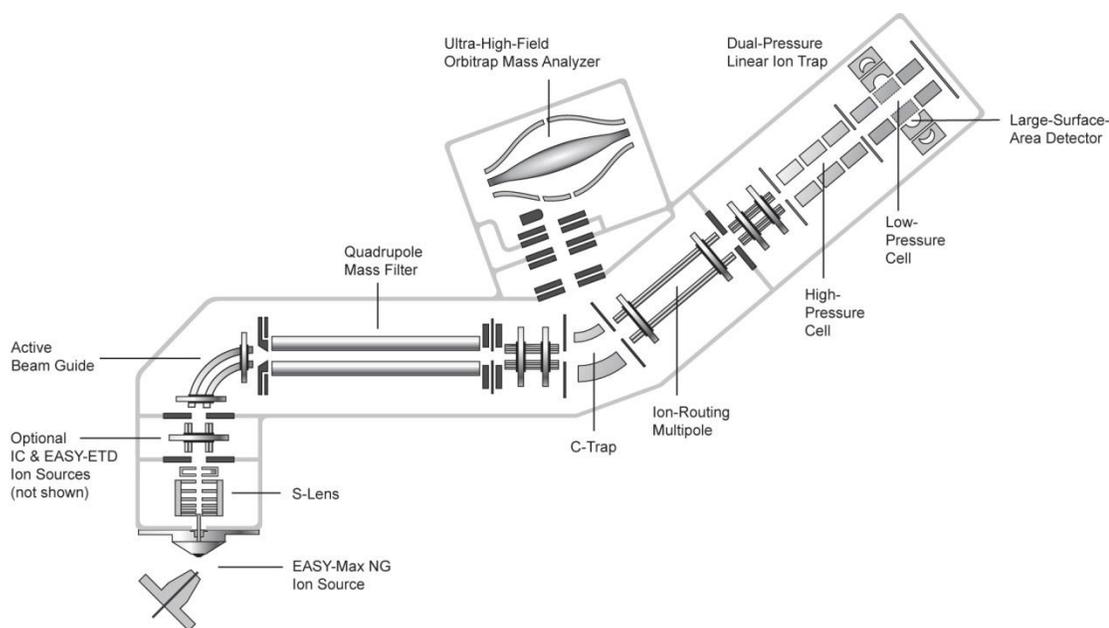
unachievable by collisional modalities (14–16). However, CAD provides a unique set of benefits for modification characterization as well. For instance, the sequential fragmentation prevalent in beam-type collisional experiments results in multiple fragmentation events of glycan side trees (6, 17). These fragmentation profiles can be useful for assigning the order of the glycan moieties or distinguishing glycan isomers, in principle. As such, the combination of fragmentation types provides useful, complementary information that can be used to better structurally evaluate a protein.

### *Instrumental Advantages of Orbitrap Fusion*

Recent advances in instrumentation have resulted in a platform far better suited for integrating a decision tree instrument method than previous iterations. The Orbitrap Fusion, the newest hybrid mass spectrometer released by Thermo Scientific, has several distinct modifications that significantly improve the ability to rapidly analyze peptide and proteins in a dynamic manner (18). A schematic representation of the Orbitrap Fusion is depicted in Figure 4.1. The instrument fundamentally combines three different mass analyzers, namely a resolving quadrupole, a dual cell linear ion trap, and a high field Orbitrap. Although previous hybrid instruments have contained these mass analyzers separately and in various configurations (19, 20), the combination of all three along with changes to their relative orientation makes for a powerful arrangement.

The primary revision to the layout compared to most previous hybrid instruments is the orientation of the two end-point mass analyzers, the ion trap and the Orbitrap. In previous instrument designs, for example the Orbitrap Elite used in Chapters 2 and 3, the ion trap is placed between the Orbitrap and the ion source at the front of the instrument.

This provided a distinct advantage, namely, that precursor ions could be isolated and/or fragmented on the way to the Orbitrap. This comes at a cost, however. It prevents any MS1 scans from being collected during any procedure which utilizes the ion trap, for example ion isolation, fragmentation, or low resolution mass analysis; this is particularly detrimental when performing ion-ion MS<sup>2</sup> reaction events, which can be quite lengthy when using sequential ion-ion reactions. As a result, the instrument's duty cycle is significantly slowed, reducing the ability to efficiently sample all eluting peptides.

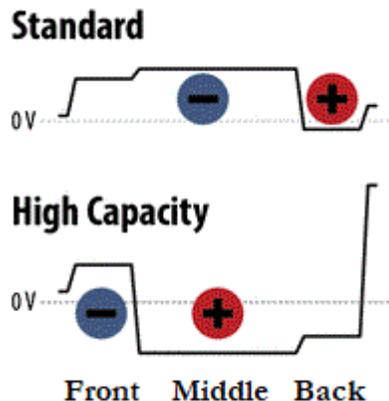


**Figure 4.1: Orbitrap Fusion Schematic.** Adapted from (18)

In the Fusion, this problem is circumvented by the addition of the front end quadrupole and the modified geometry of the ion trap and Orbitrap. The ion trap is placed after the Orbitrap rather than in front of it, even falling behind the de facto HCD collision cell, the ion-routing multipole (IRM). The foremost advantage of this modification is that ion isolation can then be predominantly performed in the more time efficient quadrupole rather than the ion trap (18). This alone substantially decreases the

operational time of the instrument. Additionally, this change also facilitates the parallelization of multiple processes to accelerate the analysis. Under this orientation, multiple types of fragmentation or mass analysis may be performed simultaneously in a way that is unachievable on previous instruments; all functions of the ion trap can be performed concurrently with all Orbitrap scans (21). This is particularly advantageous when considering multiple fragmentation types, as high resolution HCD MS2 scans can be performed in the IRM and Orbitrap while ions are also being fragmented by ETD in the ion trap.

