Ion-ion Chemistry and Parallel Ion Parking for Advances in Mass Spectrometric Analysis of Intact Proteins

Scott Andrew Ugrin
Lake Forest, CA

B.A Chemistry, California State University, San Bernardino, 2012

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, 2017

Acknowledgments

I would like to first thank Professor Donald Hunt for allowing me to join his lab and for the absolute freedom to explore my scientific curiosity. Don, I will never forget the email I received from you when I first expressed an interest in joining your group. I hope I was able to meet your expectations and reward your willingness to take a chance on me. Next, I need to thank Dr. Jeffrey Shabanowitz for his invaluable mentorship and support during my time in graduate school. Jeff, you have taught me everything I know about mass spectrometry machines (please don't hurt me). You taught me to respect them and how to think through problems systematically. I will always miss our lunches together and our very important meetings in your office to discuss fantasy football roster moves. I'd also like to thank the emotional support of the Hunt lab, Dr. Dina Bai. Dina, this lab would fall apart without the escape that is your office. Whether I needed help with a program, looking at my data, or editing a paper, or just needed to vent for a moment, your office has been a refuge. I appreciate all the advice you've given me over the years.

I would be remiss if I didn't thank all the Hunt lab members who have come before me and contributed greatly to my graduate school experience. Specifically, I would like to thank Drs. Paisley Myers, Stacy Malaker, and Amanda Wriston. You three have been amazing friends and it has been awesome getting to know you youngins. I'd also like to thank Drs. Lissa Anderson, Michelle English, and Weihan Wang. You all taught me so much and were such great colleagues. I owe a majority of my success to the three of you.

I appreciate the collaboration and friendship of all of the current members of the lab. Specifically, I'd like to thank Benjamin Barnhill and Stephanie Miller Lehman. SSC for life! Ben, I truly enjoyed working with you and especially enjoyed going to the basketball games together. Stephanie, you are a great scientist and I appreciate you putting up with all my jokes. Thanks for being a good sport.

I would also like to thank several people outside the Hunt lab that contributed to my graduate work. Dr. John Syka is possibly the smartest person I've ever met. John, it was fun and very informative working those late nights in the lab with you. I probably absorbed about 10% of the knowledge you tried to instill in me and I know it frustrated you at times but I thank you very much for your patience with me. Lee Earley and Drs. Chris Mullen, Chad Weisbrod, and Will McGee provided insight and ideas for experiments during our weekly conference calls. You all provided a great sounding board for my (sometimes dumb) ideas. I would also like to thank Drs. Scott McLuckey and Jim Stephenson. I only briefly met Scott, but the work of his group at Purdue University, which included Jim, provided the foundation for the work reported in this dissertation.

Finally, I would like to thank my family. Mom, Dad, Spoon, and Molly: I want to thank you for all of the love and support that you have given me my entire life. Most importantly, I need to thank my wife, Lindsey. I could not have finished graduate school without you by my side. I love you and can't wait to see what life has in store for us next.

Table of Contents

List of Figures and Tables

List of Abbreviations

ºC: Degrees celcius

•   Radical species

Å: Angstroms

AA: Amino acid

AC: Alternating current

AGC: Automatic gain control

Ala, A: Alanine

amu: Atomic mass units

API: Atmospheric pressure ionization

Arg, R: Arginine

Asn, N: Asparagine

Asp, D: Aspartic acid

c: Centi (1×10^{-2})

CAD: Collision-activated dissociation

CDR: Complimentarity determining region

CI: Chemical ionization

CID: Collision induced dissociation

CSIT: Charge-sign independent trapping

Cys, C: Cysteine

Da: Dalton

DC: Direct current

DTT: Dithiothreitol

ECD: Electron capture dissociation

ESI: Electrospray ionization

ETD: Electron transfer dissociation

ETnoD: Electron transfer with no dissociation

f: femto $(1x10^{-15})$

FETD: Front-end electron transfer dissociation

FT: Fourier transform

G: Gauge or gravity

Gln, Q: Glutamine

Glu, E: Glutamic acid

Gly, G: Glycine

HC: Heavy chain

HCD: Higher energy collisionally-activated dissociation

HCl: Hydrochloric acid

His, H: Histidine

HPLC: High-performance liquid chromatography

hr: Hour

HV: High voltage

Hz: Hertz

i.d.:   Inner diameter

IgG: Immunoglobulin gamma

IIPT: Ion-ion proton transfer

Ile, I: Isoleucine

IPC: Isotope pattern calculator

ITCL: Ion trap control language

kDa: kilodalton

L: Liter

LC: Liquid chromatography or Light chain

Leu, L: Leucine

LMCO: Low mass cutoff

LTQ: Linear trapping quadrupole

Lys, K: Lysine

μ: mico ($1x10^{-6}$)

m: Milli ($1x10^{-3}$) or meter

M: Mass or molar

m/z: Mass-to-charge ratio

M+nH: Molecular ion with n charges

MALDI: Matrix-assisted laser desorption ionization

Met, M: Methionine

min: Minute

mol: Mole ($6.022x10^{23}$)

M: Molar

MS: Mass spectrometry

MS/MS: Tandem mass spectrometry

MSn: Tandem MS carried to the nth stage

MW: Molecular weight

n: Nano $(1x10^{-9})$

NAEM: N-(2-aminoethyl)maleimide

NL: Normalized level

o.d.:   Outer diameter

p: Pico $(1x10^{-12})$

PDCH: Perfluoro-1,3-dimethylcyclohexane

PEEK: Polyether(ethylene)ketide

PFD: Perfluorodecalin

PFMD: Perfluoromethyldecalin

Phe, F: Phenylalanine

PMSF: Phenylmethanesulfonyl fluoride

ppm: Parts per million

Pro, P: Proline

PTM: Post-translational modification

PUGNAc: N-acetylhexosaminidase inhibitor

R: Resolution

RF: Radio frequency

RPM: Revolutions per minute

S: Svedberg

Ser, S: Serine

SRIG: Stacked ring ion guide

TCEP: Tris(2-carboxyethyl)phosphine

TEABC: Triethylammonium bicarbonate

Thr, T: Threonine

TIC: Total ion current

TMG: 1, 1, 3, 3-tetramethylguanidine

TMK: Tris-HCl, $MgCl_2$, and KCl Buffer

Trp, W: Tryptophan

Tyr, Y: Tyrosine

Val, V: Valine

Abstract

Mass spectrometry (MS) is considered the preferred tool for the structural characterization of proteins. Typical protein studies by MS involve breaking proteins into small pieces prior to analysis. Recently, there has been a concerted effort in the field of protein mass spectrometry to enable analyses of proteins in their intact form. This type of analysis allows for the ability to differentiate between closely related species, a feature not offered by peptide-based methods. Ion-ion chemistry, namely electron transfer dissociation (ETD) and ion-ion proton transfer (IIPT), has become a staple in interrogating the primary structure of intact proteins. However, detection of whole proteins using traditional mass spectrometry methods has proven difficult given their high molecular weight. The ideal MS method would allow for the identification of all proteins in a sample mixture. This includes the characterization of the entire amino acid sequence and the site-localization of all post-translational modifications (PTMs) on each protein species. However, low signal-to-noise ratios of intact proteins in both the full mass spectra and tandem mass spectra (MS/MS) limit our ability to perform this type of analysis.

Here we describe the implementation and application of novel MS methods for enhancing the signal-to-noise ratio of intact proteins and their component fragment ions. We accomplish this by selectively controlling the kinetics of products of gas phase ion-ion reactions using a technique called parallel ion parking. Using these methods we demonstrate near-complete sequence coverage of the intact protein apomyoglobin (17 kDa). In addition, we are able to preserve large fragment ions during ETD of the light

chain of a monoclonal antibody (25 kDa). Further, we introduce parallel ion parking during IIPT which allows us to concentrate the signal of several unknown proteins concurrently. We use this technique to develop an automated method of identifying intact proteins in a mixture, which range in mass from 4.5 to 61 kDa. We accomplish this on a chromatographic time-scale using IIPT and data-dependent higher energy collisionally activated dissociation (HCD).

Chapter 1. Introduction to the Dissertation

1.1 Introduction

Proteins are essential to cellular function. As such, abnormalities in protein structure or expression are implicated in most diseases. Proteomics is the field of study that aims to characterize and quantify an organism's protein compliment [1]. This discipline relies heavily on mass spectrometry (MS), which is uniquely appropriate for the study of biomolecules due to its high sensitivity, wide dynamic range, and ability to perform tandem MS (MS/MS) experiments [2-4]. Typical protein MS studies involve enzymatic digestion of proteins to form smaller peptides prior to analysis. Recently, however, there has been a push in the field of proteomics to study proteins as intact molecules. The goal of intact protein mass spectrometry is to unambiguously identify proteins, including their sequence and post-translational modifications (PTMs) [5, 6]. To be of practical application, this must be achieved on a chromatographic timescale (approximately 1 to 2 minutes) to allow for adequate separation of biological mixtures. However, for reasons which will be discussed, this goal is not yet within reach. This dissertation will attempt to address several problems associated with two aspects of MS proteomics: 1) determination of the amino acid sequence of a single protein, including all post-translational modifications, and 2) unambiguous identification of all proteins in a biological mixture. We will focus on various problems associated with the analysis of intact proteins using current techniques and our efforts to address these problems. We will discuss novel techniques for employing and controlling ion-ion chemistry to advance our ability to reach these goals. Finally, we will demonstrate the ability of these

techniques to overcome the challenges associated with analyzing biological samples.

Prior to a discussion of the challenges we look to address, this chapter will explore concepts and technology currently employed in modern protein mass spectrometry. We will discuss ionization techniques, mass analyzers, and fragmentation methods which are essential to understanding the novel techniques we employ in later chapters. We aim to address instrumental limitations by controlling the rate of gas-phase ion-ion reactions through selective manipulation of product ions in a linear trap quadrupole (LTQ) ion trap. Therefore, we will discuss the equations of ion motion in the LTQ and the kinetics of ion-ion reactions, with the goal of understanding ways in which these reactions may be controlled. With this knowledge in mind, we will explore the advantages and limitations offered through the analysis of proteins as intact molecules in comparison to peptide-based techniques.

1.2 Ionization Techniques

The first step in mass spectrometric analysis is converting a solution phase molecule into a gas-phase ion. Modern protein mass spectrometry methods typically employ one of two methods for analyte ionization: matrix-assisted laser desorption ionization (MALDI) or electrospray ionization (ESI). MALDI is performed by sublimating samples out of a solid crystalline matrix using high energy laser pulses [7]. Analytes ionized by MALDI typically gain a single proton and exhibit a 1+ charge state. This limits the size of molecules that can be analyzed in ion trap/Orbitrap hybrid

instruments because these instruments have a scan range of up to 4,000 m/z. Because of

this, MALDI is usually coupled with time-of-flight instruments, which are not limited in

their m/z range. Additionally, samples ionized by MALDI must be crystallized prior to

analysis, making chromatographic separation impossible. Thus, samples which are to be

ionized by MALDI must be highly purified.
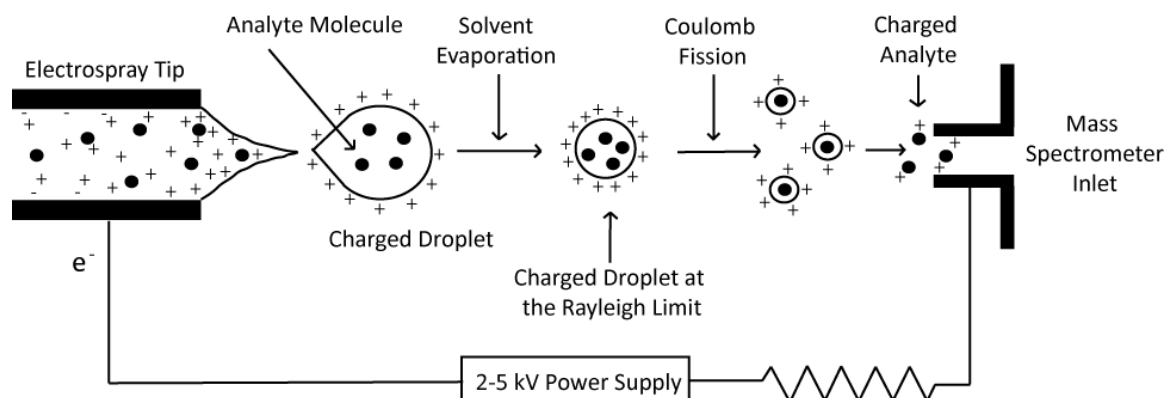
Figure 1.1: Cartoon depiction of the electrospray ionization process. Droplets are desolvated by high capillary temperatures until they reach the Rayleigh limit. Once a droplet is completely desolvated, protons are deposited onto basic sites on the analyte and converted from solution to the gas phase.

An alternative to MALDI is ESI, which can be coupled with liquid

chromatography separation methods because the analytes are ionized directly from

solution [8]. ESI is a soft ionization technique that was first reported in 1989 by John

Fenn [9]. The process of ESI is outlined in **Figure 1.1**. Briefly, proteins are ionized by

applying a high voltage (2-5 kV) to an electrospray tip through which a solution

containing the analyte is passed. The inlet to the mass spectrometer is kept at ground. The

potential difference causes a buildup of protons at the tip interface. Droplets are expelled

from the tip toward the mass spectrometer inlet due to the **c**oulombic repulsion of the

protons. As the droplets travel through the heated capillary, the water molecules that

compose the droplet to evaporate in a process called desolvation. During this process, the

repulsive forces between the protons overcome the surface tension of the water molecules in the droplet. At this point, known as the Rayleigh limit, the droplet explodes into smaller droplets that contain, ideally, a single analyte and several protons. As the droplet continues to desolvate, conditions become energetically favorable for protons to inhabit basic sites on the analyte (lysine, arginine, or histidine side chains and the N-terminal amine) [10, 11]. Because multiple protons can be deposited onto a single analyte, ESI allows for the detection of very large protein species in instruments that offer 4,000 m/z scan ranges. For example, a 30 kDa proteins that obtains 30 protons during ESI will have an m/z of 1,000.
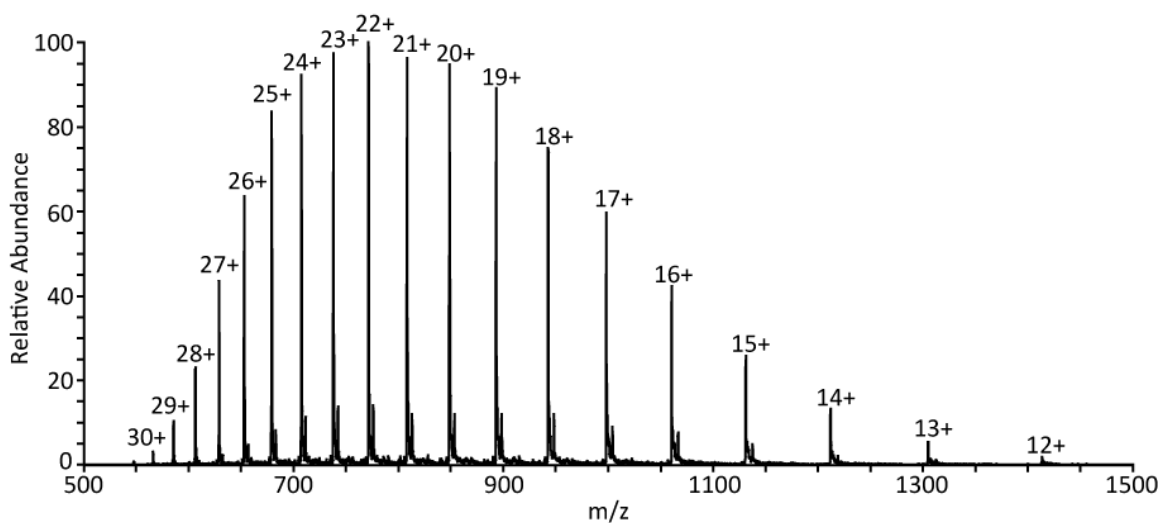
Figure 1.2: Electrospray ionization mass spectrum of the 17 kDa protein apomyoglobin. Protein analytes with a large number of basic sites will exhibit a wide range of charge states by ESI. The same protein that has accepted a different number of protons during ESI will have differing integer z values and be detected at a different m/z.

It is important to note that all basic sites will not necessarily become protonated during ESI. Side chain basicity is affected by several factors [12, 13]. For example, it is energetically unfavorable for two adjoining amino acids to be protonated due to

coulombic repulsion. Therefore, analytes which contain several basic side chains can exhibit a wide range of charge states. **Figure 1.2** shows the ESI spectrum of the 17 kDa protein apomyoglobin. Although the sequence contains 33 basic sites, the protein exhibits charge states from $[M+12H]^{12+}$ to $[M+30H]^{30+}$.

1.3 Mass Analyzers

The essential function of a mass spectrometer is to measure the mass-to-charge ratio (m/z) of ions in the gas phase. Time-of-flight (TOF), Fourier transform ion cyclotron resonance (FT-ICR), and ion mobility mass analyzers are commonly used in protein mass spectrometry but are beyond the scope of the work discussed here. For our purposes, we will focus on the linear ion trap and Orbitrap mass analyzers.

1.3.1 Linear Ion Trap

The most common mass analyzer for protein mass spectrometry is the linear trap quadrupole (LTQ) ion trap. These analyzers are inexpensive, highly sensitive, and operate at a rapid speed, allowing for data to be acquired on a chromatographic time scale [4]. An important advantage of the LTQ is that it affords the ability to perform tandem MS experiments [14]. **Figure 1.3** shows a schematic representation of a linear ion trap. In gas phase chemistry, the LTQ can be thought of as the reaction vessel, similar to a beaker in solution chemistry.

The function and theory of the LTQ will be discussed in detail in **Section 1.4**.

Briefly, the m/z of an ion is determined by the frequency at which it oscillates in the trap under a given set of trapping parameters. Each ion oscillates at a different frequency as it is trapped in the radio frequency (RF) field which is generated by applying alternating positive and negative potentials to opposing rods of the trap. In addition to the trapping potential, a single frequency supplemental RF is applied to the rods in the x-dimension of the trap. Ions with an oscillation frequency that match the frequency of the supplemental potential are ejected from the trap through slits in the x-rods. Ejected ions collide with a conversion dynode that emits electrons which are then detected by electron multipliers. The trapping voltage amplitude is adjusted such that the entire m/z range is scanned through the ejection frequency. The time at which an ion is detected in the scan is determinate of its m/z.

Several characteristics of the LTQ have led it to become the workhorse mass analyzer of protein mass spectrometry. Data is acquired quickly, resulting in the acquisition of several hundred scans per minute. Additionally, electron multipliers produce very little electronic noise, making the LTQ one of the most sensitive mass analyzers available. Finally, the LTQ offers the ability to perform MS/MS experiments. Specifically, the linear ion trap enables concurrent trapping of oppositely charged particles, a feature essential in performing ion-ion chemistry. MS/MS fragementation provides a deeper level of protein characterization that is essential to identifying biomolecules.
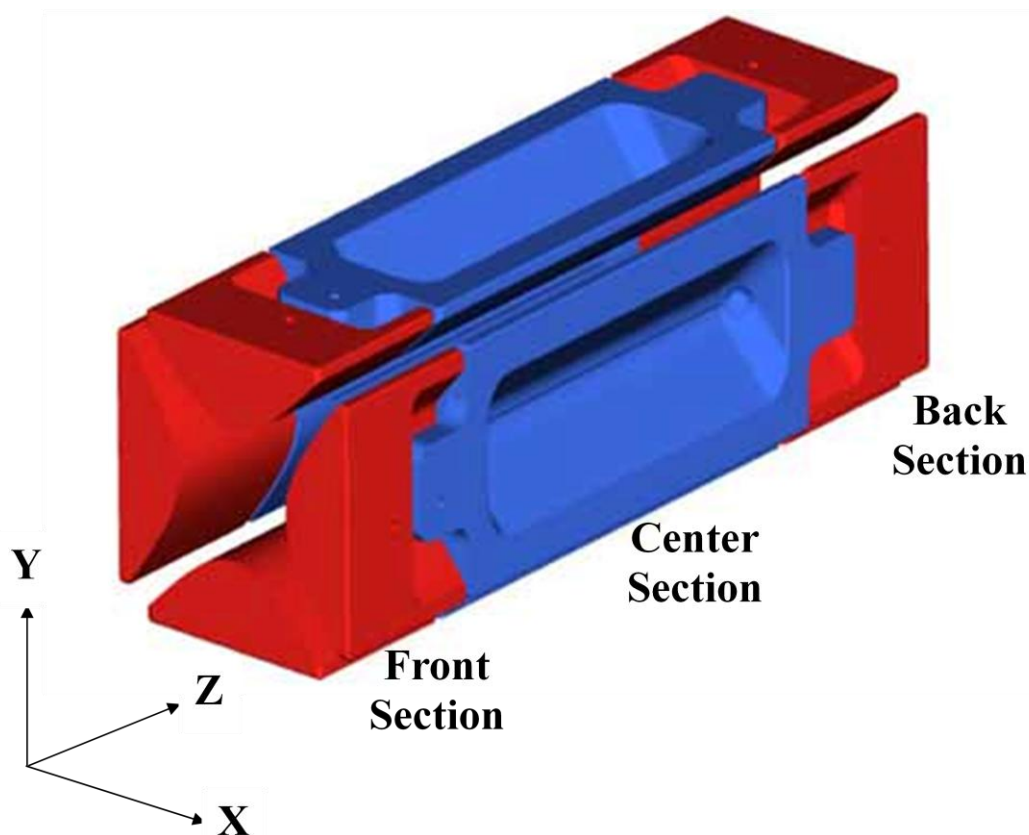
Figure 1.3: Schematic of a linear ion trap. Linear ion trap with separate sections annotated. Ions enter along the z-axis from the front section. Slight DC offsets applied to each section allow manipulation of ion populations from one section to another. This figure was modified from [14].

1.3.2 Orbitrap

One disadvantage of the LTQ is the inability to perform high resolution mass analysis. The LTQ is only able to resolve ions that differ by 1 m/z. This capability is sufficient for peptide-based approaches where detected ions may carry only two or three charges and thus produce singly or doubly charged fragments. However, when detecting large proteins that are highly charged, a higher resolving power is necessary. The Orbitrap (**Figure 1.4**) consists of a central spindle electrode, around which ions orbit, and

an outer barrel electrode that serves as the detector [15]. Prior to injection into the

Orbitrap, ions to be detected are held in the C-trap, a device similar to the LTQ but with a

curved shape. This architecture serves to "squeeze" the ions into a compact packet prior

to injection, assuring that all ions reach the detector at the same time [16]. Ions are guided

from the C-trap into the Orbitrap by a series of electrodes and enter the detector via a slit

that is offset from the center of the spindle electrode [17]. At this time, the voltage of the

spindle electrode is ramped from ground to a negative potential in a carefully timed

manner. This causes the ion packet to enter into an orbit around the spindle electrode,

inducing its radial motion which is independent of m/z.

Axial motion of the ions is m/z-dependent and is induced by a restoring force

toward the center of the Orbitrap. The electrostatic field produced by the spindle

electrode is uneven along the electrode's axis due to the electrode's shape. The field is at

a maximum at the center of the trap, causing ions to oscillate along the electrode as they

orbit it [18]. The frequency of an ion's oscillation is m/z-dependent, as seen in **Equation
1.1**:

$$\textbf{(1.1)} \quad \omega = \sqrt{\frac{z}{m}} \; x \; k$$

where ω is the frequency, and k is the force constant of the Orbitrap field. The outer

barrel electrode is split into two sections along its central axis. Ions that are oscillating in

the Orbitrap induce an image current in the time domain that is detected by the outer

electrode. The image current is then converted to the frequency domain by Fourier

Transform (FT), and then to corresponding m/z values.
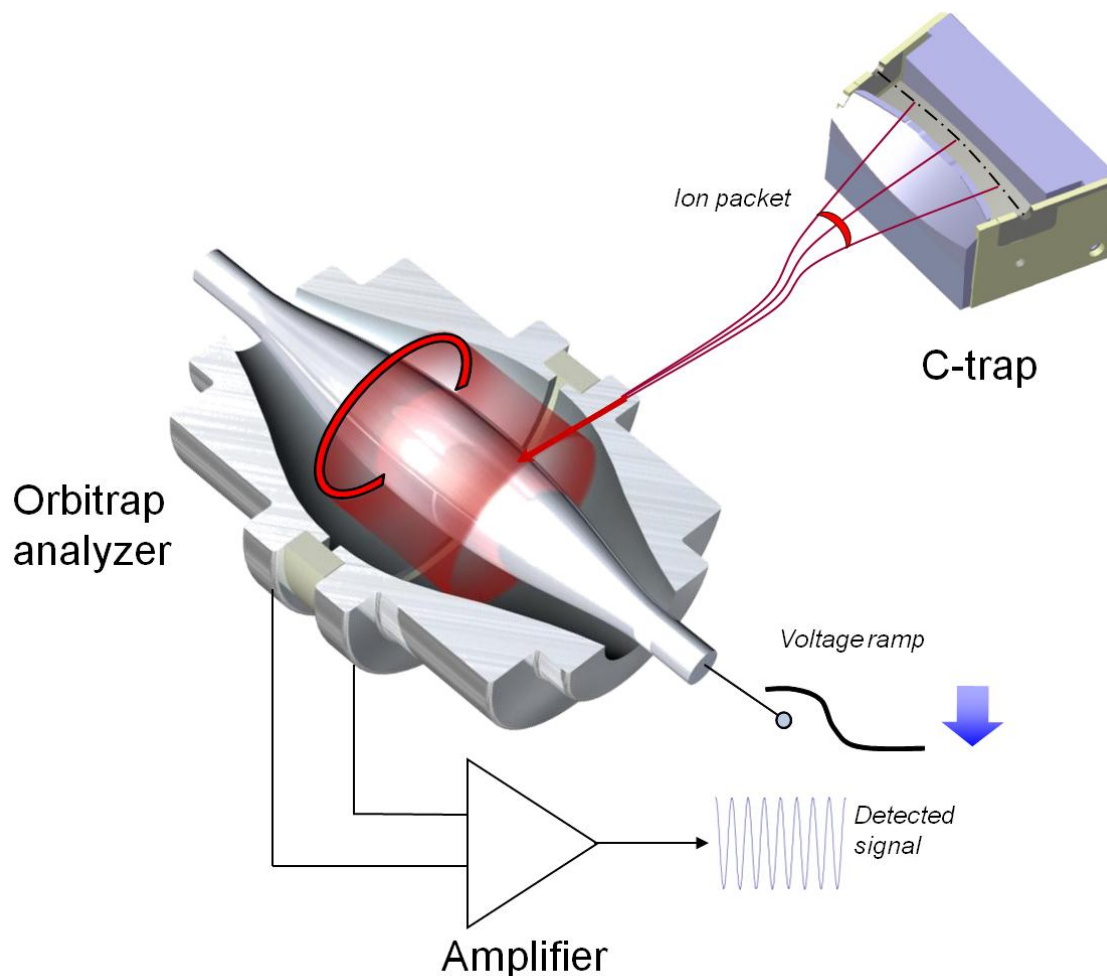
Figure 1.4: Cutaway view of an Orbitrap mass analyzer. The Orbitrap measures the oscillation frequency of ion packets as they orbit the central electrode. Ions are injected from the C-trap off axis to the central electrode and oscillate at a frequency corresponding to their mass-to-charge ratio. This image was modified from [18]

The Orbitrap mass analyzer theoretically offers unlimited resolution. There is a linear relationship between image current (also known as transient) acquisition time and resolution. Increasing resolution *ad infinitum* simply requires the ions to remain in orbit and oscillating along the central electrode. However, this would require a complete vacuum in the Orbitrap chamber. In reality, the pressure in the Orbitrap is approximately

2 x 10^{-11} torr. While very low, this places limits on the practical resolution for the Orbitrap system. Accurate mass detection in the Orbitrap requires that ions of the same m/z remain in coherent rings while they are detected. As the transient acquisition time is increased to achieve higher resolution, some ions collide with the background gas. This causes the rings to spread out axially, leading to the introduction of noise to the image current. The frequency of collisions is dependent on the surface area of the ions being detected. Ions with a larger surface area do not maintain coherence as long as smaller ions.

The mass of large ions is determined by multiplying its m/z by the charge state it occupies. In high resolution mass spectra, the charge state is determined by calculating the m/z difference between isotope peaks (i.e. an ion with all ^{12}C carbons versus an ion with one ^{13}C carbon): Isotopes of a $[M+5H]^{5+}$ ion will differ by 0.2000 m/z, isotopes of a $[M+10H]^{10+}$ ion will differ by 0.1000 m/z, etc. For practical use, Orbitrap mass analyzers can achieve resolutions sufficient to determine the isotopic m/z difference of ions with several dozen charges. The ability to distinguish charge states of ions is especially important when analyzing complex spectra resulting from fragmentation of intact proteins.

1.4 Linear Ion Trap Theory

To measure the m/z of ions, the LTQ first has to trap and hold the population of ions to be analyzed. As the ions enter the LTQ, they become trapped in the z-dimension

(**Figure 1.3**) by a combination of slight DC potentials applied to each section and the end

lens of the trap along with a bath gas of helium [14,19]. The ions are held in the x- and y-

dimension by a quadrupole field produced by applying an RF trapping potential to the

rods of the LTQ [20]. The strength with which each ion feels the quadrupole field is

described by solutions of the Mathieu equation [21] (**Equation 1.2**):

$$(\mathbf{1.2}) \quad \frac{\delta^2 \mu}{\delta \xi} + \left(a_\mu - 2q_\mu cos2\xi\right)\mu = 0$$

Where μ is the ion's displacement in the x, y, or z dimension; ξ is equal to $\Omega t/2$ (Ω is the

RF frequency and t is time); and a and q are unit-less parameters that describe the

stability of an ion in a DC and RF field, respectively (**Equation 1.3 and 1.4**):

$$(\mathbf{1.3}) \quad a_u = \frac{-8eU}{m\Omega^2 r_0{}^2} \qquad (\mathbf{1.4}) \quad q_u = \frac{4eV}{m\Omega^2 r_0{}^2}$$

Here, *e* is the charge of an electron, U and V are DC and RF field amplitudes,

respectively, *m* represents the mass-to-charge ratio (m/z) of the ion, Ω is the trapping RF

frequency, and r_0 is the trap radius.

The solutions to the Mathieu equation, when plotted in terms of a- and q-values

provide the Mathieu stability diagram (**Figure 1.5**). Combinations in the shaded areas are

considered stable in a given dimension, with the overlapping region denoting the

combinations that are stable in both dimensions.
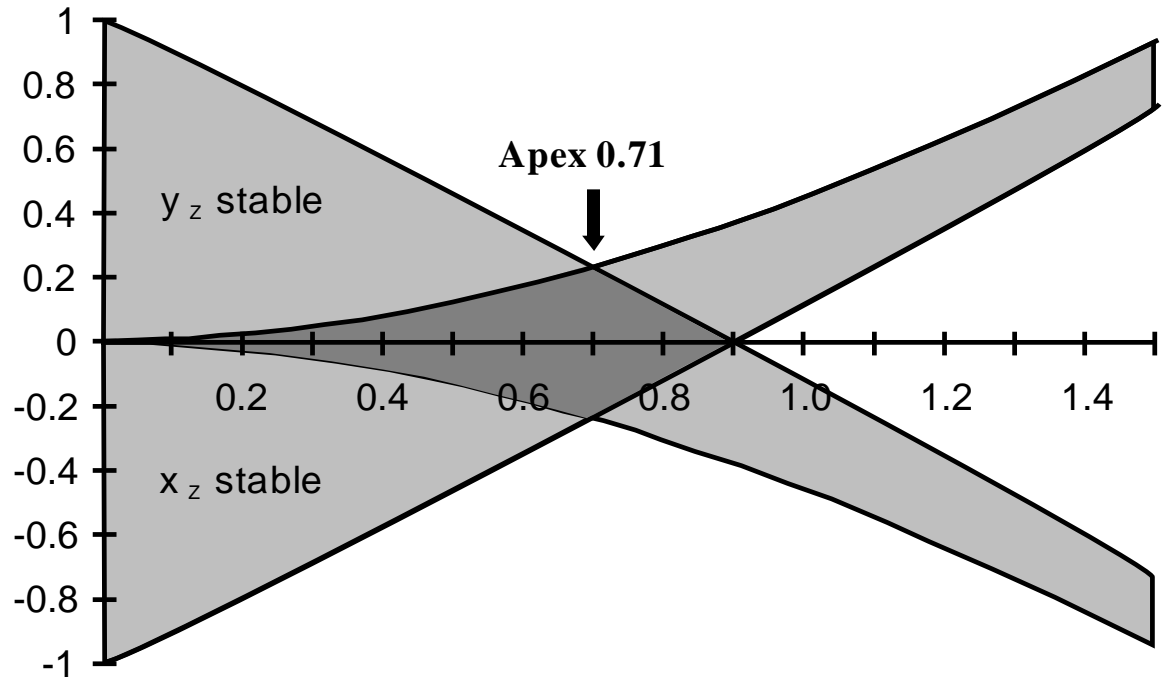
Figure 1.5: Stability diagram showing combined radial stability of ions within a linear ion trap. An ion with a Mathieu q-value (x-axis) and a-value (y-axis) combination is considered stable in the x- or y-dimension if it falls within the shaded regions. For an ion to be stable in both dimensions, it must have an a, q combination which falls within the darkly shaded area. Ions with a Mathieu a-value of zero are stable if they have a q-value between 0 and 0.908.

In the linear ion trap the amplitude of the aforementioned DC potentials used to trap ions are low enough (typically a few volts) to be considered negligible. Therefore, q alone determines an ion's stability. An ion with a q between 0 and 0.908 is considered to be stable. A trapped ion is directed toward the center of the trap by the alternating attractive and resistant forces induced by the RF field. The restoring force toward the center of the trap is similar to the spring constant in classical mechanics. Ions with a high m/z have more momentum as they reach the apex of their motion and feel the effects of the field less than ions with a low m/z. The restorative force felt by each ion induces a secular frequency in the x- and y-dimension. The frequency (ω) in each direction is approximated by **Equation 1.5**:

$$(\mathbf{1.5}) \quad \omega = \frac{q_x \Omega}{2\sqrt{2}}$$

Substituting for q, we can determine the frequency of a given m/z using **Equation 1.6**:

$$(\mathbf{1.6}) \quad \omega = \frac{4eV}{m\Omega r_0{}^2 2\sqrt{2}}$$

In this equation are three variables that are relevant to mass analysis in the LTQ: frequency (ω), trapping RF amplitude (V), and m/z (m). During a mass scan, an additional RF potential is applied to only the x-rods of the LTQ. This supplemental RF is lower in frequency than the main trapping RF (~200 kHz vs. 1.145 MHz). Ions with a secular frequency that match the frequency of the supplemental RF are resonantly excited and ejected from the trap through the slits in the x-rods and detected by the electron multiplier. As the amplitude of the trapping RF is ramped, each ion is ejected as its frequency matches the supplemental RF, meaning the m/z is determined by the time at which it is detected in the ramp. The supplemental RF potential and the excitation it imparts on resonant ions are important to many aspects of MS/MS analyses and is the basis for the novel methods that will be presented in this dissertation.