This does come with a disadvantage in that one the strategy of performing multiple C-trap fills, a particularly useful strategy of ETD signal amplification, is no longer possible because ions must pass through the c-trap to reach the ion trap (22). However, this is largely offset by the Fusion's capability to use high capacity ETD (23). This modified ETD injection scheme is depicted in Figure 4.2. Rather than injecting precursor ions into the smaller, back section of the ion trap and reagent ions into the middle section of the trap, precursor ions are instead stored in the center section while reagent ions are stored in the front. Approximate ion capacity estimates place the front and back trap section capacities are  $\sim 2 \times 10^5$  ions and the middle section at  $\sim 1 \times 10^6$  ions (24). As such, the equivalent of 5 multiple fills (which itself is nearing the limit of what is chromatographically feasible) can be performed using this procedure. The inversion of ion populations has a slight impact on ETD reaction kinetics, but is largely be compensated for and is far less pronounced when working with larger peptides since fewer copies are present within the trap in a given scan (23).



**Figure 4.2 Difference between ion trap DC offsets in standard vs high capacity ETD.** Storage of precursor ions in the center section of the trap enables significantly higher ion targets, amplifying the spectrum signal-to-noise ratio despite comparable reaction times. Adapted from (23)

Taken together, these advantages make the use of a decision tree based shotgun methodology far more attractive in light of the improved depth of analysis enabled on this platform. As such, coupling this kind of analysis to the wealth of sequence informative peptides generated by the immobilized Aspergillopepsin reactor provides a promising opportunity to adopt an alternative sequencing strategy. Rather than being aimed at depth of analysis with respect to precursor fragmentation information, this strategy instead prioritizes depth in terms of total number of peptides evaluated in the analysis. This chapter illustrates how the use of this kind of strategy in combination with the Aspergillopepsin I immobilized enzyme reactor combines achieves sufficient information to unambiguously assign the entire sequence of a monoclonal antibody.

## 4.2 Materials and Instrumentation

### Agilent Technologies (Palo Alto, CA)

1100 Series high performance liquid chromatograph

1100 Series vacuum degasser

**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

Sizes: (360 µm o.d. x 75 & 150 µm i.d.)

**Sigma Aldrich (St. Louis, MO)**

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

Glacial acetic acid, ≥99.99% trace metal basis

Fluoranthene, >99% purity

2-propanol, LC-MS grade

Tris(2-carboxyethyl)phosphine, (>98.0%)

**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 Fusion™ Tribrid™ Mass Spectrometer

Pierce® Water, LC-MS Grade

Urea (>99.0%)

### 4.3 Methods

#### *Antibody sample preparation*

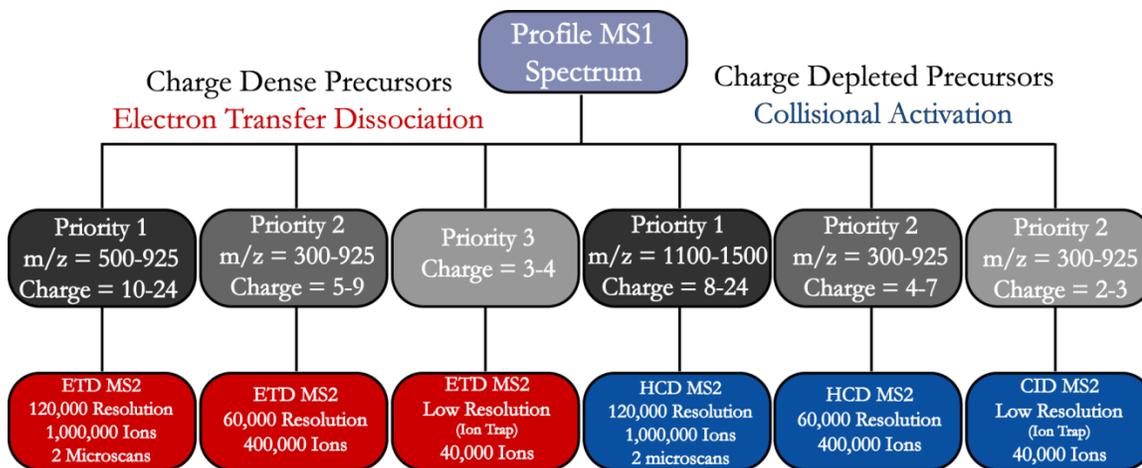
Preparation of the adalimumab sample proceeded similarly to the steps described in section 3.3.1, albeit with some slight modifications. A 2  $\mu\text{L}$  aliquot of the adalimumab stock (50  $\mu\text{g}/\mu\text{L}$ ) was similarly diluted into 50 mM ammonium bicarbonate and digested with 100 units of IdeS for 30 minutes at 37° centigrade at a final concentration of 1  $\mu\text{g}/\mu\text{L}$ . In contrast to the previously described steps, the Fc removal step by protein A following digestion was omitted so the Fc portion of the antibody remained in solution, allowing for full sequence characterization and glycan site localization in the same analysis. The sample was instead immediately dried to completeness and reconstituted in a 10  $\mu\text{L}$  solution of 10 mM TCEP in 8M urea and 0.5% acetic acid and allowed to react at 50° for 12 minutes. The sample was then neutralized to pH 6.5 with 0.2M ammonium hydroxide before being diluted with 10  $\mu\text{L}$  of 20 mM NAEM in 8M urea and 500 mM ammonium acetate and alkylated for 10 min at room temperature. Finally, the solution was diluted to 0.2  $\mu\text{g}/\mu\text{L}$  of total antibody in 0.5% acetic acid and 8M urea, reaching a final pH of 4.

The reduced and alkylated adalimumab sample was digested in the immobilized enzyme reactor. A 2.0 mm bed of 1  $\mu\text{m}$  Aspergillopepsin-conjugated particles was packed into a piece of 360  $\mu\text{m}$  OD x 150  $\mu\text{m}$  ID fused silica. The sample digested ~0.91s based on equation 3.3 and LC-MS analysis was then performed on the sample without performing any further preparation.