1.5 Fragmentation and Gas-Phase Reactions

In addition to intact mass measurement, unambiguous identification of a protein requires the ability to determine the protein's amino acid sequence. The sequence of a protein determines the way it folds and interacts with other proteins. It is thus an important piece of information that protein mass spectrometry looks to unlock. In

addition, amino acid side chains can be covalently decorated by chemical modifications that are not encoded by the genome and are added after protein translation. These post-translational modifications (PTMs) affect the function of the protein, thus their detection is an essential component to full characterization of the protein [5, 22-24]. In order to determine an amino acid sequence, proteins are fragmented along the peptide backbone. Different fragmentation mechanisms result in the breaking of various bonds along the protein backbone. The nomenclature for these fragments is illustrated in **Figure 1.6**. Fragmentation of the Cα-C bond results in one N-terminal-containing a-ion and one C-terminal-containing x-ion. Separation of the peptide bond results in one N-terminal-containing b-ion and one C-terminal-containing y-ion. Separation of the N-Cα bond results in one N-terminal-containing c-ion and one C-terminal-containing z-ion. The combined mass of a complimentary a and x pair, b and y pair, or c and z ion pair will equal the mass of the intact protein. Ideally, fragmentation of a population of proteins will result in fragmentation at each position in the protein. The difference in mass between subsequent fragments (i.e. $c_3 - c_2$) is calculated to determine the mass of the amino acid at that position. To induce fragmentation of the peptide bond, several fragmentation methods, each with their own benefits, have become commonplace in tandem mass spectrometry.
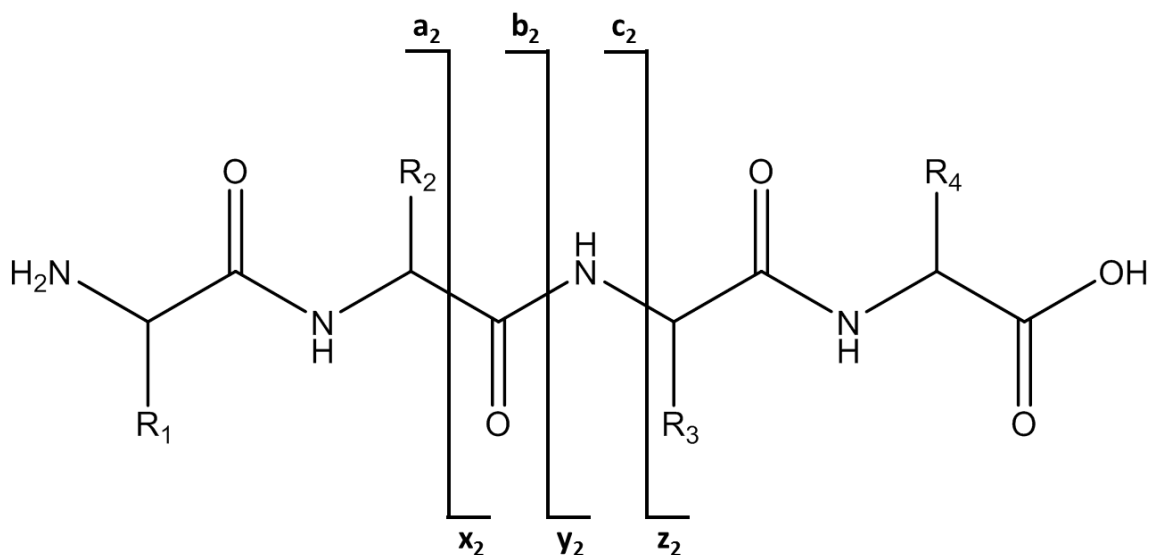
Figure 1.6: Fragmentation scheme of the peptide backbone. Peptidic species fragmented at one of the bonds that make up the peptide backbone form two complimentary fragment ions of a- and x-type, b- and y-type, or c- and z-type.

1.5.1 Collisionally Activated Dissociation (CAD)

CAD imparts energy into the protein via thousands of collisions with helium molecules. To begin, the precursor is accelerated using a single frequency RF potential to increase its kinetic energy [3, 25, 26]. Upon colliding with a helium molecule, a small amount of that kinetic energy is converted to internal energy. Eventually, the energy necessary to fragment the peptide bond and produce b- and y-ions is reached. CAD is considered a "slow heating" fragmentation method in that several thousand collisions are necessary to reach the energy barrier for fragmentation. This leads to preferred fragmentation pathways in which the weakest bond is generally the first bond cleaved. In this case, the resulting MS/MS spectrum is dominated by fragments produced by this preferred fragmentation. Many important post-translational modifications, such as phosphorylation and O-linked N-acetylglucosamine, are weakly bound to their

substituents [27]. This makes the loss of the modification the preferred pathway and results in an MS/MS spectrum in which the primary peak is the mass of the precursor less the mass of the modification. Two difficulties arise in this situation: (1) very few sequence-informative fragments are detected and (2) site-localization of the modification is impossible if more than one potential modification site exists.

Sequence coverage by CAD is also limited by the inherent low mass cutoff introduced by the mechanism of fragmentation. To ensure maximum stability and fragmentation efficiency, precursor ions are activated at a Mathieu q-value of 0.25 during CAD. The Mathieu q stability limit of ions in the ion trap is 0.908, or approximately 27% of the precursor m/z. Thus, low mass fragment ions are unstable in the trap and are not detected [21].

One advantage of CAD is that secondary fragmentation is rare. As mentioned previously, a single frequency is used to accelerate the precursor. The frequency of the activation potential matches the oscillation frequency of the precursor only. Thus, products that differ in m/z from the precursor are no longer excited and do not undergo any further collisions.

CAD is the ideal technique for fragmentation of tryptic peptides. Trypsin is a serine protease that cleaves proteins C-terminal to the basic residues arginine and lysine. A majority of the peptides resulting from a tryptic digest thus have two sites available for

protonation by ESI: the amino terminal and the side chain of the C-terminal residue. However, some tryptic peptides are triply protonated due to a missed cleavage by trypsin, a histidine residue which can also be protonated, or because the cleavage site is next to proline [28].

Intact proteins do not fragment well by CAD. Investigations into the mechanism of CAD fragmentation have theorized that a mobile proton - a proton that is delocalized along the peptide backbone - is necessary to induce fragmentation [29, 30]. Basic residues present a problem in that any nearby charge is localized to the amino acid side chain. Thus, the proton is unavailable to participate in CAD fragmentation therefore reducing the chance that fragmentation will occur. Fragmentation by CAD is especially problematic for the analysis of intact proteins because of the many basic residues upon which a proton can be localized.

1.5.2 Higher Energy Collisional Dissosciation (HCD)

HCD is an alternative vibrational fragmentation technique that is not hindered by some of the limitations inherent to CAD. The energy of the collision is increased by using a higher mass bath gas, such as nitrogen or argon as opposed to helium [31]. Rather than inducing fragmentation via several hundred low energy collisions between the cation and the bath gas, the collisional energy resulting from a single collision is sufficient to fragment the precursor. Thus, each bond position is equally likely to be cleaved. Like CAD, HCD results in b- and y-ions. HCD has the added benefit of not being limited by a

low mass cutoff for product ions. Precursor ions are accelerated in the axial direction from the C-trap to the HCD collision cell by a DC potential offset. Because the RF trapping potentials are not changed during this process, no stability limit is introduced that could lead to the ejection of low mass ions.

Fragmentation of intact proteins by HCD provides richer MS/MS spectra than CAD; however, some problems remain. First, fragments resulting from HCD can undergo further collisions with the bath gas to produce internal fragments. This complicates the MS/MS spectra with non-sequence-informative signals. In addition, fragments derived from cleavage of peptide bonds in certain areas of the protein are much more abundant than others. Rather than the single preferred cleavage site seen with CAD, proteins fragmented by HCD cleave more readily at certain regions of the protein. For example, experimental evidence suggests that acidic regions of a protein are more preferably fragmented by HCD than basic regions. HCD provides several adjoining fragments and allows for identification of the protein, but limits the degree of characterization that is able to achieve.

1.5.3 Electron Transfer Dissociation (ETD)

An alternative to CAD is Electron transfer dissociation (ETD), which was developed by the Hunt laboratory in 2004 [32]. The reaction mechanism for ETD is shown in **Figure 1.7**. First, an electron is provided by a gas phase radical reagent anion, typically a cyclic aromatic compound. The two most common reagents are azulene or

fluoranthene, but several alternative compounds have been reported [33-35]. During

ETD, the reagent anion cloud and analyte cation cloud are allowed to interact in the

center of the ion trap. An electron is transferred from the reagent to an amide bond along

the peptide backbone. A carbonyl radical anion is generated which abstracts a proton

from a nearby positively charged basic residue. The resulting product contains a radical at

the carbonyl carbon which triggers bond rearrangement, eventually fragmenting the

amide bond to produce one c-ion and one z•-ion.

ETD is ideal for site-localizing labile PTMs, in that phosphorylation and O-

GlcNAcylation modifications are retained on peptides and proteins which have

undergone ETD fragmentation [36]. In theory, ETD provides random fragmentation

along the peptide backbone, with an equal chance that fragmentation will occur at each

N-Cα bond. However, this is not always the case. It is possible for a precursor ion to

undergo ETD but remain intact. In these cases, referred to as electron transfer with no

dissociation (ETnoD), the product will have a mass one Dalton less than the precursor.

Fragmentation N-terminal to proline, for example, will result in an intact precurosor due

to the N-Cα ring structure of proline. ETD efficiency is also highly dependent on charge

density. A protonated site must be nearby to provide a proton for the intermediate step of

the reaction mechanism – for this reason, ETD lends itself well to the fragmentation of

highly charged intact proteins [37]. However, some limitations do exist and will be

described in further detail in a later section.
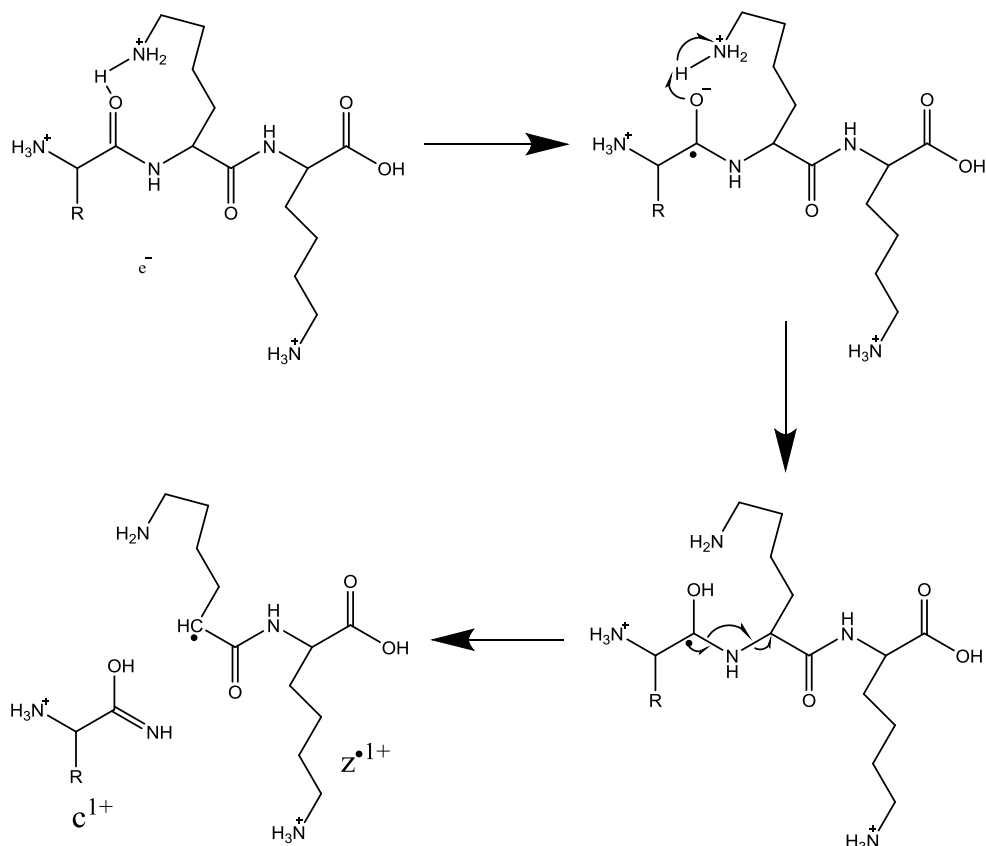
Figure 1.7: ETD reaction mechanism. Transfer of an electron to the carbonyl of the peptide backbone induces bond rearrangement that results in the formation of one c-ion and one z•-ion.

1.5.4 Ion-ion Proton Transfer (IIPT)

Ion-ion proton transfer (IIPT) has become an important tool in intact protein analysis. IIPT is analogous to ETD in that it is an ion-ion reaction between a multiply protonated cation and a singly charged reagent anion. However, the radical reagent anion abstracts a proton from the cation during IIPT rather than donating an electron [38-40]. Several studies have been carries out to explore the characteristics that would affect proton transfer versus electron transfer [35, 41]. Proton transfer from a peptide cation is significantly more exothermic than electron transfer from the reagent. Electron transfer therefore has a higher potential energy requirement and will occur at a longer distance as

the reagent anion orbits the cation. Reagents with a high electron affinity (>50 kcal/mol) will retain their electron past this point and result in proton transfer. However, it was found that several reagents with electron affinities below 50 kcal/mol performed proton transfer exclusively. In 2005, McLuckey *et al.* determined that the Franck-Condon overlap (the transition energy between vibrational states) of the anionic and neutral reagent ground electronic states could be used as a guide to determine reagents which would efficiently produce electron transfer products. In short, reagents with low electron affinities and low Franck-Condon factors would produce exclusively proton transfer products. In practice, these tend to be perfluoronated polycyclic hydrocarbons.

The product of IIPT differs from the reactant by a single charge (z) and the mass of a proton. For example, a 1,200 Da fragment with 3 charges (m/z = 400) that has undergone IIPT once is now 1,199 Da with 2 charges (m/z = 599.5). IIPT is commonly used in conjunction with ETD to charge-reduce fragment ions but can also be used to reduce the integer charge state multiply protonated precursor ions. This will be described in further detail in **Section 1.7**.

1.6 Ion-ion Reaction Kinetics

Ion-ion reactions (ETD and IIPT) occur when the ions involved come into a stable orbit with one another - more specifically, when the reagent anion comes into a stable orbit with the peptidic cation [42]. The probability that this will occur is given by the collisional cross-section (**Equation 1.7**):

$$(\mathbf{1.7}) \qquad \sigma = \frac{z_1{}^2 z_2{}^2 e^4}{\varepsilon_0 (\mu v^2)^2}$$

where z_1 and z_2 are the integer charge states of the two ions involved, e is the charge of

an electron, ε_0 is the vacuum permittivity, v is the relative velocity between the two

species, and μ is the reduced mass of the system. The energy contributed to the reaction

by the velocity and a geometric term convert the collisional cross-section to a reaction

rate constant given by **Equation 1.8**:

$$(\mathbf{1.8}) \qquad k = v\pi\sigma = v\pi \left(\frac{z_1{}^2 z_2{}^2 e^4}{\varepsilon_0 (\mu v^2)^2} \right)$$

The reaction rate is largely dependent on two factors. The rate is proportional to the

square of the precursor charge (assuming a single charged anionic reagent) and inversely

proportional to the velocity cubed. We will discuss the impact and implications of the

former in the next section. The latter is the basis for how we intend to control ion-ion

reactions. Slight increases in the relative velocity of either of the two species will

drastically decrease their reactivity with one another [43]. The following chapters will

explore the theory of parallel ion parking in more detail. We will also describe the

implementation and practice of parallel ion parking in the analysis of intact proteins.

1.7 Protein vs. Peptide Mass Spectrometry

The study of intact proteins offers a big picture view of unique proteoforms not

available with peptide-based techniques. The ability to differentiate protein species that

contain multiple post translational modifications or amino acid substitutions is essential

to understanding the role they play in biological function or disease. Further, analysis of

proteins as intact molecules allows us to differentiate naturally truncated proteins from experimentally introduced degradation products. Sequence determination of intact proteins grants access to information that might not be available with peptide-based methods. For instance, tryptic peptides that are very small may be impossible to detect by LC-MS methods. In addition, large peptides that carry few charges will not fragment efficiently. To date, the advantages offered by mass spectrometric analysis of intact proteins have been outweighed by limitations inherent to several of the methodologies discussed above. Recently, several advances to come out of the Hunt lab have aimed to address some of these limitations [44, 45].

First, the signal-to-noise ratio of fragments produced by ETD of intact proteins is much lower than that resulting from fragmentation of peptides. Two factors contribute to this: 1) the number of protein ions available for fragmentation is lower than in the case of peptides, and 2) the number of unique fragment ions that can possibly be produced by ETD of proteins is higher than with peptides. To ensure consistent ion-ion kinetics, modern instruments employ automatic gain control (AGC) which fills the LTQ with a designated number of charges [46]. For a typical precursor target of 1.0×10^5 charges, this equates to approximately 3,300 protein ions, each containing 30 charges. In comparison, the same target value equates to about 33,000 peptide ions with 3 charges each. In MS/MS experiments this problem is exacerbated by the number of fragments that are possible in each case. A protein the size of apomyoglobin (153 amino acid residues) can generate 304 possible c-type and z•-type fragment ions by ETD. Thus, for

an $[M+30H]^{30+}$ precursor, 3,300 precursor ions will produce about 11 of each fragment.

To compare, 33,000 precursor ions of a standard peptide, angiotensin (10 amino acids), at

the $[M+3H]^{3+}$ charge state will generate about 1,800 of each possible unique fragment.

Recently, the Hunt lab reported the implementation of a front-end ETD (FETD)

reagent source [44]. By introducing reagent anions from the front of the instrument, the

C-trap is available to store the products from several sequential ETD reactions prior to

injection into the Orbitrap for mass analysis. Whereas summing transients returns an

increase in signal equivalent to the number of transients, multiple fills of the C-trap

results in a linear response relative to the number of fills performed. However, there are

limits to the number of product ions that can be efficiently trapped. FETD and multiple

fill technology have resulted in dramatic increases in the signal-to-noise ratio of fragment

ions and thus sequence ion coverage [6]. But fragment ions larger than 100 amino acids

are still rarely observed.

The second issue that leads to poor sequence coverage overlap is the complexity

of MS/MS spectra. As **Figure 1.8** illustrates, large fragments will likely have a mass-to-

charge ratio (m/z) similar to that of the intact precursor. Therefore, a majority of the

fragment ions are concentrated around the isolated m/z. It has become common practice

in the industry to address this problem by intentionally over-reacting fragments to spread

the product ion signal throughout the available m/z range. As outlined in **Figure 1.8**, this

leads to smaller fragments and an increase in chemical noise due to internal fragment

ions, which are not valuable for providing sequence information. Both of these lead to a reduced overlap in fragment ion coverage. An alternative method of charge reduction, IIPT, was first developed by the McLuckey group and offers the ability to simplify ETD spectra without the risk of further fragmentation [45, 47]. As discussed above in **Section 1.5.4**, IIPT is similar to ETD in that the reaction involves a singly charged radical reagent anion and a multiply protonated precursor cation. However, reagents used for IIPT, typically perfluoronated hydrocarbons, abstract a proton from the cation rather than donating an electron. IIPT can also be used to manipulate the integer charge state of intact precursor ions, but the products of uncontrolled charge-reduction inhabit a range of charge states. If the IIPT reaction is carried even further, the m/z of product ions will fall outside of the instrument's mass range, eventually becoming neutralized and undetectable.

The third problem to be addressed is the over-reaction of proteins by ETD. The rate of ion-ion reactions is dependent on the square of the charge of the ions involved. In other words, a protein with 30 charges will react one hundred times faster than a peptide with 3 charges. Large fragments will continue to fragment at a similarly high rate. Consider, for example, the same 30+ protein undergoes ETD near the C-terminus to produce a fragment that has 29 charges and contains one less amino acid. This fragment will react almost as quickly as the precursor ion ($30^2/29^2 = 1.07$). As **Figure 1.8** shows, large fragment ions undergo secondary electron transfers to give only one sequence-informative fragment ion and one non-sequence-informative internal fragment ion. As

discussed above, ETD of apomyoglobin can generate 304 c-type or z•-type fragments, but there are over 11,000 possible unique internal fragments. Thus, internal fragments produced by ETD are not useful in determining sequence and only serve to contribute to chemical noise. In order to extend sequence coverage of larger proteins we set out to preserve large fragments that have undergone only one electron transfer reaction. In 2005, McLuckey *et al.* proposed a technique for altering the reactivity of ETD products termed parallel ion parking [48]. Implemented on a modified Paul trap instrument, the group demonstrated the ability of parallel ion parking to preserve more than 90% of first-generation ETD products from several peptides between seven and twelve amino acids in length.

The final limitation of intact protein analysis that we aim to address is the low signal-to-noise ratio of proteins ionized by ESI. As discussed earlier in **Section 1.2** and shown in **Figure 1.2**, the signal of each protein is split amongst a wide range of charge states. In the example shown, the most abundant charge state contains only 12% of the total signal attributed to the protein. Modified species are often present at very low levels and are unable to be differentiated from the noise in high resolution MS spectra and are therefore not detected. Additionally, species are typically selected for fragmentation based on their abundance in the ESI spectrum. The same protein at different charge states is often selected multiple times for fragmentation while low abundance species are not selected at all.

High MW Precursor Ion

| N - Terminus | 21 kDa (30+) = 700 m/z | C - Terminus |

First-Generation ETD Products

| 18 kDa (26+) = 692 m/z |

Second-Generation ETD Products

| 15 kDa (21+) = 714 m/z |

Third-Generation ETD Products

Legend:

= Intact Protein Ion
= C-Type Ion
= Z-Type Ion
= Internal Fragment
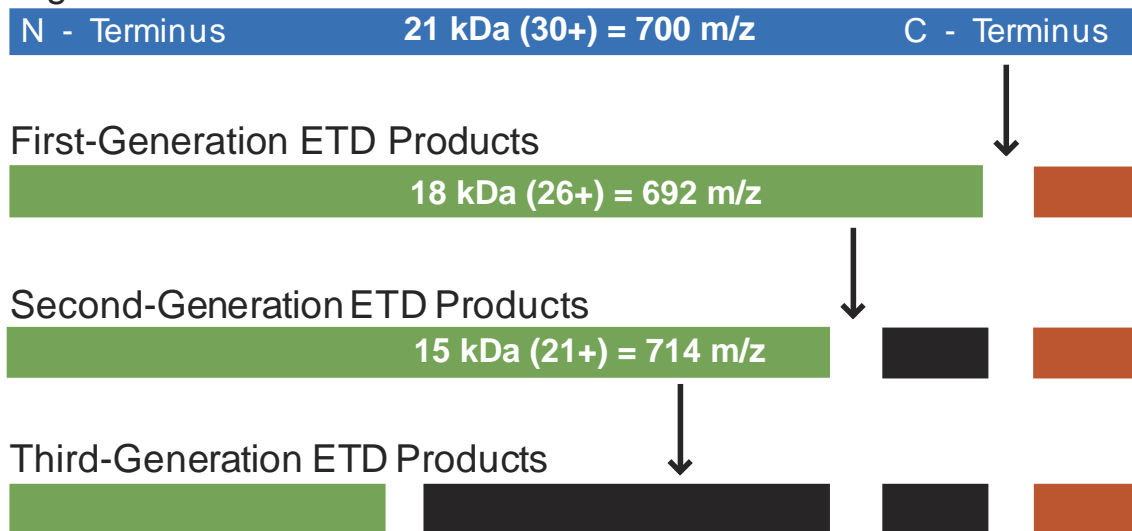
Figure 1.8: Cartoon depiction of products from several generations of ETD. When allowed to continure to react with fluoranthene, large products of ETD will continue to react. The resulting fragments are smaller sequence-informative fragments or internal fragments that do not provide sequence information. Sequence-informative fragments are at m/z's similar to that of the precursor ion and result in highly complex product ion spectra.

The remaining chapters in this dissertation will explore the implementation and application of parallel ion parking in the analysis of intact proteins. When used in conjunction with ETD, parallel ion parking can be used to preserve large fragment ions. We demonstrate increased sequence and fragment ion coverage on a chromatographic time scale, providing a more robust application of intact protein sequencing by ETD. When coupled with IIPT, parallel ion parking allows the concentration of an ion's signal into a single charge state. The result is a simplified spectrum in which each protein inhabits a single m/z peak. In addition, the signal-to-noise ratio is dramatically increased,

increasing the likelihood that less abundant species will be observed and allowing for

confident assignment of a protein's intact mass. We demonstrate the utility of this novel

method by identifying ribosomal proteins in a mixture on a chromatographic time-scale.

1.8 References

1. Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphrey-Smith I, Williams KL, Hochstrasser DF. From proteins to proteomics: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology*. **1996,** *14*, 61-65.

2. Dass C. *Principles and Practices of Biological Mass Spectrometry*. **2001,** John Wiley & Sons, Inc.

3. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR. Protein Sequencing by Tandem Mass Spectrometry. *Proc Natl Acad Sci USA*. **1986,** *83*, 6233-6237.

4. Aebersold R, Mann M. Mass spectrometry based proteomics. *Nature*. **2003,** *422*, 198-207.

5. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol*. **2003,** *21*, 255-261.

6. Anderson LC, Karch KR, Ugrin SA, Coradin M English AM, Sidoli S, Shabanowitz J, Garcia BA, Hunt DF. Analyses of histone proteoforms using front-end electron transfer dissociation-enabled Orbitrap instruments. *Mol Cell Proteomics*. **2016,** *15*, 975-988.

7. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*. **1988,** *60*, 2299-2301.

8. Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB. Electrospray interface for liquid chromatographs and mass spectrometry. *Anal Chem.* **1985,** *57*, 675-679.

9. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science*. **1989,** *246*, 64-71.

10. Fernandez de la Mora J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal Chim Acta*. **2000,** *406*, 93-104.

11. Smith RD, Loo JA, Ogorzalek RR, Busman M, Udseth HR. Principles and practices of electrospray ionization- mass spectrometry for large polypeptides and proteins. *Mass Spectrom. Rev*. **1991,** *10*, 359-452.

12. Carbeck JD, Severs JC, Gao J, Wu Q, Smith RD, Whitesides GM. Correlation between the charge of proteins in solution and in the gas phase investigated by protein charge ladders, capillary electrophoresis, and electrospray ionization mass spectrometry. *J Phys Chem*. **1998,** *102*, 10596-10601.

13. Nesatyy VJ, Suter MJF. On the conformation-dependent neutralization theory and charging of individual proteins and their non-covalent complexes in the gas phase. *J Mass Spec*. **2004,** *39*, 93-97.

14. Schwartz JC, Senko MW, Syka JE. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom*. **2002,** *13*, 659-669.

15. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG. The Orbitrap: a new mass spectrometer. *J Mass Spec*. **2005,** *40*, 430-443.

16. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, Horning S. Performance evaluation of a hybrid linear ion trap/Orbitrap mass spectrometer. *Anal Chem*. **2006,** *78*, 2113-2120.

17. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem*. **2000,** *72*, 1156-1162.

18. Eliuk S, Makarov A. Evolution of Orbitrap mass spectrometry instrumentation. Annu Rev *Anal Chem*. **2015,** *8,* 61-80.

19. Wong PS, Cooks GR. Ion trap mass spectrometry. *Curr Sep.* **1997,** *16,* 85-92.

20. Gerlich D. Inhomogenous RF fields- a versatile tool for the study of processes with slow ions. *Adv Chem Phys*. **1992,** *82*, 1-176.

21. March RE. An introduction to quadrupole ion trap mass spectrometry. *J Mass Spec*. **1997,** *32*, 351-369.

22. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. **2000,** *403*, 41-45.

23. Walsh CT. *Posttranslational Modification of Proteins: Expanding Nature's Inventory*. **2006.** Roberts & Company Publishers.

24. Deribe YL, Pawson T, Dikic I. Post-translational modifications in signal integration. *Nat Struct Mol Biol*. **2010,** *17*, 666-672.

25. Kaiser RE, Jr., Cooks RG, Syka JEP, Stafford GC, Jr.. Collisionally activated dissociation of peptides using a quadrupole ion trap mass spectrometer. *Rapid Commun*

Mass Spectrom. **1990,** *4*, 30-33.

26. Hunt DF, Buko AM, Ballard JM, Shabanowitz J, Giordani AB. Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. *Biomed Mass Spectrom.* **1981,** *8*, 397-408.

27. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal Chem*. **2004,** *76*, 3590-3598.

28. Keil B. *Specificity of Proteolysis*. **1992.** Springer-Verlag.

29. Martin, D.B.; Eng, J.K.; Nesvizhskii, A.I.; Gemmill, A.; Aebersold, R. Investigation of neutral loss during collision induced dissociation of peptide ions. *Anal. Chem.* **2005,** *77*(15), 4870-4882

30. Wysocki VH, Tsaprailis G, Smith LL, Breci LA. Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom.* **2000.** *35.* 1399-1406

31. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Meth*. **2007,** *4*, 709-712.

32. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA*. **2004,** *101*, 9528-9533.

33. Coon JJ, Shabanowitz J, Hunt DF, Syka JEP. Electron transfer dissociation of peptide anions. *J Am Soc Mass Spectrom*. **2005,** *16*, 880-882.

34. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Stafford G, Swaney DL, Syka JEP, Zabrouskov V, Coon JJ. A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-Orbitrap mass spectrometer. *J Proteome Res*. **2008,** *7*, 3127-3136.

35. Gunawardena HP, He M, Chrisman PA, Pitteri SJ, Hogan JM, Hodges BDM, McLuckey SA. Electron transfer versus proton transfer in gas-phase ion/ion reactions of polyprotonated peptides. *J Am Chem Soc*. **2005,** *127*, 12627-12639.

36. Garcia BA, Shabanowitz J, Hunt DF. Characterization of Histones and their post-translational modifications by mass spectrometry. *Curr Opin Chem Biol*. **2007,** *11*, 66-73.

37. Fornelli L, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, Tsybin YO. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol Cell Proteomics.* **2012,** *11,* 1758-1767.

38. Stephenson JL, Jr., McLuckey SA. Ion/ion proton transfer reactions for protein mixture analysis. *Anal Chem*. **1996,** *68*, 4026-4032.

39. Stephenson JL, McLuckey SA. Ion/ion reactions for oligopeptide mixture analysis: application to mixtures comprised of 0.5-100 kDa components. *J Am Soc Mass Spectrom.* **1998,** *9,* 585-596.

40. Scalf M, Westphall MS, Smith LM. Charge reduction electrospray mass spectrometry. *Anal Chem*. **2000,** *72*, 52-60.

41. Coon JJ, Syka JEP, Schwartz JC, Shabanowitz J, Hunt DF. Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions. *Int J Mass*

Spectrom. **2004,** *236,* 33-42.

42. McLuckey SA, Stephenson JL, Jr., Asano KG. Ion/ion proton-transfer kinetics: implications for analysis of ions derived from electrospray of protein mixtures. *Anal Chem.* **1998,** *70,* 1198-1202.

43. McLuckey SA, Reid GE, Wells JM. Ion parking during ion/ion reactions in electrodynamic ion traps. *Anal Chem.* **2002,** *74*, 336-346.

44. Earley L, Anderson LC, Bai DL, Mullen C, Syka JEP, English AM, Dunyach JJ, Stafford GC, Jr., Shabanowitz J, Hunt DF, Compton PD. Front-end electron transfer dissociation: a new ionization source. *Anal Chem.* **2013,** *85*, 8385-8390.

45. Coon JJ, Ueberheide B, Syka JEP, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci.* **2005,** *102*, 9463-9468.

46. Schwartz J, Kovtoun VV. Automatic gain control (AGC) method for an ion trap and a temporarily non-uniform ion beam. US 7960690 B2. June 14, 2011.

47. Stephenson JL, Jr., McLuckey SA. Simplification of product ion spectra derived from multiply charged parent ions via ion/ion chemistry. *Anal Chem.* **1998,** *70,* 3533-3544.

48. Chrisman PA, Pitteri SJ, McLuckey SA. Parallel ion parking: improving conversion of parents to first-generation products in electron transfer dissociation. *Anal Chem.* **2005,** *77*, 3411-3414.

Chapter 2. Parallel Ion Parking During ETD to Extend Fragment Ion Coverage

2.1 Abstract

Most protein studies using mass spectrometry involve the digestion of proteins into smaller peptides to simplify data acquisition and analysis. Characterizing proteins as intact molecules offers several advantages over this technique. However, low signal-to-noise ratio of product ions, highly complex fragment ion spectra and uncontrolled fragmentation are inherent to electron transfer dissociation (ETD) of intact species and have limited previous efforts in sequence interrogation. Previously, we have described a technique that increases the signal-to-noise ratio of MS/MS spectra using multiple product ion fills via implementation of a front-end ETD (FETD) source. We have also demonstrated a method of simplifying ETD spectra using ion-ion proton transfer (IIPT) to disperse product ions over the entire m/z range. First reported by the McLuckey group at Purdue University, parallel ion parking is a method of resonantly exciting specific ions during gas phase ion-ion reactions to dramatically slow their rate of reaction. Here we report the implementation of parallel ion parking during ETD of intact proteins to preserve large fragment ions and extend fragment ion coverage of intact proteins.

When combined with previously reported techniques (multiple product ion fills and ion-ion proton transfer) parallel ion parking significantly increased the number of fragment ions observed. When characterizing the standard protein apomyoglobin, large fragment ions resulting from a single electron transfer event were preserved as evidenced by fragments which had lost a single amino acid by ETD. The identity of 147 out of 153

amino acid residues (96.1%) was confirmed by the observation of 132 of the 148 possible

c-ions and 122 of the 148 possible z•-ions (254 of 296 total; 85.8%). This represents a

dramatic improvement over experiments in which parking was not employed, which

yielded only 59.5% fragment ion coverage.

To demonstrate the effectiveness of our method, we applied it to a more complex

and biologically relevant sample. Specifically, we used parallel ion parking, in

combination with charge-enhancing derivatization techniques, to characterize the light

chain (LC) of the monoclonal antibody (mAb) immunoglobulin gamma (IgG). Using

data acquired from three LC-MS experiments, we confirmed the identity of 163 out of

219 amino acid residues (74.4%). This was confirmed by the observation of 113 of the

207 possible c-ions and 77 of the 207 possible z•-ions (190 of 414 total; 44.9%).

2.2 Introduction

One of the major goals of protein analysis by mass spectrometry is the complete

characterization of a protein's primary structure, or amino acid sequence. Currently, a

majority of this analysis is performed at the peptide level [1, 2]. Proteins are typically

subjected to enzymatic digestion, which breaks them into smaller pieces prior to mass

spectrometric analysis. The component peptides are then characterized, allowing for the

determination of the protein sequence. However, as **Figure 2.1** illustrates, proteins that

contain multiple post-translational modifications, amino acid substitutions, or natural

degradation products are not easily differentiated from one another when the protein is

digested prior to analysis [3]. Protein species that are encoded by the same gene but differ

by an amino acid substitution or post-translational modification (PTM) are referred to as

proteoforms, and can differ vastly in their cellular function [4]. Thus, the ability to

identify and site-localize all modifications is crucial in order to fully understand the role

each proteoform plays within the cell.
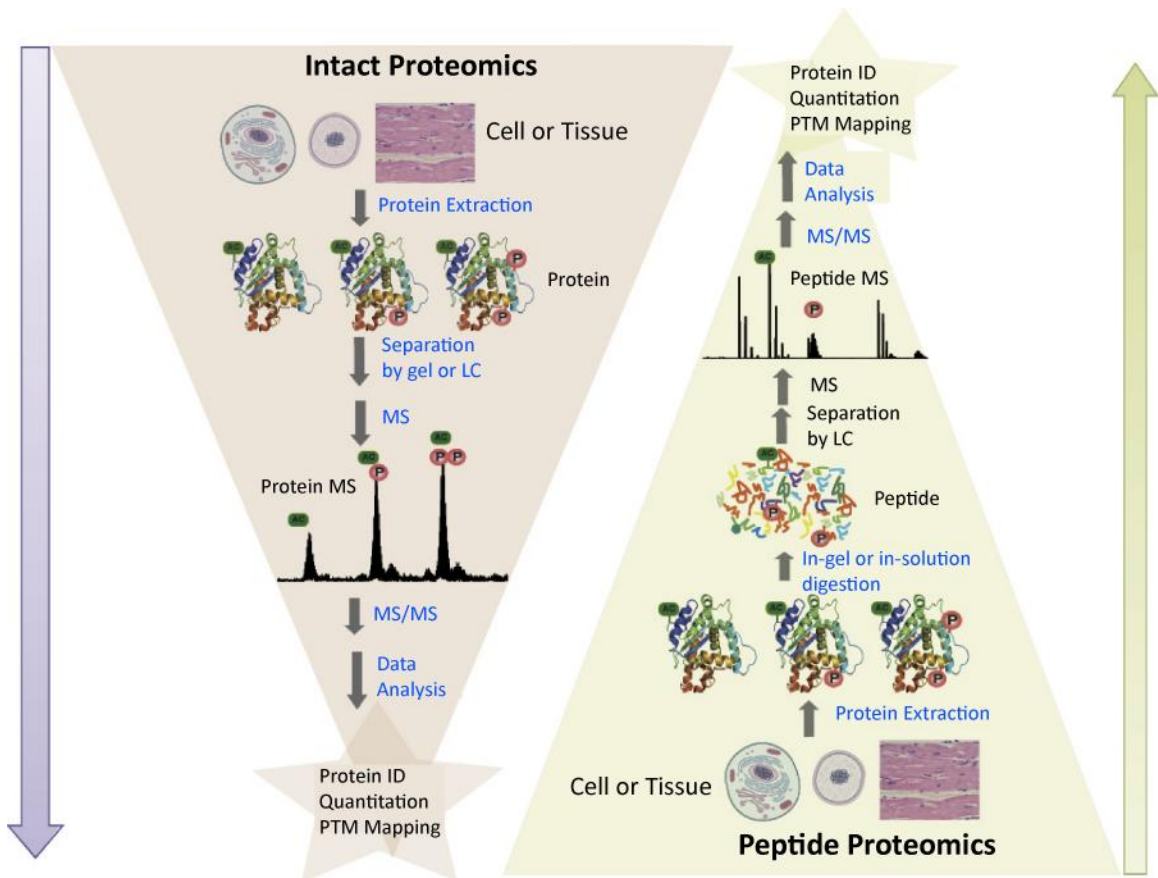
Figure 2.1. General workflow of intact versus peptide proteomics. Proteins to be analyzed by peptide proteomic methods are enzymatically digested into smaller pieces prior to MS analysis. This method allows for the site-localization of PTMs, yet makes the relationship between the presence of multiple PTMs on single proteins difficult to determine. Proteins analyzed by intact proteomic techniques are kept in their intact form, allowing for each modified species to be observed by MS analysis. Modified from [4].