### ***Decision-Tree Based Mass Spectrometry***

LC-MS analysis of the ~0.91s adalimumab sample was conducted using a 10 cm PLRP-s reverse phase chromatography column (75  $\mu\text{m}$  inner diameter, 10 cm bed, 3  $\mu\text{m}$  diameter particles) using a solvent system of 0.3% formic acid in water for solvent A and 72% ACN, 18% IPA, 10% water, 0.3% formic acid for solvent B. Six hundred fmol of the digest was pressure loaded onto the column and was desalted by rinsing with solvent A at 50 bar for 30 minutes. Peptides were eluted from the column using a gradient of 25%-50%-100% solvent B in 5-80-85 minutes while heated to 50° centigrade. Eluting peptides were ionized by ESI, and analyzed on an Orbitrap Fusion Mass Spectrometer.

The eluting peptides were analyzed using a decision tree methodology, represented diagrammatically in Figure 4.3. Masses of eluting peptides were first determined using a 60,000 resolution profile Orbitrap MS1 scan. Resulting masses were then prioritized for fragmentation using data-dependently optimized parameters on the basis of both their size and charge density. Peptides of varying properties were divided into 6 non-mutually exclusive categories for fragmentation, 3 using ETD and 3 using collisional dissociation. These categories were prioritized such that peptides of a higher priority class would be fragmented first even if peptides of a lower priority class were present and more abundant while prioritization within a class was defined by peptide abundance. Each category also possessed its own dynamic exclusion list to enable peptides with charge states meeting multiple criteria to be fragmented by both relevant fragmentation techniques.



**Figure 4.3: Orbitrap Fusion Decision Tree Method.** Peptides were prioritized on the basis of their size to dynamically increase resolution, ion target, and scan averaging. High-charge precursors were subjected to ETD while low-charge precursors were subjected to CAD. Most peptides are likely to contain charge-states in both categories, allowing them to be characterized complementarily by both types of dissociation.

Highest priority precursors were large peptides with charge states  $\geq 10$  and  $m/z$  from 500-925 (indicating high charge density) or charge states  $\geq 8$  and  $m/z$  from 1100-1500. These size ranges were fragmented by ETD or collisional activation types, respectively. Scans in these categories were analyzed in the Orbitrap at a resolution of 120,000 with two microscans, both using ion populations of  $1e6$ . Elevated ETD ion targets were enabled by the use of high capacity ETD and a reagent population of  $2e5$  ions. Collisional fragmentation used stepped collision energy HCD fragmentation with normalized collision energy of 22-25-28%, each for one third of the total ion population. Second priority precursors consisted of medium-size peptides, falling between charge states 5-9 and  $m/z$  300-925 for ETD fragmentation or charge states 4-7 for stepped HCD fragmentation with no  $m/z$  requirement. These peptides were analyzed at 60,000 resolution with no additional microscans using ion populations of  $4e5$  and the same fragmentation parameters. Lowest priority precursors were peptides with charge-states from 3-5 for ETD and 2-4 for collisional activation. ETD fragmentation was performed

as in the other categories, but collisional activation instead used ion trap CID fragmentation with a normalized collision energy of 30%. Lowest priority scans were analyzed in the linear ion trap at “normal” scan rate. All ETD fragmentation used for all fragmentation categories used charge-dependent ETD reaction times based off of the instrument’s kinetics calibration, under these conditions corresponding to Equation 4.1:

$$t_{etd} = \frac{3^2}{Charge\ State^2} \times 57.28 \quad (4.1)$$

Provisional identification of analyzed peptides was performed using the Byonic (Protein Metrics) (25) search algorithm node housed within Proteome Discoverer version 2.2.0.386 (Thermo Scientific). The data was searched against a database containing the sequence of adalimumab using fully nonspecific cleavage pattern to allow possible cleavage sites between all residues of the protein. High resolution (FT) and low resolution (IT) MS/MS scans were segmented into separate search categories and searched with different parameters. FT scans were searched with a 10 ppm precursor mass tolerance and a 15 ppm fragment mass tolerance for both ETD and HCD fragmentation types. IT scans were also searched with a 10 ppm precursor mass tolerance, but a 0.35 Da fragment ion mass tolerance was used for fragment ions due to the low resolution MS/MS scans. The maximum allowed precursor mass was 20,000 Da for FT scans and 10,000 Da for IT scans.

The search algorithm allowed for a variety of modifications to the peptides. The static mass shift associated with the NAEM modification was assigned to cysteines, and the lack of an NAEM modification was included as a variable modification (in case incomplete alkylation occurred within the sample). Other variable modifications used in both searches included N-terminal pyroglutamylation of glutamine or glutamic acid,

oxidation of methionine, hexose glycation at serine or threonine, HexNAc glycation at serine or threonine, and N-linked glycosylation at asparagine and glutamine. Deamidation was allowed for IT scans only. N-linked glycan trees searched were G0 (4 HexNAc, 3Hexose), G0F (4 HexNAc, 3 Hexose, 1 Fucose), G1 (4 HexNAc, 4 Hex), G1F (4 HexNAc 4, 4 Hexose, 1 Fucose), G2 (4 HexNAc, 5 Hexose), and G2F (4 HexNAc, 5 Hexose, 1 Fucose).

Following automated identification of peptides, composite sequence coverage was derived from a subset of peptides identified within the search. Peptides were evaluated on the basis of their 2 dimensional posterior error probability (PEP-2D) from the Byonic output as well as the automated peak annotation presented in the Byonic viewer. Peptides positively evaluated in both categories were selected for manual annotation. The relevant MS/MS spectra under the chromatographic peak of each peptide species were averaged and the peaks in the resulting averaged spectra were inspected and peaks were assigned to peptide fragment ions. In total, 40 of the 777 uniquely identified peptides were used for manual annotation, equaling ~5% of the dataset. The composite sequence coverage across each of the antibody chains was then generated by superimposing the complementary sequence information from the amino acid assignments made in the manually annotated spectra.

#### **4.4 Results and Discussion**

One of the most distinct characteristics of the Aspergillopepsin I enzyme reactor is the sheer complexity of the resulting digests; proteins like antibodies can contain several hundred amino acids, and the lack of a well-defined cleavage specificity (26–28)

creates a bewildering number of potential peptide products. Consider a protein with 500 amino acids. Naturally, if enzymatic cleavages along the backbone can occur at any of these locations, it results in a total 499 potential cleavage sites. Given that a given pair of enzymatic cleavages can occur randomly across the molecule, digesting a protein of this size could in principle produce 124,251 unique digestion products. This doesn't even consider the further complexity associated with variable post translational modifications being present on those peptides as well.