Structural interrogation of proteins as intact molecules offers the ability to

identify and characterize each unique proteoform. Traditionally, 100% sequence

coverage by MS/MS is defined as the ability to assign the identity of an amino acid via the observation of either an N-terminal (a, b, or c) or C-terminal (y or z•) containing fragment ion [5, 6]. However, full confidence in the primary structure requires the observation of all possible fragmentation products (i.e. all c-ions and all z•-ions). The ability to read the sequence of a protein both forward and backward offers greater confidence in the determination of modification sites or amino acid substitutions. In addition, sequence information that may be unavailable (i.e. small c-terminal fragments that contain no basic residues) can be accessed by the observation of larger complimentary fragments.

In **Chapter 1** we discussed several aspects of mass spectrometry instrumentation that present difficulties in the sequence analysis of intact proteins. Efforts in attaining sequence ion coverage of intact proteins by ETD to date have demonstrated limited sequence overlap at best [7, 8]. Recently, we reported the most comprehensive sequence coverage of intact apomyoglobin (17 kDa) to date – 94% sequence coverage via derivatization techniques [9]. However, only 199 of 296 (67%) possible fragment ions were observed. Additionally, a majority of large fragment ions (>100 amino acids in length) were absent from the ETD spectra. This finding was not unexpected given the various aspects of intact protein sequencing that make it difficult to detect large fragments. Several of these are inherent to intact protein sequencing in general, while others are specific to the mechanism of ETD. Previously, our lab has addressed several of these difficulties in an attempt to improve sequence coverage. We have enabled multiple

product ion fills of the C-trap prior to mass analysis in the Orbitrap, granting dramatic gains in the signal-to-noise ratio of fragment ion spectra [10]. Additionally, we have utilized IIPT reactions subsequent to ETD fragmentation to spread the product ion signal throughout the available mass range [9, 11]. In this chapter, we will discuss a proposed solution for retaining large fragments resulting from a single ETD reaction.

Recall that the rate of ETD is proportional to the square of the precursor's charge. Because of this, highly charged products of ETD will continue to react. This is the primary reason that fragment ion coverage by ETD is limited. In order to observe large fragments and extend sequence coverage, it is necessary to stop large product ions from reacting. In 2006, McLuckey *et al.* proposed a technique for altering the reactivity of ETD products termed parallel ion parking [12]. Implemented on a modified Paul trap instrument, the group demonstrated the ability of parallel ion parking to preserve more than 90% of first-generation ETD products from several peptides between seven and twelve amino acids in length. Here we describe the implementation of parallel ion parking on an FETD-enabled Orbitrap Elite™ to preserve large fragment ions that have undergone a single ETD event. The aim of this work is to utilize parallel ion parking, in conjunction with multiple product ion fills and IIPT to address the problems discussed above. We ultimately show that very large ETD product ions can be preserved during fragmentation to extend fragment ion coverage of various intact proteins.

2.3 Parallel Ion Parking Theory

As discussed in **Chapter 1**, ion motion in RF ion trapping devices is approximated by the Mathieu equation, such that for a given combination of trapping RF field frequency and magnitude, ions inherit secular frequencies of motion that are primarily a function of m/z [13, 14]. LTQ devices enable the application of an auxiliary AC voltage in addition to the trapping field. An auxiliary voltage with a frequency which matches that of a trapped ion will resonantly excite (activate) that ion. Resonant activation is regularly employed in typical tandem MS functions such as collisionally activated dissociation (fragmentation) and mass analysis (ejection) [15, 16]. Activation energies below those necessary to induce fragmentation will increase the velocity at which the ion oscillates but does not contribute the kinetic energy necessary for fragmentation of the peptide bond. Recall that the reaction rate is largely dependent on two parameters: 1) the analyte charge and 2) the relative velocity between the analyte and reagent ions. In fact, the rate is inversely proportional to the cube of the relative velocity as is seen in **Equation 1**:

$$(\mathbf{1}) \quad k \; = \; v\pi \left(\frac{z_1 z_2 e^2}{4\pi\varepsilon_0 \mu v^2} \right)^2$$

Where v is the relative velocity between the analyte and reagent, $z\bullet_1$ and $z\bullet_2$ are the integer charge states of the two species involved, and µ is the reduced mass of the two reactants [17]. Even small increases in the velocity of one species will dramatically reduce its reaction rate [18].

Ion parking was reported as a means to slow the rate of reaction of specific ions trapped in an RF field while leaving other ions unaffected [19]. The desired ion is activated, based on its m/z, by applying a supplemental RF frequency that matches the ion's oscillation frequency. The specified ion is thus resonantly excited and it is effectively "parked" in its current state. While the reaction of an ion can-not be completely stopped, it can be slowed dramatically. This is illustrated in **Figure 2.2** which shows the IIPT reaction rate of the $[M+13H]^{13+}$ charge state of ubiquitin (7.6 kDa) as it relates to the activation amplitude of the IIPT reagent anion (perfluoromethyldecalin; m/z 512). With only a small amount of activation the reaction rate is reduced by almost an order of magnitude. The ability to arrest the kinetics of specific ions is especially beneficial to the analysis of intact proteins and can be utilized during both ETD and IIPT reactions.

"Parallel" ion parking is an extension of ion parking. Rather than targeting a single m/z using a single frequency, parallel ion parking involves activating a broad m/z range by applying several RF frequencies concurrently [12]. The signal of the frequencies which are activated is represented by a waveform in the frequency domain. Waveforms are typically used to eject ions that inhabit a wide m/z range in the linear ion trap of commercial mass spectrometers. However, the amplitude of the ejection waveform is much higher than those used for parking.
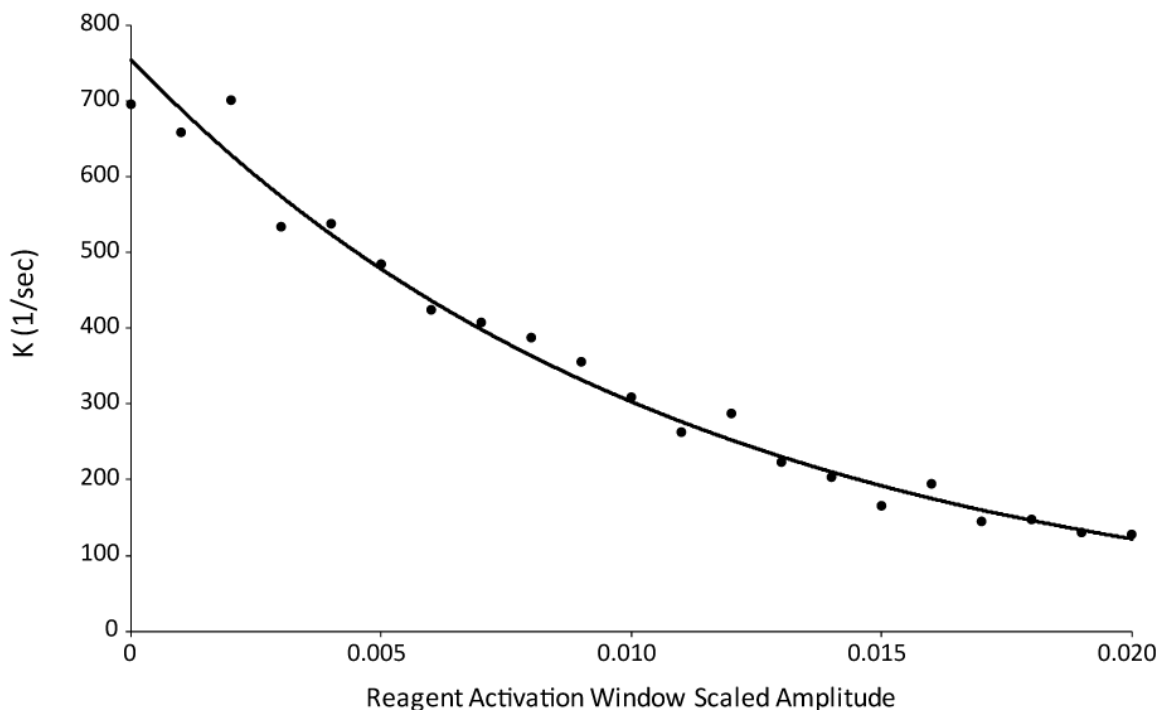
Figure 2.2. Effect of reagent activation amplitude on experimental ion-ion reaction rate constant. Experimentally determined rate constant of the IIPT reaction of the $[M+13H]^{13+}$ charge state of Ubiquitin (m/z 660) with perfluoromethyldecalin reagent anions as a function of amplitude (arbitrary units) of the frequency components for kinetic excitation of the reagent. The Mathieu q of the reagent (m/z 512) was set to 0.55.

Parallel ion parking during ETD will preserve product ions that have undergone a single fragmentation event by activating ions within the entire m/z range. The frequency components corresponding to the reagent anion and the precursor cation are lower in amplitude (approximately 10%) to ensure the desired reaction occurs. Fragment ions that differ in m/z from the precursor ion will be resonantly activated by the waveform and become parked, unable to continue reacting. The ETD products are then charge-reduced using IIPT to reduce any overlap in fragment ion signal.

2.4 Materials and Instrumentation

Agilent Technologies (Palo Alto, CA)

1100 Series high performance liquid chromatograph

1100 Series vacuum degasser

POROSHELL 300SB-C18 resin (5-µm, 300-Å)

Branson (Danbury, CT)

Branson 1200 ultrasonic bath

Eppendorf (Hauppauge, NY)

5414R Benchtop centrifuge

Fisher Scientific (Hampton, NH)

Tris(2-carboxyethyl)phosphine hydrochloride (TCEP•HCl)

GTS Welco (Allentown, PA)

10 ppm SF_6 in nitrogen gas

Honeywell (Morristown, NJ)

Burdick and Jackson® Acetonitrile, LC-MS grade

Labconco Corporation (Kansas City, MO)

Centrivap centrifugal vacuum concentrator

Millipore (Billerica, MA)

Amicon® Ultra-0.5 10K centrifugal filter

Molex (Lisle, IL)

Polymicro Technologies™ polyimide coated fused silica capillary

Sizes: 360 µm o.d. x 25, 50, 75 & 150 µm i.d.

Oakwood Chemical (Estill, SC)

Perfluoromethyldecalin

Omega Engineering (Stamford, CT)

Temperature controllers

Parker Hannifin Corporation (Cleveland, OH)

Porter back pressure regulator

Porter mass flow controller

PQ Corporation (Valley Forge, PA)

Kasil – Potassium silicate solution

SGE Analytical Science (Melbourne, Australia)

PEEKsil tubing 1/16" o.d., 0.025 mm i.d.

Sigma Aldrich (St. Louis, MO)

2-propanol, LC-MS grade

Ammonium Acetate

Ammonium Hydroxide

Apomyoglobin from equine skeletal muscle,

protein sequencing standard, lyophilized powder

Fluoranthene

Glacial acetic acid

N-(2-aminoethyl)maleimide trifluoroacetate salt

Trichloroacetic acid

Ubiquitin from bovine erythrocytes

Sutter Instrument Co. (Novato, CA)

P-2000 microcapillary laser puller

Swagelok (Solon, OH)

316 stainless steel plug for 1/8" Swagelok® tube fitting

Tektronix (Beaverton, OR)

MDO3014 Mixed domain oscilloscope

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

Calibration mixture

Formic acid, LC-MS grade

Orbitrap Elite™ mass spectrometer (custom modified with front-end ETD)

Upgraded from Velos Pro™ July, 2014

Pierce® water, LC-MS grade

Urea

Waters Corporation (Milford, MA)

Murine immunoglobulin gamma 1 (IgG1) monoclonal antibody (mAb)

Zeus Industrial Products (Orangeburg, SC)

Teflon tubing, 0.012" i.d. x 0.060" o.d.

2.5 Methods

2.5.1 Instrument Modification and Operation

To enable parallel ion parking, custom instrument control (ITCL) software was written for implementation of ion parking on an FETD-enabled Orbitrap Elite™ capable

of performing multiple product ion fills of the C-trap [10]. Supplemental waveforms were

designed for application during the charge-sign independent trapping portion of IIPT gas

phase reactions. Shown in **Figure 2.3** is the parallel ion parking waveform in the

frequency domain. The waveform included all frequencies, spaced at 500 Hz increments,

corresponding to the full 4,000 m/z range of the instrument. For experiments in which the

parking waveform was not applied during ETD, the Mathieu q value of the reagent (m/z

202) was set to 0.40. For experiments that utilized the parking waveform, the Mathieu q

value of the reagent was set to 0.80. The reasons for this will be described in the results

and discussion section.
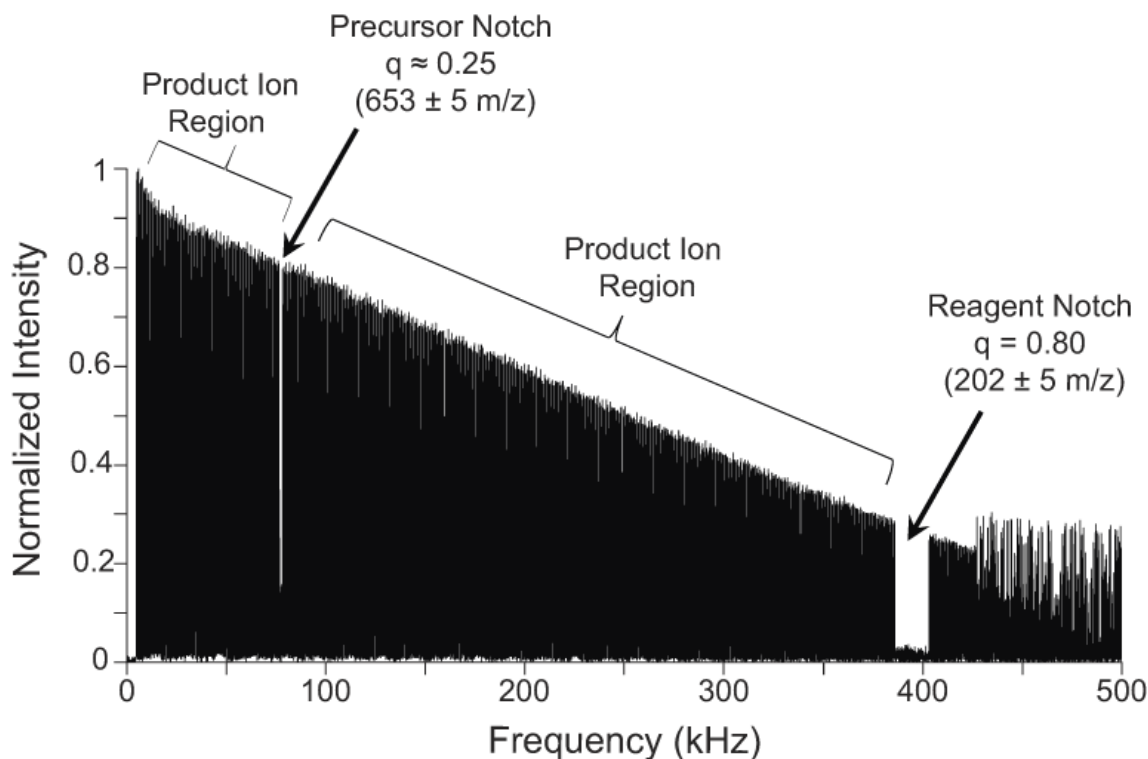
Figure 2.3. Parking waveform applied during charge-sign independent trapping portion of ETD.
The frequency domain waveform applied during ETD includes frequency components corresponding to
the entire m/z range. Product ions with an m/z that is resonantly excited by the waveform do not
fragment further. The precursor and reagent ions are slightly activated by a low amplitude voltage
within their respective frequency notches.

The amplitude of each frequency component that comprised the waveform was a function of frequency. The slope of the function was translated from the ejection amplitude calibration that correlates the optimum ejection amplitude to several m/z values. Ions with a high m/z (low frequency) require higher activation amplitudes relative to low m/z ions to feel the same degree of excitation. Frequency notches corresponding to the ETD reagent anion (fluoranthene; m/z 202 +/- 10) and a 10 m/z window centered on the precursor cation (m/z varied depending on experiment) were applied at significantly lower amplitudes (~10%) than the frequencies corresponding to product ions.

2.5.2 Direct Infusion of Ubiquitin for Reagent Activation Kinetics

Ubiquitin (bovine) was purchased from Sigma Aldrich (St. Louis, MO). One nanomole of lyophilized powder was reconstituted in a solution containing 40% acetonitrile and 60%, 0.1% acetic acid in water to a final concentration of 2.5 pmol/µL (all percentages expressed as v/v). The standard solution was directly infused into the instrument via a fused silica capillary emitter tip (360 µm o.d. x 75 µm i.d.) at a flow rate of approximately 100 nL/min and ionized by micro-electrospray ionization.

IIPT MS/MS scans were acquired in the ion trap mass analyzer under the following conditions: 300-4000 m/z scan range; 1E4 ion counts precursor automatic gain control (AGC) target; 3E5 reagent (perfluoromethyldecalin) AGC target; 1 scan/spectrum; 5 m/z precursor isolation window centered at 660 ($[M+13H]^{13+}$ charge state of ubiquitin); 0.03-100 ms IIPT. During IIPT, a parking waveform was applied that

included frequency components corresponding to a 5 m/z window centered on the reagent anion (502 m/z). The rate constant, k, of IIPT for the $[M+13H]^{13+}$ charge state of ubiquitin was experimentally determined as a function of the normalized amplitude of the frequency components that made up the parking waveform.