In a way, this has resulted in inefficient data extraction from the sample in all previous methodologies used to analyze these digests. The previously reported work on the reactor that targeted several species in certain elution windows as well as the targeted parking experiments discussed in Chapter 3 both leave the vast majority of peptides in the analysis entirely uncharacterized. Although many of these peptides may contain redundant sequence information, they allow for both complementary fragmentation information owing to different chemical characteristics, as well as further confirmation for cleavages which may be difficult to confidently assign from a single MS/MS spectrum.

Much of this analytical insufficiency is the result of hardware limitations present in older generations of mass spectrometers. Mass spectrometers, like the Orbitrap Elite used for previously discussed parking experiments, are limited in their ability to analyze multiple precursors in a time-dependent manner; the instrument architecture limits the number of processes which can be completed simultaneously due to properties like needing to pass ion through the ion trap to get to the Orbitrap. As a result, whenever any process using the ion trap was being conducted, no other procedures, whether HCD cell

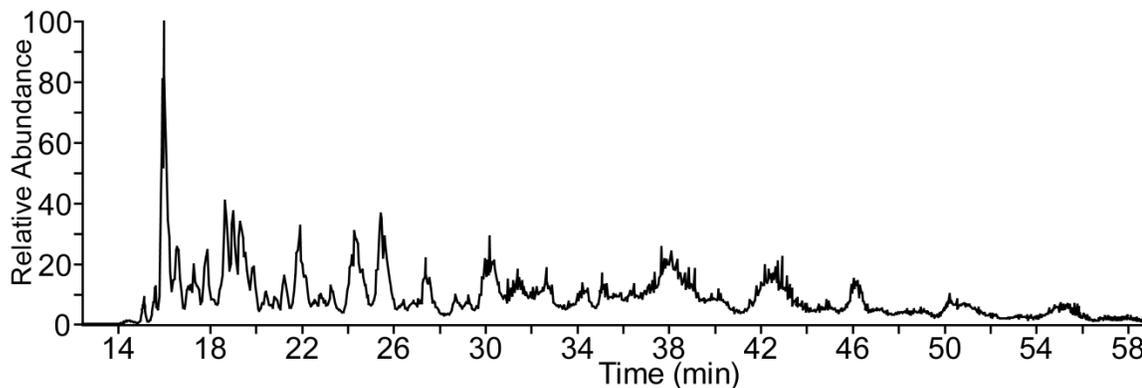
fragmentation or MS1 analysis, could be conducted at the same time. By comparison, the architecture of the Orbitrap Fusion enables these kinds of processes to be conducted synchronously, resulting in a much higher analyte throughput and increasing depth of analysis for very complex samples.

Capitalizing on the improved analysis time enabled by this kind of instrumentation, we developed a tiered decision tree methodology in order to maximize the number of peptides analyzed across a run. The primary aim of this strategy was to achieve protein sequencing through the composite information across the population of peptides rather than deep analysis of a few. This strategy makes a trade-off in that the Orbitrap Fusion currently lacks some of the powerful MS/MS tools like ion parking and IIPT, but the complementary information provided from multiple peptides and fragmentation types is likely to compensate for this loss.

Peptides were prioritized based on the analytical requirements to extract useful sequence information from them (29), primarily modifying resolution, ion population size, and signal averaging to spend as little time as possible analyzing a particular peptide. For example, larger peptides are more likely to require higher resolution and more signal averaging in order to produce highly informative MS/MS spectra. Large peptides eluting in a given window were therefore both prioritized for fragmentation and analyzed using much higher ion populations and resolutions than their smaller counterparts, saving valuable instrument time and enabling better depth of analysis. Furthermore, despite being analyzed at the lowest priority, peptides considered small enough to be analyzed in the ion trap rather than the Orbitrap were analyzed largely simultaneously with the large peptides.

Fragmentation conditions were also modified on the basis of the eluting peptide's properties. Peptides were required to possess relatively high charge density ( $>925$  m/z) to be selected for ETD fragmentation in order to minimize analysis time lost by fragmenting lower quality precursors. Similarly, charge-states at or below  $\sim 1200$  m/z were selected for HCD fragmentation to minimize charge-remote fragmentation pathways while also avoiding the highly selective cleavages that occur from high charge density. By tailoring fragmentation conditions to a particular analyte in this manner, the information extracted from any individual peptide was maximized.

To illustrate the utility of this approach, this strategy was applied to a  $\sim 0.9$ s adalimumab digestion similar to the one used for parking experiments in Chapter 3. The most noteworthy difference is the inclusion of Fc portion of the molecule. Although this further increases the sample complexity, it allows for the confirmation of the full antibody sequence as well as Fc glycan characterization. The total ion chromatogram (TIC) of this analysis is depicted in Figure 4.4. The digestion profile is primarily dominated by many large peptides toward the end of the gradient, but it also contains a plethora of much smaller digestion products which tend to elute much earlier in the gradient, on average.



**Figure 4.4: Total ion chromatogram of  $\sim 0.9$ s enzyme reactor digestion**

Peptides present within the dataset were identified automatically by searching against the antibody's sequence in Byonic. In total, over 700 unique peptides were identified within the run at a 1% false discovery rate. Often, as many as a dozen different unique peptides cover a given section of the protein, offering multiple opportunities to extract sequence information out of a particular region. Given the immense amount of information present in the analysis, a subset of peptides was selected for manual annotation. These peptides were selected on the basis of both the score assigned by the Byonic search algorithm and their automatically annotated sequence coverage shown within the Byonic data viewer. Only forty total peptides, equating to ~5.5% of the dataset, were selected for manual inspection. Ten of the peptides were from the light chain, fourteen were from heavy chain Fd' peptides, and sixteen were from heavy chain Fc/2, as depicted in Figure 4.5.

DIQMTQSPSSLSASVGDRVTITCRASQGIR  
NYLAWYQQKPGKAPKLLIYAASSTLQSGVPS  
RFSGSGSGTDFTLTISSLQPEDVATYYCQR  
YNRAPYTFGQGTKVEIKRTVAAPSVFIFPP  
SDEQLKSGTASVVCLLNNFYPREAKVQWKV  
DNALQSGNSQESVTEQDSKDYSLSSSTLT  
LSKADYEKHKVYACEVTHQGLSSPVTKSFN  
RGEC

EVQLVESGGGLVQPGRSLRLSCAASGFTFD  
DYAMHWVRQAPGKGLEWVSAITWNSGHIDY  
ADSVEGRFTISRDNKNSLYLQMNSLRAED  
TAVYYCAKVSYLSTASSLDYWGQGTLVTVS  
SASTKGPSVFPLAPSSKSTSGGTAALGCLV  
KDYFPEPVTVSWNSGALTSGVHTFPAVLQS  
SGLYSLSSVVTVPSSSLGTQTYICNVNHKP  
SNTKVDKKVEPKSCDKTHTCPPCPAPELLG  
  
GPSVFLFPPKPKDTLMISRTPEVTCVVVDV  
SHEDPEVKFNWYVDGVEVHNAKTKPREEQY  
NSTYRVVSVLTVLHQDWLNGKEYKCKVSNK  
ALPAPIEKTISKAKGQPREPQVYTLPPSRD  
ELTKNQVSLTCLVKGFYPSDIAVEWESNGQ  
PENNYKTTTPVLDSDGSFFLYSKLTVDKSR  
WQQGNVFSCSVMHEALHNHYTQKSLSLSPG  
K

**Figure 4.5: Map of total peptide overlap and the peptide subset selected for manual sequence evaluation for light chain (top), heavy chain Fd' (middle), and heavy chain fc/2 (bottom) from adalimumab ~0.9s reactor digestion.**

DIQMTQSPSSLSASVGDRVTITCRASQGIR\_  
NYLAWYQQKPGKAPKLLIYAASTLQSGVPS\_  
RFSGSGSGTDFTLTISSLQPEDVATYYCQR\_  
YNRAPYTFGQGTKVEIKRTVAAPSVFIFPP\_  
SDEQLKSGTASVVCLLNNFYPREAKVQWKV\_  
DNALQSGNSQESVTEQDSKDSTYSLSSTLT\_  
LSKADYEKHKVYACEVTHQGLSSPVTKSFN\_  
RGEC

EVQLVESGGGLVQPGRSLRLSCAASGFTFD  
 DYAMHWVRQAPGKGLEWVSAITWNSGHIDY  
 ADSVEGRFTISRDNAKNSLYLQMNSLRAED  
 TAVYYCAKVSYLSTASSLDYWGQGTLVTVS  
 SASTKGPSVFPPLAPSSKSTSGGTAALGCLV  
 KDYFPEPVTVSWNSGALTSGVHTFPAVLQS  
 SGLYSLSSVVTVPSSSLGTQTYICNVNHKP  
 SNTKVDKKVEPKSCDKTHTCPPCPAPELLG  
  
 GPSVFLFPPKPKDTLMISRTPEVTCVVVDV  
 SHEDPEVKFNWYVDGVEVHNAKTKPREEQY  
 NSTYRVVSVLTVLHQDWLNGKEYKCKVSNK  
 ALPAPIEKTISKAKGQPREPQVYTLPPSRD  
 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQ  
 PENNYKTTTPVLDSDGSFFLYSKLTVDKSR  
 WQQGNVFSCSVMEALHNHYTQKSLSLSPG  
 K

**Figure 4.6:** Composite sequence coverage derived from the annotated subset of light chain (top) heavy chain Fd' (middle) and heavy chain Fc/2 (bottom) subunits. ETD and collision derived cleavages are denoted in red and blue, respectively.

Following sequence evaluation, the resulting sequence coverage maps from both different peptides and different fragmentation types were superimposed to generate the composite sequence coverage achieved by only this subset of peptides. A condensed version of this cleavage map is depicted in Figure 4.6.

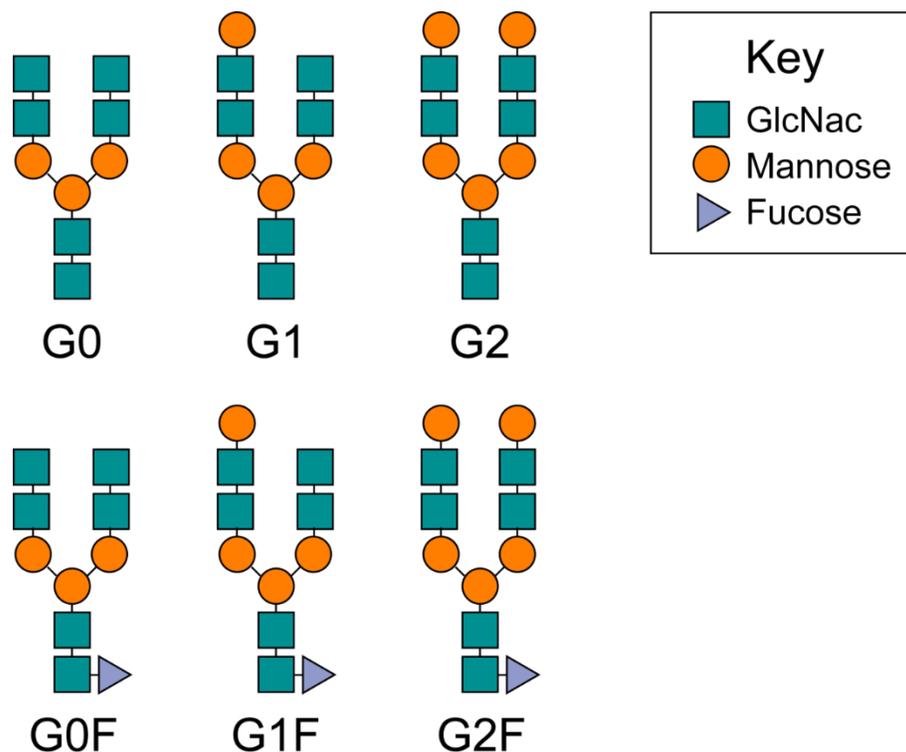
Significantly, a diagnostic fragment ion for nearly every bond cleavage in the molecule was present in at least one of the analyzed peptides, and the vast majority of bonds were characterized across multiple peptides and fragmentation types. The only