2.5.3 Direct Infusion of Apomyoglobin for ETD Sequence analysis

Apomyoglobin (equine) was purchased from Sigma-Aldrich (St. Louis, MO). One nanomole of lyophilized powder was reconstituted in a solution containing 40% acetonitrile and 60%, 0.1% acetic acid in water to a final concentration of 2.5 pmol/μL. Protein standards were directly infused into the instrument via a fused silica capillary emitter tip (360 μm o.d. x 75 μm i.d.) and ionized by micro-electrospray ionization.

MS/MS spectra were acquired using the following instrument parameters: resolution (R) = 120,000 at 400 m/z; 300–4000 m/z scan range; 1E5 ion counts precursor AGC target; 3E5 reagent (fluoranthene) AGC target; 1 scan/spectrum; 5 m/z precursor isolation window centered at 653 ($[M+26H]^{26+}$ charge state of apomyoglobin); 3-50 ms ETD followed by 10-30 ms IIPT. For experiments in which parking was employed, the parking waveform was applied during the charge-sign independent trapping portion of the ETD reaction. For several experiments, narrow windows were selected from the ETD products for charge reduction by IIPT. The windows were between 12 and 20 m/z in width and included all ETD products from m/z 580-860. Isolation windows were selected to ensure that unreacted precursor ions or products resulting from ETnoD were not

selected. To ensure all ETD products were selected, the isolation windows overlapped where appropriate (i.e. m/z 619-631 and m/z 629-641). **Figure 2.4a** shows a visual representation of all isolation windows with respect to the ETD product ion spectrum.

2.5.4 Reduction and Alkylation of IgG1

Murine immunoglobulin gamma 1 (IgG1) monoclonal antibody (mAb) was provided by Waters Corporation (Milford, MA). Disulfide bonds were reduced by dissolving 20 µg of the antibody in 10 µL tris(2-carboxyethyl)phosphine hydrochloride (TCEP•HCl) prepared in a 0.1% acetic acid solution containing 8 M urea (all percentages expressed as v/v). The sample was incubated at 50ºC for 10 minutes, after which the sample was neutralized to pH 6.8 by adding 2 µL of 0.2 M ammonium hydroxide. In some instances, cysteine residues were alkylated by adding 10 µL of 20 mM N-(2-aminoethyl)maleimide trifluoroacetate salt (NAEM) freshly prepared in 0.5 M ammonium acetate buffer containing 8 M urea (pH 6.8). The solution was then incubated at 50ºC for 10 minutes. The sample was immediately acidified by adding 1 µL of 25% formic acid. The alkylated mAb was diluted to 0.2 µg/µL by adding 77 µL of 0.1% acetic acid solution containing 8 M urea (final pH 3.9 with 8 M urea).

2.5.5 HPLC-MS/MS of IgG1

Approximately 1 pmol of total protein (1 µL) was pressure-loaded onto a precolumn (150 μm i.d. x 360 μm o.d. fused silica capillary packed with 14 cm POROSHELL 300SB-C18 5 μm diameter (Agilent)). The sample was desalted by

flowing solvent A over the sample-loaded precolumn for one hour (Solvent A: 0.3%

formic acid in water). After desalting, the precolumn was connected to an analytical

column (75 μm i.d. x 360 μm o.d. fused silica capillary packed with 22 cm POROSHELL

300SB-C18 5 μm diameter (Agilent)) which was equipped with a laser-pulled

nanoelectrospray emitter tip.

An Agilent (Palo Alto, CA) 1100 Series binary HPLC system was interfaced with

the Orbitrap Elite™ mass spectrometer for online protein separation and MS/MS

analysis. Proteins were eluted at 100 nL/min using the following gradient: 0-100%

solvent B in 60 minutes (Solvent B: 0.3% formic acid, 72% acetonitrile, 18%

isopropanol, 9.7% water).

MS scans were taken in the Orbitrap mass analyzer from 300-2,000 m/z (R =

120,000 at 400 m/z) and in the ion trap from 300-2,000 m/z (normal scan rate). MS/MS

scans were taken in the Orbitrap under the following conditions: r = 120,000 at 400 m/z;

200-4,000 m/z scan range; 1e5 precursor and 3e5 reagent AGC targets; 10 fragment ion

fills of the C-trap; 1 scan/spectrum; 5 m/z isolation window centered at m/z 757.25

($[M+32H]^{32+}$ charge state of IgG light chain) or m/z 780 ($[M+32H]^{32+}$ charge state of IgG

light chain with cysteines alkylated by NAEM); 5-20 ms ETD followed by 10 ms IIPT.

For experiments in which parking was employed, the parking waveform was applied

during the charge-sign independent trapping portion of the ETD reaction.

2.5.6 Data Analysis

High resolution full MS and MS/MS data were manually inspected using Qual Browser software (Thermo Fisher Scientific). Interpretation of all MS/MS data was performed manually on the unprocessed raw spectra. Theoretical fragment ion masses were calculated using in-house developed fragment mass calculator software. Theoretical isotopic distributions were modeled using Isotope Pattern Calculator (IPC) software (Pacific Northwest National Laboratories). Mass difference in parts per million (ppm) was calculated by comparing the observed mass of the most abundant isotope to the theoretical mass of the most abundant isotope predicted by IPC. Percent sequence coverage was calculated by dividing the number of observed N-Cα bond cleavages by the total number of possible N-Cα bond cleavages. Percent fragment ion coverage was calculated by dividing the number of observed fragment ions (c- or z•-type) to the total number of possible fragment ions. Note that cleavage of the N-Cα bond that is N-terminal to proline does not produce an observable fragment.

2.6 Results and Discussion

Three experiments were conducted to demonstrate the increase in sequence coverage offered by parallel ion parking. Experiment 1 was conducted using only ETD, IIPT, and multiple fill techniques. Parking was not employed for this experiment. For experiment 2, the parallel ion parking waveform was applied during the ETD portion of the ion-ion reaction. The entire population of ETD product ions was subsequently charge-reduced using IIPT to simplify data interpretation. Experiment 3 was conducted similarly

to Experiment 2 in that parking was performed during ETD. However, IIPT was performed on narrow m/z windows isolated from the ETD products. For all experiments, ten MS/MS spectra were averaged to provide the results presented. When assessing the effectiveness of experiments in which parking was employed during ETD, two types of coverage were calculated. "Sequence coverage" was defined as the observation of at least one fragment ion, c-type or z•-type, which confirmed the identity of the amino acid at that position. Confirmation of all 153 residues of apomyoglobin would be necessary to achieve 100% sequence coverage. Sequence coverage is the type most commonly calculated and reported in literature. However, confident assignment of amino acid identity requires the ability to read the sequence of the protein forward and backward. We therefore also determined the "fragment ion coverage" achieved by each experiment. When fragmentations N-terminal to proline are discounted, 100% fragment ion coverage of apomyoglobin required the observation of 148 c-ions and 148 z•-ions (296 ions total).

2.6.1 ETD & IIPT of Intact Apomyolgobin

Shown in **Figure 2.4A** is a 400 m/z range of the MS/MS spectrum resulting from Experiment 1. The $[M+26H]^{26+}$ ions (m/z 653) of the protein apomyoglobin (MW 16952.52 Da) were allowed to react with fluoranthene radical anions for 3 msec. ETD product ions were then allowed to react with sulfur hexafluoride ($SF_6^{-•}$) radical anions for 30 msec. The spectrum shown was the result of post-FT averaging of ten individual spectra acquired over the course of 66 seconds. Fragment ions observed in the spectrum are shown in **Figure 2.6** in red. Experiment 1 resulted in a relatively high degree of

sequence coverage of 88.9%. Sequence ions that were not observed were primarily fragments near the protein termini or near proline residues. Fragments containing one or two amino acids were not expected to be observed because their mass is lower than the lowest mass measured in the MS/MS spectra (m/z 300). Additionally, low mass $z\bullet$-ions were sparse due to the lack of a basic residue within the last six residues near the C-terminus. Two C-terminal fragments (z_2 and z_5) were observed at very low abundance, presumably the result of ETD events that occurred in the rare case in which a mobile proton resided within the last six residues. Further sequence coverage of this region required the observation of complimentary c-ions. Confirmation of proline residues was made difficult by the fact that ETD fragmentation N-terminal to a proline residue does not occur. Thus, the proline residue at that position can be inferred by the observation of fragment ions around that position, but both the c-ion and $z\bullet$-ion are necessary to unambiguously confirm the presence of a proline residue. While this experiment provided a high degree of sequence coverage, only 89 c-ions and 84 z-ions were observed for a total of 59.5% fragment ion coverage. The largest c-type and $z\bullet$-type fragments observed contained only 117 and 115 amino acids, respectively.
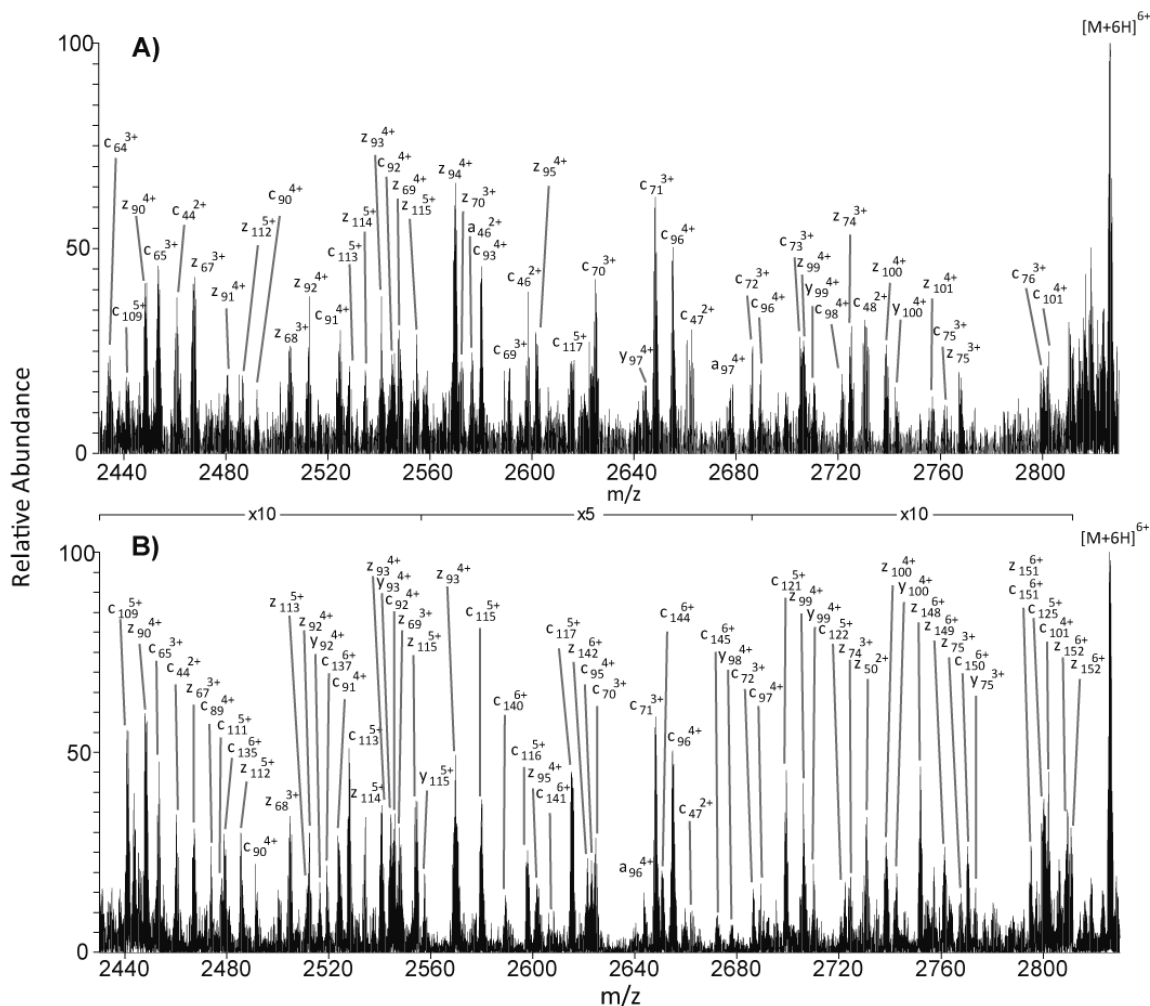
Figure 2.4. ETD/IIPT fragment ion spectra of intact apomyoglobin. (a) 3 ms ETD with no parking waveform applied followed by 30 ms IIPT of the $[M+26H]^{26+}$ charge state of apomyoglobin. (b) 50 ms ETD with the parking waveform applied followed by 30 ms IIPT.

2.6.2 ETD Parking & IIPT of Intact Apomyoglobin

Shown in **Figure 2.4B** is the MS/MS spectrum resulting from Experiment 2. The same 400 m/z range from **Figure 2.4A** is shown. Once again, the $[M+26H]^{26+}$ ions of apomyoglobin were reacted with fluoranthene radical anions for 50 msec during which a parallel ion parking waveform was applied. The ETD products were then allowed to react with $SF_6^{-\bullet}$ radical anions for 30 msec. The excess ETD reaction time was necessary

because the reagent and precursor ions are slightly activated by the parking waveform during the ETD portion of the experiment. The total acquisition time to acquire ten individual MS/MS spectra was 75 seconds. Fragment ions observed as a result of Experiment 2 are shown in blue in **Figure 2.6**. This experiment alone yielded 87.5% sequence coverage. Large fragment ions were observed in these spectra, including products resulting from the loss of a single amino acid from the N- and C-termini. In total, 120 unique c-ions and 98 unique z•-ions were observed, giving fragment ion coverage of 73.6%. Combined with the results from Experiment 1, 94.1% sequence coverage and 79.1% fragment ion coverage were acquired in less than two and a half minutes of total acquisition time. Observation of c_{148} through c_{152} confirmed the C-terminal sequence of the protein. This information was previously unavailable due to the lack of a basic residue within the first six amino acids. A significant gap in fragment ion coverage was observed between $z•_{116}$ and $z•_{145}$. It was presumed that these fragments were being produced due to the observation of complimentary c-ions within the same region. However, these c-ions were at a comparatively low abundance suggesting that fragmentation within this region is not preferred. This can be explained by the relatively low charge density in the region, with only two basic amino acids within the first 30 residues. For ETD to occur, a basic residue must be near the carbonyl that accepts the electron. We hypothesize that the fragment z•-ions in the region were present but were not being observed because they were at very low abundance and could not be differentiated from the noise.