bond lacking a confirmatory fragment ion was the K<sub>151</sub>-D<sub>152</sub> bond on the heavy chain Fd'. This cleavage appears to be a highly favorable site for aspergillopepsin proteolysis, being present on the vast majority of peptides within that region. Spectra of peptides spanning this region of the protein were typically much lower in quality due to their very low abundance. Fortunately, the order of amino acids in this region may still be reasonably assigned on the basis of overlapping peptides. Flanking peptides are clearly identified which contain the adjacent K and D cleavages, identifying K as being on the N-terminal side of the proteolysis site, while D is present on the C-terminal side. Further, large, overlapping peptides can be used to exclude the possibility that additional amino acids are present between the K and D residues, indirectly confirming the K|D cleavage site despite the lack of direct fragmentation data.

Significantly, neither ETD nor HCD were individually capable of achieving complete sequence assignment within the subset of selected peptides. ETD and HCD each achieved 80% and 89% sequence coverage on the selected peptides, respectively, when only considering data of each fragmentation type independently. However, in combination, these two fragmentation techniques are powerful enough to enable the full sequence mapping of the protein with such a limited subset of the data. Further, additional sequencing redundancies are likely present in the remaining 94.5% of the dataset as well, allowing one to corroborate any weak or unambiguous assignments if necessary.

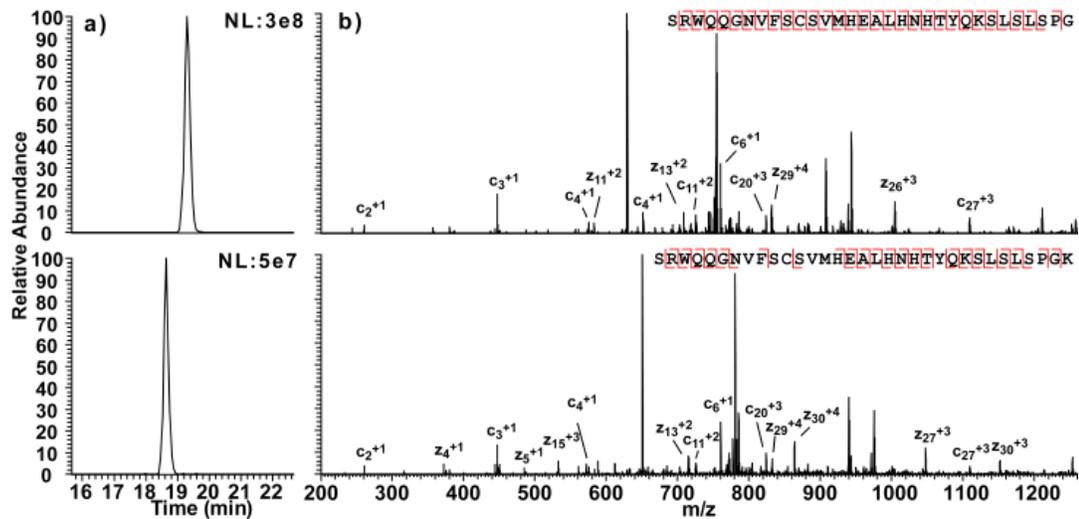
The glycan tree was also successfully site localized to Fc/2 residue N31 (or N271 of the entire heavy chain). Several different glycan trees were searched for during the run, namely the G0, G0F, G1, G1F, G2, and G2F glycan trees. The structure of these trees is

pictured in Figure 4.7. The predominant glycan tree identified within the analysis was the G0F tree; however, the G1F tree was also identified within the run. Glycan site localization was conducted in the G0F variant. The same G1F peptides were successfully identified within the analysis, but MS/MS spectral quality was noticeably lower as a result of their reduced abundance, and as such, the G1F glycan could not be definitively site localized. Notably, sequence characterization of glycosylated peptides occurred almost entirely through the use of ETD. HCD spectra tended to exhibit the dominant loss of the glycan tree, and significantly more sparse fragment ions, particularly for fragment ions also containing an intact glycan tree. While in principle the antibody could be deglycosylated prior to analysis, these results suggest that glycan identification remains a possibility in glycosylated samples when ETD is used as a mode of fragmentation.



**Figure 4.7: Structure for common antibody glycans searched for in adalimumab digestion**

C-terminal lysine clipping was found in the sample based on the peptides identified (30). Peptides identified on the c-terminal portion of the antibody contain two apparent variants, one which contains a c-terminal lysine residue and one which doesn't. One example of such a peptide pair which was fragmented and identified, namely S<sub>179</sub>-K<sub>211</sub> and S<sub>179</sub>-G<sub>210</sub>, is pictured in Figure 4.8. Both peptides generated fragment-rich MS/MS spectra upon dissociation, clearly illustrating the 128 Da mass shift associated with the additional lysine residue. Further, they are also present at similar abundances, suggesting that both variants are present within the sample to a significant capacity. This is strongly suggestive of these variants being a result of Fc C-terminal lysine clipping, a common modification among antibodies. However, the nonspecific nature of the enzymatic digestion makes it difficult to know definitively that these peptides are genuine lysine clipping modifications on the original sample and not artifacts from the digestion. While this is plausible given their similar abundances and the ubiquity of the modification on antibodies, these truncations cannot be definitively assigned within the molecule as it isn't unmistakably clear that the truncation was a feature on the intact molecule rather than a product of the nonspecific digestion.



**Figure 4.8: Lysine clipped and unclipped peptide identified within the analysis.** Peak The extracted ion chromatogram peak intensity (left) was about 10-fold more abundant for the clipped species compared to the unclipped species. Both species were well characterized by ETD fragmentation, confirming that the mass shift was derived from an additional lysine residue on the C-terminus.