2.6.3 IIPT of Parked ETD Product Slices

Shown in **Figure 2.5A** is the MS/MS spectrum resulting from experiment 3. After ETD was performed with the parking waveform applied small "slices" (10-20 m/z windows) were isolated from the ETD products and IIPT was conducted on the smaller ion population. This was done to eliminate product ions resulting from an electron transfer but no dissociation (ETnoD) that dominated the MS/MS spectra in Experiment 2. This increased the sensitivity of the product ion spectra by ensuring that only sequence-informative ions were being detected. **Figure 2.5B-D** show the fragment ion products resulting from 10 msec IIPT by $SF_6^{-\bullet}$ of several of the slices. Fragment ions observed as a result of Experiment 3 are shown in orange in **Figure 2.6**. This method alone gave limited sequence coverage of 79.1%. Fragment ion coverage was equally low with only 94 c-ions and 80 z•-ions observed (58.4%). However, this method uncovered fragment ions not previously observed in bulk MS/MS spectra. Several fragment ions within the N-terminal region were observed in Experiment 3 but a large gap in fragment ion coverage still remained. Work is ongoing to confirm the cause of this gap. The acquisition time required for this experiment was approximately two hours. When combined with the results from Experiment 1 and Experiment 2, a total of 97.4% sequence ion coverage was achieved. In total, 132 c-ions and 122 z•-ions were observed resulting in a combined fragment ion coverage of 85.8%. The combined results and results from each experiment are presented in **Figure 2.6**.
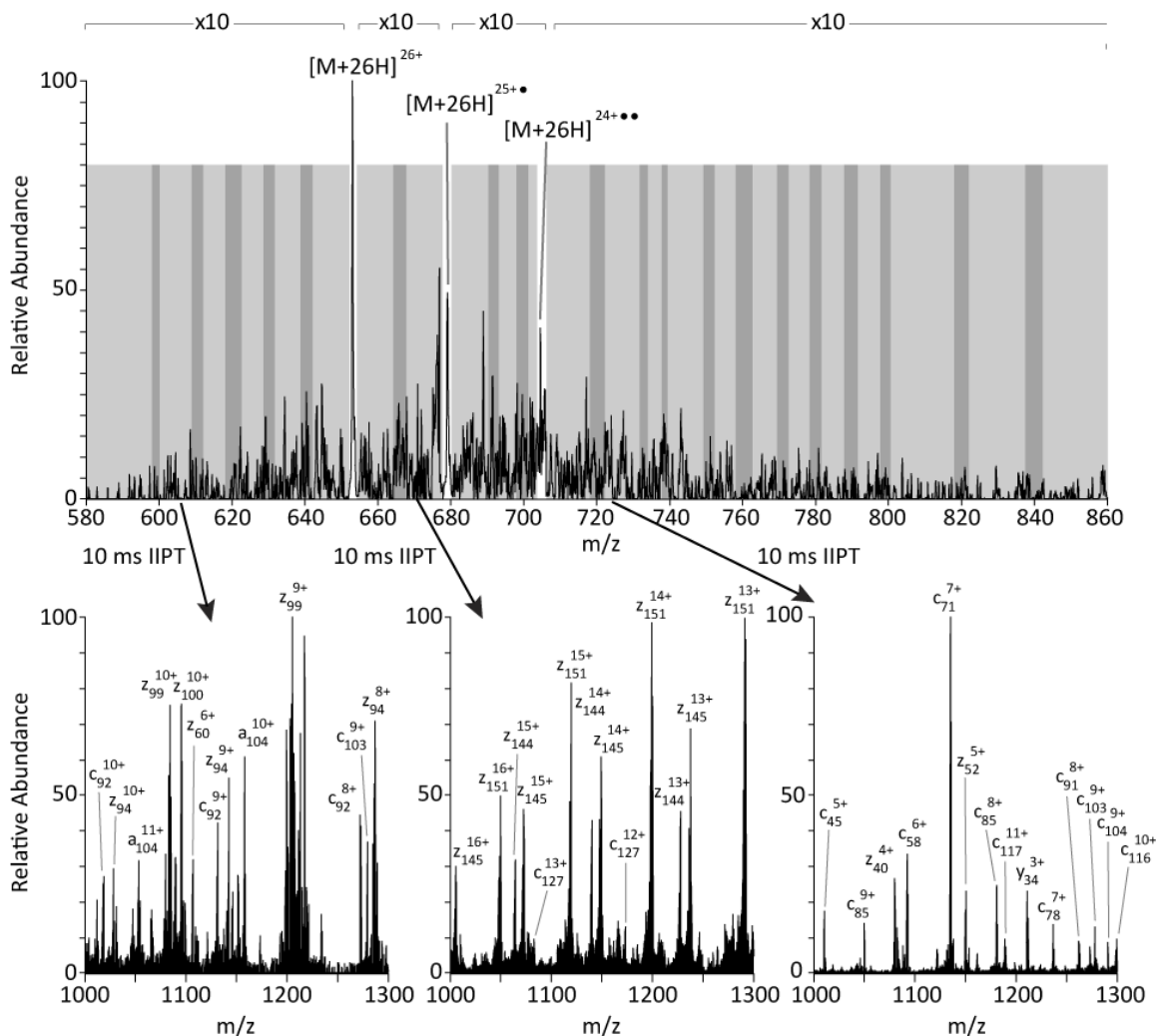
Figure 2.5. IIPT MS/MS spectra of narrow windows isolated from ETD products. (a) 3 ms ETD of the $[M+26H]^{26+}$ charge state of apomyoglobin with the parking waveform applied. (b) 10 ms IIPT performed on all ETD product ions between m/z 599 – 611. (c) 10 ms IIPT performed on all ETD product ions between m/z 664 – 678. (d) 10 ms IIPT performed on all ETD product ions between m/z 719 – 731.
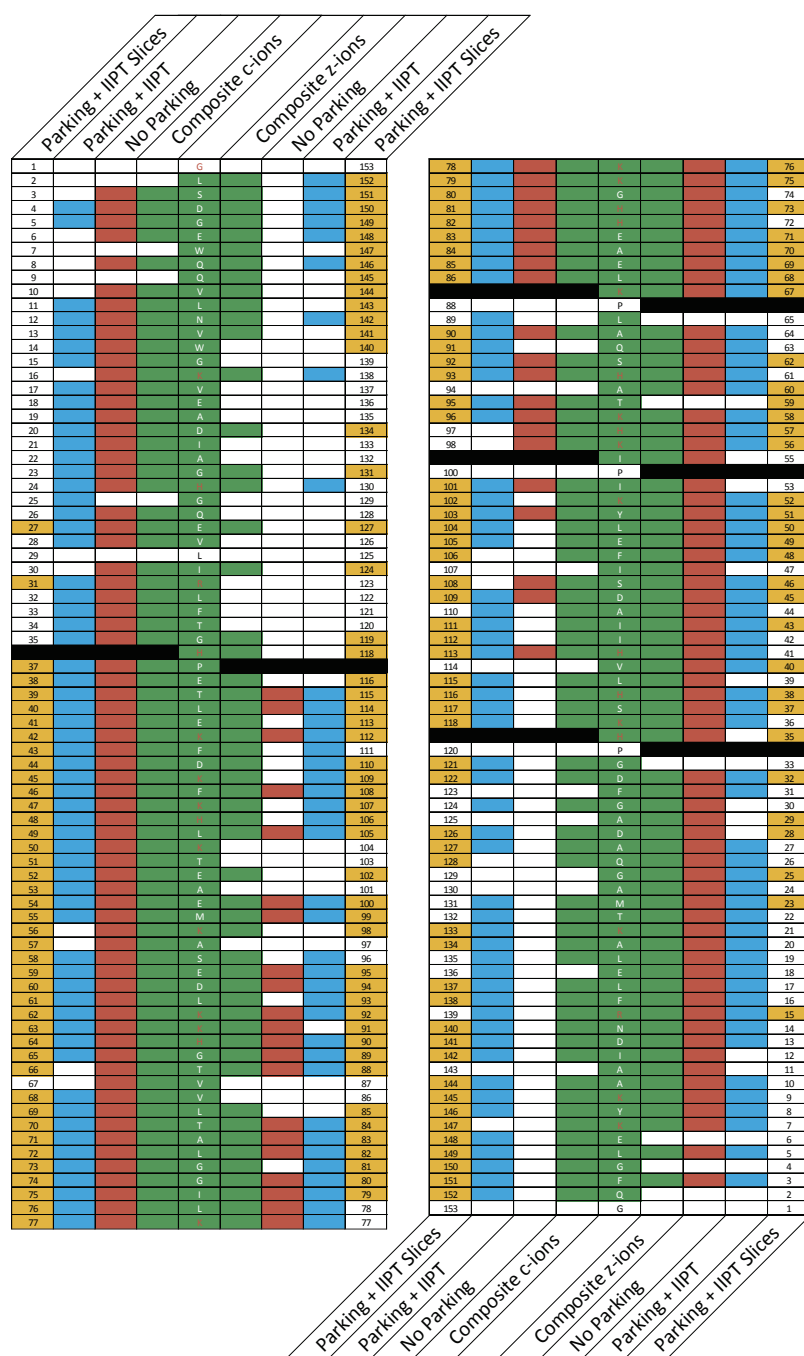
Figure 2.6. Combined sequence coverage map of apomyoglobin. Fragment ions observed as a result of ETD and IIPT only (red), ETD with parking followed by IIPT of all products (blue), and ETD with parking followed by IIPT on small windows isolated from the products (orange) are presented. Combined sequence and fragment ion coverage resulting from all experiments is shown in green. Amino acids with a basic side chain are denoted by red letters. A black bar indicates fragment ions that will not be observed due to proline. An expanded version of this figure can be found on the following pages.

#	Parking + IIPT Slices	Parking + IIPT	No Parking	Composite c-ions	Composite z-ions	No Parking	Parking + IIPT	Parking + IIPT Slices	#
1				G					153
2				L					152
3				S					151
4				D					150
5				G					149
6				E					148
7				W					147
8				Q					146
9				Q					145
10				V					144
11				L					143
12				N					142
13				V					141
14				W					140
15				G					139
16				K					138
17				V					137
18				E					136
19				A					135
20				D					134
21				I					133
22				A					132
23				G					131
24				H					130
25				G					129
26				Q					128
27				E					127
28				V					126
29				L					125
30				I					124
31				R					123
32				L					122
33				F					121
34				T					120
35				G					119
				H					118
37				P					116
38				E					116
39				T					115
40				L					114
41				E					113
42				K					112
43				F					111
44				D					110
45				K					109
46				F					108
47				K					107
48				H					106

#	Parking + IIPT Slices	Parking + IIPT	No Parking	Composite c-ions	Composite z-ions	No Parking	Parking + IIPT	Parking + IIPT Slices	#
49	blue	red	green	L	green	red	blue		105
50	blue	red	green	K	white	white	white		104
51	blue	red	green	T	white	white	white		103
52	blue	red	green	E	green	white	white		102
53	blue	red	green	A	white	white	white		101
54	blue	red	green	E	green	red	blue		100
55	blue	red	green	M	white	red	white		99
56	white	red	green	K	white	white	white		98
57	blue	red	green	A	white	white	white		97
58	blue	red	green	S	white	white	blue		96
59	blue	red	green	E	white	red	white		95
60	blue	red	green	D	white	white	white		94
61	blue	red	green	L	green	white	white		93
62	blue	red	green	K	white	red	white		92
63	blue	red	green	K	white	white	white		91
64	blue	red	green	H	green	red	blue		90
65	blue	red	green	G	green	red	blue		89
66	white	red	green	T	green	red	blue		88
67	white	white	green	V	white	white	white		87
68	blue	red	green	V	green	white	white		86
69	blue	red	green	L	green	white	blue		85
70	blue	red	green	T	green	red	blue		84
71	blue	red	green	A	green	red	white		83
72	blue	red	green	L	green	red	white		82
73	blue	red	green	G	green	white	white		81
74	blue	red	green	G	green	red	blue		80
75	blue	red	green	I	green	red	blue		79
76	blue	red	green	L	green	red	blue		78
77	blue	red	green	K	green	red	blue		77
78	blue	red	green	K	green	red	blue		76
79	blue	red	green	K	green	red	blue		75
80	blue	red	green	G	green	red	white		74
81	blue	red	green	H	green	red	blue		73
82	blue	red	green	H	green	red	blue		72
83	blue	red	green	E	green	red	blue		71
84	blue	red	green	A	green	red	blue		70
85	blue	red	green	E	green	red	blue		69
86	blue	red	green	L	green	red	blue		68
	black	black	black	K	green	red	blue		67
88	white	white	white	P	black	black	black	black	
89	blue	white	white	L	green	white	white		65
90	blue	red	green	A	green	red	blue		64
91	blue	white	white	Q	green	red	blue		63
92	blue	red	green	S	green	white	blue		62
93	blue	red	green	H	green	red	blue		61
94	white	white	white	A	green	red	blue		60
95	blue	red	green	T	white	white	white		59
96	blue	red	green	K	green	red	blue		58

	Parking + IIPT Slices	Parking + IIPT	No Parking	Composite c-ions		Composite z-ions	No Parking	Parking + IIPT	Parking + IIPT Slices	
97					H					57
98					K					56
					I					55
100					P					
101					I					53
102					K					52
103					Y					51
104					L					50
105					E					49
106					F					48
107					I					47
108					S					46
109					D					45
110					A					44
111					I					43
112					I					42
113					H					41
114					V					40
115					L					39
116					H					38
117					S					37
118					K					36
					H					35
120					P					
121					G					33
122					D					32
123					F					31
124					G					30
125					A					29
126					D					28
127					A					27
128					Q					26
129					G					25
130					A					24
131					M					23
132					T					22
133					K					21
134					A					20
135					L					19
136					E					18
137					L					17
138					F					16
139					R					15
140					N					14
141					D					13
142					I					12
143					A					11
144					A					10

	Parking + IIPT Slices	Parking + IIPT	No Parking	Composite c-ions	Composite z-ions	No Parking	Parking + IIPT	Parking + IIPT Slices	
145				K					9
146				Y					8
147				K					7
148				E					6
149				L					5
150				G					4
151				F					3
152				Q					2
153				G					1

2.6.4 IgG Background

Monoclonal antibodies (mAbs) make up one of the most rapidly growing classes of therapeutic pharmaceuticals [19]. As an important aspect of the adaptive immune system, mAbs are highly specific to their target antigen. Therapeutic mAbs are designed to identify and bind to a specific protein or peptide on the cell surface. Once an antibody (also known as immunoglobulin) has bound to its target, it blocks signaling pathways or recruits additional components of the immune system to attack the cell. The antigen target of an antibody is determined by the amino acid sequence of the variable region of the protein. Small changes in amino acid sequence can negatively affect the safety of the drug by influencing the immunogenic properties, therapeutic efficacy, or clearance rate [20, 21]. It is therefore important to determine the amino acid sequence of therapeutic mAbs. IgG1 is the most common type of antibody, making up about 50% of the antibody compliment in humans, and it will be the focus of the work presented here.
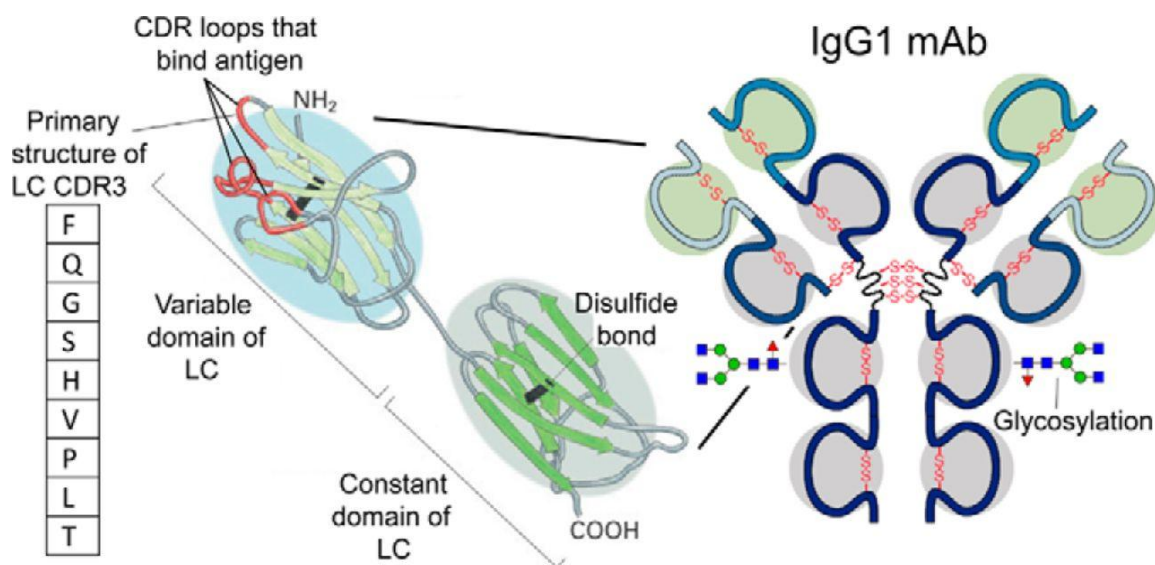
Figure 2.7. Cartoon depiction of the structure of immunoglobulin gamma 1 monoclonal antibody. IgG1 is a y-shaped tetramer made up of two identical heavy chains and two identical light chains. Proteins are bound to one another by Cys-Cys disulfide bonds. The variable domain of each chain is made up of the first approximately 100 amino acids. The sequence of the variable domain determines the antigen identified by a given antibody. This figure is modified from [28].

Antibodies exist in several classes (e.g. alpha, beta, gamma, etc.) that differ in the immune response they stimulate and functional location. Immunoglobulin gamma (IgG) is the class responsible for identifying and attacking invading pathogens such as bacteria and viruses. Each class in turn has several types (i.e. 1, 2, 3, or 4) that mostly differ in the sequence of the hinge region. **Figure 2.7** shows the structure of the IgG1 monoclonal antibody [22]. IgG1 is a 150 kDa tetrameric molecule made up of 4 separate protein chains: 2 identical 50 kDa heavy chains (HCs), and 2 identical 25 kDa light chains (LCs). The two HCs are connected to one another by three interchain cysteine-cysteine disulfide bonds in the middle of each chain [23]. Similarly, each LC is connected to the HC dimer through a single disulfide bond. The resulting tetramer is a y-shaped molecule with two identical halves. Each chain contains a constant domain, depicted by the dark blue

regions, and a variable domain, depicted by the light blue regions. The variable domain

of each chain is made up of the first approximately 100 amino acids. Within this domain

on each chain are three complimentarity determining regions (CDRs), each between 7

and 16 amino acids in length. The antigen identified by the antibody is determined by the

sequence of the three CDRs.

2.6.5 ETD & IIPT of IgG Light Chain

The antibody sample was incubated with tris(2-carboxyethyl)phosphine (TCEP)

prior to analysis by LC-MS/MS. As **Figure 2.8** illustrates, treatment of a protein by

TCEP reduces cysteine (Cys) disulfide bonds within the protein [24]. In IgG1, this

separates the sample into a mixture of light and heavy chains. The light chain and heavy

chain of IgG1 exhibit similar m/z ratios when ionized by ESI. Thus, the sample could not

be directly infused. Rather, the sample was separated by online high performance liquid

chromatography (HPLC) prior to MS analysis. **Figure 2.9A** shows the total ion current

(TIC) chromatogram from the HPLC-MS experiment. The trace of several of the most

abundant charge states of the light chain and heavy chain are shown in **Figure 2.9B** and

2.9C respectively. Shown in **Figure 2.9D** is the averaged ESI spectrum recorded as the

light chain eluted. Charge states of the light chain (●) carried between 16 and 34 positive

charges. In the high resolution Orbitrap spectrum, the most abundant charge state

contained 24 charges (most abundant isotope = m/z 1009.2493). The theoretical isotopic

distribution was modeled using Isotope Pattern Calculator software (Pacific Northwest

National Laboratory, http://omics.pnl.gov/). The software predicted the most abundant

isotope to contain 14 ^{13}C atoms with an m/z of 1009.2525. The theoretical and observed

mass of the intact light chain differ by only -3.2 ppm. **Figure 2.9E** shows the averaged

ESI spectrum recorded as the heavy chain elutes. Several of the charged species in this

spectrum were determined to be from the light chain. Additionally, the abundance of

heavy chain signal (♦) was lower than that of the light chain (●).
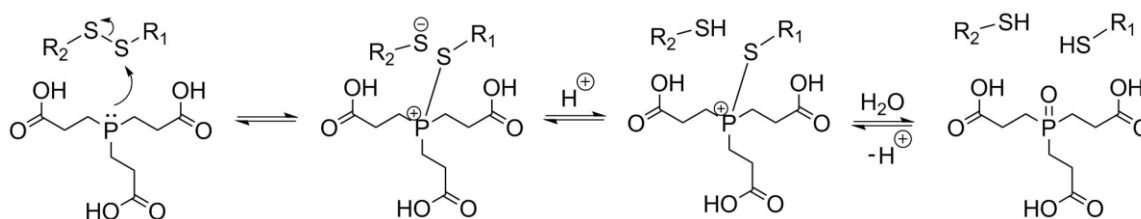
Figure 2.8. Mechanism of disulfide bond reduction by TCEP. Treatment of disulfide bonds by tris(2-carboxyethyl)phosphine (TCEP) results in reduction of the disulfide bond. Phosphine acts as the nucleophile forming a bond with one of the sulfides in what is the rate limiting step.

We determined that the $[M+32H]^{32+}$ (m/z 757.25) charge state was the highest

charged species of sufficient abundance to provide efficient fragmentation by ETD.