Despite this slight limitation, these results show that the use of kinetically controlled, nonspecific proteolysis coupled to a decision tree-based methodology is indeed a powerful tool for sequence characterization. Compared to the current field standard, the sample preparation is incredibly rapid and possesses significant sample flexibility based on the use of highly denaturing conditions; this ensures more consistent cleavages and improved solubility over a wider variety of proteins than more typical strategies allow. Further, the depth of information extracted is superior by allowing for amino acid localization based on explicit fragmentation data rather than simple peptide mapping as is the standard. This enables better discrimination of amino acid inversions and localization of additional modifications. Given these advantages, kinetically controlled proteolysis represents a promising technology for detailed characterization of large, highly diverse proteins with antibodies representing one of many useful applications.



## References

1. Swaney DL, McAlister GC, Coon JJ (2008) Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* 5(11):959–964.
2. Mikesch LM, et al. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta - Proteins Proteomics* 1764(12):1811–1822.
3. Watson DJ, McLuckey SA (2006) Charge state dependent ion trap collision-induced dissociation of reduced bovine and porcine trypsin cations. *Int J Mass Spectrom* 255–256(1–3):53–64.
4. Brodbelt JS (2014) Photodissociation mass spectrometry: New tools for characterization of biological molecules. *Chem Soc Rev* 43(8):2757–2783.
5. 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* 101(26):9528–9533.
6. Hogan JM, Pitteri SJ, Chrisman PA, McLuckey SA (2005) Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J Proteome Res* 4(2):628–632.
7. Zubarev RA, Zubarev AR, Savitski MM (2008) Electron Capture/Transfer versus Collisionally Activated/Induced Dissociations: Solo or Duet? *J Am Soc Mass Spectrom* 19(6):753–761.
8. Good DM, Wirtala M, McAlister GC, Coon JJ (2007) Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol Cell Proteomics* 6(11):1942–1951.
9. Gu C, Tsaprailis G, Breci L, Wysocki VH (2000) Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Anal Chem* 72(23):5804–5813.
10. Foreman DJ, Dziekonski ET, McLuckey SA (2019) Maximizing Selective Cleavages at Aspartic Acid and Proline Residues for the Identification of Intact Proteins. *J Am Soc Mass Spectrom* 30(1):34–44.
11. Palzs B, Suhal S (2005) Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* 24(4):508–548.
12. Wysocki VH, Tsaprailis G, Smith LL, Breci LA (2000) Mobile and localized protons: A framework for understanding peptide dissociation. *J Mass Spectrom* 35(12):1399–1406.
13. 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.

14. Riley NM, Coon JJ (2018) The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal Chem* 90(1):40–64.
15. Makarov A, et al. (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* 78(7):2113–2120.
16. Chi A, et al. (2007) Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci* 104(7):2193–2198.
17. Malaker SA, et al. (2017) Identification and Characterization of Complex Glycosylated Peptides Presented by the MHC Class II Processing Pathway in Melanoma. *J Proteome Res* 16(1):228–237.
18. 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.
19. Michalski A, et al. (2011) 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.
20. 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.
21. Brunner AM, et al. (2015) Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization. *Anal Chem* 87(8):4152–4158.
22. Earley L, et al. (2013) Front-end electron transfer dissociation: A new ionization source. *Anal Chem* 85(17):8385–8390.
23. 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.
24. Rose CM, et al. (2013) Multipurpose dissociation cell for enhanced ETD of intact protein species. *J Am Soc Mass Spectrom* 24(6):816–827.
25. Bern M, Kil YJ, Becker C (2012) Byonic: Advanced peptide and protein identification software. *Curr Protoc Bioinforma* 40(SUPPL.40):13.20.1-13.20.14.
26. Ahn J, Cao MJ, Yu YQ, Engen JR (2013) Accessing the reproducibility and specificity of pepsin and other aspartic proteases. *Biochim Biophys Acta - Proteins Proteomics* 1834(6):1222–1229.
27. Templin MF, Høltje J-V (2013) *Handbook of Proteolytic Enzymes* doi:10.1016/B978-0-12-382219-2.00317-3.
28. Zhang H-M, et al. *Enhanced Digestion Efficiency, Peptide Ionization Efficiency,*

*and Sequence Resolution for Protein Hydrogen/Deuterium Exchange Monitored by FT-ICR Mass Spectrometry* Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2784605/pdf/nihms86585.pdf> [Accessed April 21, 2019].

29. 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.
30. Cai B, Pan H, Flynn GC (2011) C-terminal lysine processing of human immunoglobulin G2 heavy chain in vivo. *Biotechnol Bioeng* 108(2):404–412.

## **Conclusion**

The protein sequencing strategies presented in this work represent a substantial advancement compared to current state-of-the-art technology. Each strategy possesses its own strengths and weaknesses, but applying each judiciously can result in superior characterization compared to current methodologies.

Chapter 2. Further, this set of experiments is particularly novel as it represents an implementation of ETD parallel ion parking on a chromatographic timescale as well as making use of selective IdeS digestion and reduction to simplify the analysis of the antibody subunits(1). While IdeS digestion strategies have been reported previously, they typically use fairly standard fragmentation strategies and only achieve ~50% sequence coverage in a single LC-MS analysis (2, 3). Instead, using pipETD/IIPT in this work dramatically improves the results compared to these analyses, enabling the observation of fragment ions corresponding to ~81% of all ETD cleavable bonds (~75% of all total bonds) across the antibody. Advancements with respect to fragmentation include the chromatographic implementation of pipETD when analyzing these subunits. To date, pipETD experiments have been performed by infusing the protein of interest directly into the instrument. Here, we instead make use of this technology on a chromatographic timescale which inherently comes with disadvantages in both cycle time and variation in ion flux (which can negatively impact the parking process). Nonetheless, pipETD was successfully applied in this context, illustrating its capability as a chromatographically feasible fragmentation strategy. Further advancements of the both ion parking and ETD as a whole are likely to further improve the capabilities of this technique. While the selective enzymatic digestion is exclusively useful for antibody analysis, the implementation of

these fragmentation tools will be broadly applicable to the study of any large peptides or proteins.

Chapter 3 illustrates the potential of using a precisely controlled digestion with a nonspecific protease in order to liberate large subunits of a particular molecule. Under normal circumstances, these proteases would typically digest proteins into relatively small peptides, even dipeptides if allowed to proceed unimpeded. However, the use of precise sub-second digestion on an immobilized reactor allows fine enough control of this process that the average digestion product can be prevented from undergoing more than one or two enzymatic cleavages. To date, digestion strategies aimed at generating subunit-sized pieces have typically been limited to the kind of single-cleavage digestion used in Chapter 2, or else they succeed in co-localizing 2 CDRs at most (4). This work shows how using precisely timed cleavages on an immobilized, nonspecific enzyme can circumvent these problems in a largely sequence-independent manner. Further, this strategy would have likely been successful had improvements not been made to our previous enzyme reactor scaffold. As noted, the peptide size distribution from digestions on our previous 20  $\mu\text{m}$  porous scaffold generated a very broad distribution in terms of peptide size, resulting in low yield of particularly large peptides (5). Instead using a 1  $\mu\text{m}$  solid scaffold as illustrated in this work improved the yield of this process substantially enough that these subunit-sized pieces could be generated far more efficiently.

This strategy is likely to be somewhat broadly applicable, and has already been applied separately to other molecules in a similar manner. One open question for this type of strategy is whether or not sufficiently complex samples or modifications to the sample matrix will substantially impede the ability to generate relevant peptides. One reason for

the high yields of these large peptides is likely a sufficiently high concentration such that the intact protein can out-compete the largest peptides until they can diffuse away from the bead surface. As such, proteins digested in a mixture may behave quite differently than when digested alone, making extrapolation of appropriate digestion parameters quite difficult. Further, it is unknown how well the yield on peptides in this size range will scale as protein size increases, as the largest molecule that has been analyzed using this strategy was ~29 kDa. It may be more difficult to generate high-yield, equivalently sized pieces when generating those pieces is reliant on more than a single enzymatic cleavage given the diversity of cleavage sites along the molecule.

Finally, the enzyme reactor shotgun sequencing strategy illustrated in Chapter 4 represented a substantial improvement over standard peptide-mapping strategies owing to a few unique features. Most attempts at exhaustive peptide mapping of an antibody rely on multiple, separate digestions and analyses in order to generate an overlapping sequence reconstruction in the composite dataset (6, 7). As these in-tube reactions are typically lengthy, lasting hours, and multiple LC-MS analyses need to be performed in order to generate the necessary information, this is a very lengthy process. The use of the overlapping peptide series generated by our reactor in effect condenses these many experiments to a single experiment which can be performed in its entirety in about 4 hours. Further, the improved extent of peptide overlap as well exploitation of peptide complementarity fragmentation (both in terms of dual fragmentation techniques and profiles of different peptides) allows for an additional level of confidence in this data. Protein sequences are identified on the basis of observed fragment ions rather than simply the purported peptide identification, as is the norm. As such, this experimental strategy

represents a substantial improvement in both sample analysis time as well as data quality compared to standard techniques.

This strategy also represents a substantial improvement relative to our previous implementations of the reactor digestions. Original proof-of-concept experiments were performed using three separate LC-MS runs in order to perform analysis on a set of 39 selected peptides using both ETD and CID to achieve 95% sequence coverage after combining all of these runs (5). The new decision tree strategy, conversely, identifies ~777 peptides in a single LC-MS run and also achieves full sequence characterization using only a subset of analyzed peptides. As such, it represents over an order of magnitude increase in depth of analysis (on the basis of unique peptides characterized) in one third of the previous instrumental analysis time.

However, the most important aspect of this strategy may be that it possesses the greatest extensibility of the three digestion strategies proposed here. As this strategy relies strictly on peptide overlap and complementarity, deviations in reactor performance will have a smaller effect than they would otherwise. While an overreacted sample will generate smaller segments of overlap and fail to co-localize more distant portions of the molecule, fragmentation parameters will be automatically optimized for this smaller peptide size range, and a composite sequence may still be reconstructed from a larger set of smaller peptides. In this way, offloading a portion of the experiment optimization to instrumental portion of the analysis adds a level of flexibility to the experimental procedure which may help compensate for complications associated with using the limited reactor digestion. In its current state, the decision tree methodology currently fails to achieve the kind of optimal analysis of the largest peptides, primarily as a result of lacking pipETD and IIPT on the

Orbitrap Fusion platform. However, with future instrument advancements, these two strategies can be merged in principle, creating a compound strategy which possesses the benefits of both.

In summary, the work presented in this dissertation represents a substantial improvement protein sequencing. Both precisely controlled limited proteolysis and gas-phase chemistry represent powerful tools which are largely underutilized when analyzing proteins, but which offer significant advantages over existing technology, as illustrated in this work. Hopefully, these strategies will prove useful in answering biological questions which remain elusive to current methodologies, and further expansion on these techniques will establish a new, useful toolkit for the proteomics community.

## References

1. An Y, Zhang Y, Mueller HM, Shameem M, Chen X (2014) A new tool for monoclonal antibody analysis Application of IdeS proteolysis in IgG domain-specific characterization. *MAbs* 6(4):879–893.
2. 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.
3. Fornelli L, et al. (2018) Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal Chem* 90(14):8421–8429.
4. Srzentić K, et al. (2014) Advantages of Extended Bottom-Up Proteomics Using Sap9 for Analysis of Monoclonal Antibodies. *Anal Chem* 86(19):9945–9953.
5. Zhang L, et al. (2015) 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. Resemann A, et al. (2016) mAbs Full validation of therapeutic antibody sequences by middle-up mass measurements and middle-down protein sequencing. doi:10.1080/19420862.2015.1128607.
7. Sen KI, et al. Automated Antibody De Novo Sequencing and Its Utility in Biopharmaceutical Discovery. doi:10.1007/s13361-016-1580-0.