Figure 2.13 shows the combined sequence and fragment ion coverage of the light chain

of IgG1 under several experimental conditions. Coverage for ETD and IIPT alone is

shown in red. A basic residue (K, R, H) is denoted by a red letter. The MS/MS spectra

resulting from 5 ms ETD followed by 10 ms IIPT yielded only 58.4% sequence coverage

(128 of 219 amino acids). Only 68 out of 207 c-ions and 60 of a possible 207 z•-ions (128

of 414 fragment ions) were observed. This amounted to a total fragment ion coverage of

30.9%. Very little overlap in sequence coverage was observed. The largest fragment ion

observed contained 129 amino acids.
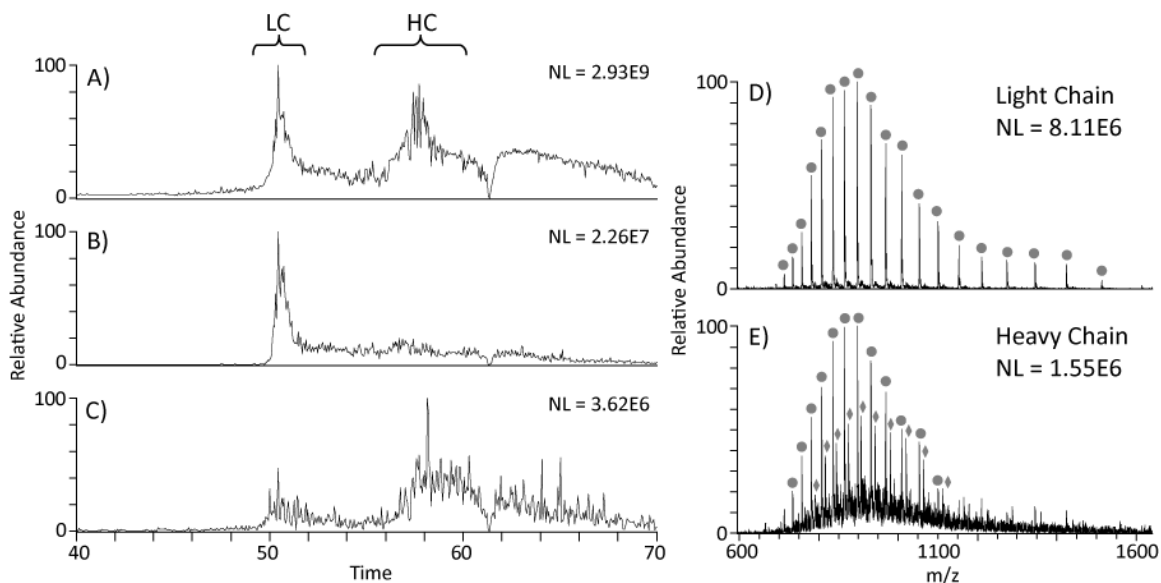
Figure 2.9. HPLC-MS chromatogram of reduced IgG1. The total ion chromatogram (A) shows effective separation of the light chain from the heavy chain. Extracted ion chromatograms of several of the most abundant charge states of the light chain (B) and the heavy chain (C). Averaged MS spectra from under the light chain (D) and heavy chain (E) chromatographic peaks.

It is important to note that the light chain of IgG only contains 24 basic amino acids. Including the N-terminus, this means that there are only 25 sites available for protonation. The precursor selected for ETD contained 32 charges. Thus, a number of protons were likely delocalized along the peptide backbone. As we discussed in **Chapter 1**, ETD efficiency is highly dependent on a nearby protonated basic side chain to provide a proton to the ETD reaction [25]. Typically, one basic residue for every seven amino acids is sufficient. The light chain of IgG1 contains several large gaps without a basic residue, including the first 23 amino acids. Additionally, we observed gaps in fragment ions observed near Cys residues (denoted by a * in **Figure 2.13**). Combined with the fact that a large amount of light chain was co-eluting with the heavy chain, this suggested that some portion of the disulfide bonds were remaining bound in solution. To mitigate this,

we treated the sample with an alkylating reagent subsequent to reducing the disulfide bonds.

2.6.6 Reduction, Alkylation, and Charge Enhancement by NAEM

To ensure that the light and heavy chains would remain separated in solution, the sample was alkylated after reduction of the disulfide bonds by TCEP. In typical proteomic mass spectrometry experiments, iodoacetamide is used as the alkylating reagent [26]. Recently however, we have introduced a new alkylating reagent, N-(2-aminoethyl)maleimide trifluoroacetate salt (NAEM). We have previously shown that alkylation by NAEM improves ETD of peptides containing Cys residues by adding a basic site to thiol groups [23]. As **Figure 2.10** illustrates, NAEM complexes with sulfhydryl compounds produced by reducing disulfide bonds. The primary amide on the NAEM group is readily protonated during ESI. In addition to fully separating the light chain from the heavy chain, alkylation by NAEM increased the total number of basic sites available for protonation.
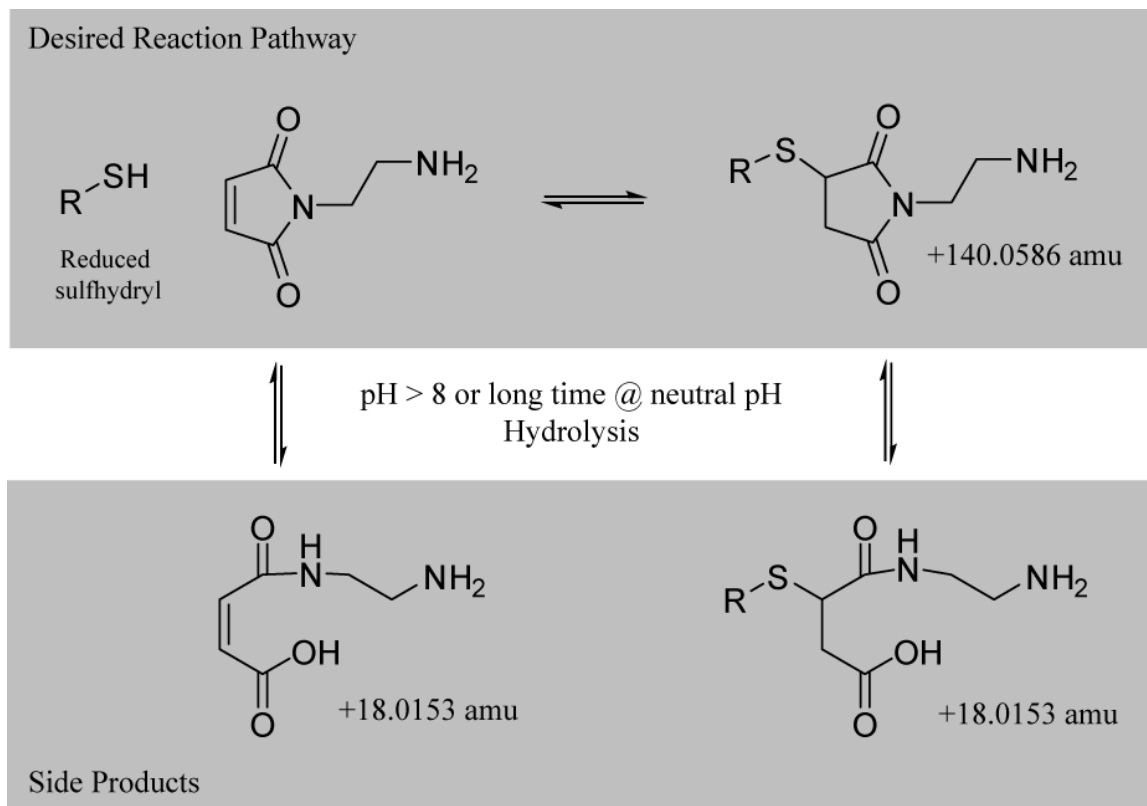
Figure 2.10. Mechanism of sulfhydryl alkylation by NAEM. Reduced sulfhydryls are derivatized when incubated with NAEM, resulting in the addition of 140.0586 amu to cysteine residues. Under certain conditions, the NAEM can be hydrolyzed, adding an additional 18.0153 amu.

Shown in **Figure 2.11A** is the TIC chromatogram of the NAEM treated IgG1. In this experiment, the light chain and heavy chain were fully separated from one another. **Figures 2.11B** and **2.11C** show traces of several of the most abundant charge states of the light chain and heavy chain respectively. Likewise, the averaged ESI MS spectra of the light chain and heavy chain peaks are shown in **Figure 2.11D** and **2.11E** respectively. In contrast to the experiment on the underivatized sample (**Figure 2.9**) no charge states from the light chain are evident in the heavy chain spectrum. This suggests that the light chain and heavy chain were indeed separated in solution.
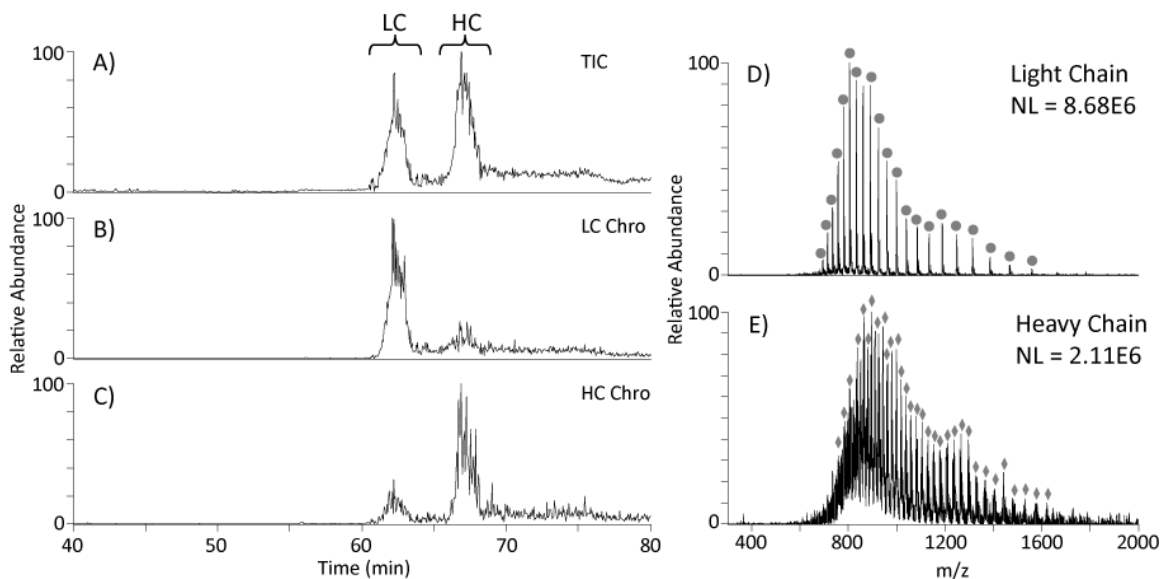
Figure 2.11. HPLC-MS chromatogram of NAEM treated IgG1. Total ion current chromatogram (A) of the chromatographic gradient is shown. Extracted ion chromatograms of several charge states of the light chain (B) and heavy chain (C) show efficient separation by NAEM treatment. Averaged MS spectra of the light chain (D) and heavy chain (E) chromatographic peaks show no carryover of the light chain into the heavy chain peak.

Figure 2.12 shows the most abundant charge state (27+) from the averaged high resolution spectrum of the light chain chromatographic peak. Four isotopic clusters are observed in this spectrum. The sequence of the light chain contains five Cys residues. Assuming all five Cys side chains are derivatized, the most abundant isotope at this charge state is expected to be at m/z 923.1992. The most abundant isotope of one observed isotopic cluster is at m/z 923.1922, a difference of -7.6 ppm. The remaining isotopic clusters each exhibit a mass shift of approximately 18.0 amu. NAEM can be hydrolyzed (+18.0153 amu) if the reaction is performed at a pH above 8 or is allowed to progress for a long period of time (approximately an hour). The hydrolyzed side products are shown in **Figure 2.10**. Indeed, almost every fragment ion containing a Cys residue exhibited a mass shift corresponding to 18.01 amu in addition to the expected 140.06

amu. Fragment ions containing multiple Cys residues exhibited multiple fragment ion

signals corresponding to the number of Cys present.
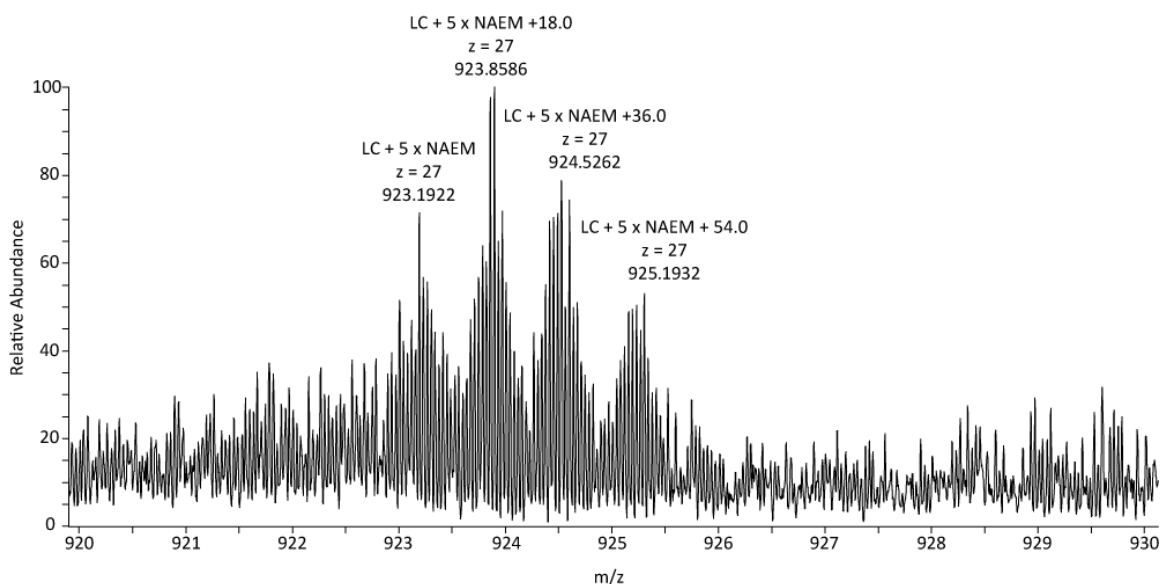
Figure 2.12. Isotopic clusters of [M+27H]$^{27+}$ charge state of NAEM treated light chain. A zoomed in view of a small section of the high resolution Orbitrap scan of intact light chain at the [M+27H]$^{27+}$ charge state. Each subsequent cluster results from the addition of approximately 18.01 amu.

Once again, the [M+32H]$^{32+}$ charge state (m/z 780.25) was selected for

fragmentation by 5 ms ETD and charge reduction by 10 ms IIPT. Fragment ions observed

in the MS/MS spectrum are shown in blue in **Figure 2.13**. The experiment resulted in

sequence ion coverage of 62.6% (137 of 219 amino acids). Fragment ion coverage of

33.6% (74 c-ions and 57 z•-ions) was achieved. Both metrics showed modest increases

from the previous experiment (58.4% and 30.9% respectively). The most notable result

from this experiment was the increase in fragment ion coverage near several Cys

residues. Without alkylation, we observed a gap in coverage around Cys93, from Gly89

to Ser97. ETD of the NAEM derivatized sample yielded two new fragment ions, Cys93

and Glu94. Likewise, z•-ion coverage from Tyr197 to Cys199 was observed in the

derivatized sample. Interestingly, a fragment ion containing 135 amino acids was observed during this experiment.

Although several newly observed fragments were realized by ETD of IgG1 treated with NAEM, only a slight increase in sequence and fragment ion coverage were achieved. Several fragment ions that were present in the untreated sample were not observed in the treated sample. Upon further investigation, the fragment ions that were no longer present were those that were in low abundance in the previous experiment. These ions were likely lost due to the incomplete hydrolysis of the NAEM derivative mentioned earlier. A single fragment ion was present at several different m/z values, further splitting the fragment ion signal to the point that the signal of any one peak was not above the noise level. In our experiments with apomyoglobin, large fragments such as those resulting from the loss of a single amino acid were present at very low S/N. Further reduction in S/N may result in large ions being unobserved and was therefore not a desirable trade off in exchange for the minimal increase in coverage gained by NAEM treatment. This, combined with the fact that sufficient separation of the light chain was achieved with the untreated sample, prompted us to perform ion parking experiments with an untreated IgG1 sample.

2.6.7 Parallel Ion Parking of IgG Light Chain

To demonstrate the online capabilities of parallel ion parking, the $[M+32H]^{32+}$ charge state of the untreated IgG1 light chain (m/z 757.25) was selected for

fragmentation. During the charge-sign independent trapping portion of the ETD reaction, a waveform was applied which included frequencies corresponding to the entire mass range except those corresponding to the precursor +/- 5 m/z and the reagent (fluoranthene; m/z 202) +/- 5 m/z. The reagent and precursor were slightly activated by the waveform, as evidenced by the increased ETD time necessary (20 ms versus 5 ms) in comparison to the experiment in which parking was not employed. Fragments with an m/z outside the precursor and reagent notches fell in resonance with the main frequencies of the waveform. These ions were excited by the waveform, causing their velocities to increase and dramatically slowing their ability to further react with the reagent anion. After ETD, IIPT was performed on the entire fragment ion population, similar to experiment 2 with apomyoglobin. The light chain eluted over the course of two minutes, during which ten scans were acquired, each with ten fills of the C-trap. The results presented were derived by averaging these ten spectra. Fragment ions observed by parallel ion parking during ETD are shown in orange in **Figure 2.13**. Sequence coverage as a result of this experiment was 61.6% (135 of 219 amino acids). The true benefit of this technique is demonstrated by the fragment ion coverage achieved due to the preservation of very large fragment ions. 68 z•-ions and 93 c-ions were observed (161 of 414 possible fragments; 38.9%). While previous experiments yielded fragment ions containing 135 and 97 c- and z•-ions respectively, the ion parking waveform successfully preserved fragment ions containing up to 204 amino acids. Taken together, over the course of three 60 minute HPLC gradients, sequence coverage of 74.4% was achieved through the observation of 190 of 414 possible fragment ions (45.9%).

Figure 2.13. Combined coverage map of IgG1 light chain. Fragment ions observed as a result of ETD/IIPT of reduced IgG1 light chain (red), ETD/IIPT of NAEM alkylated IgG1 light chain (blue), and ETD with parking followed by IIPT of the entire product ion population (orange) are presented. Combined sequence and fragment ion coverage resulting from all experiments is shown in green. Amino acids with a basic side chain are denoted by red letters. Cysteine residues are denoted by (*). A black bar indicates fragment ions that will not be seen due to proline. An expanded version of this figure can be found on the following pages.

#	ETD Parking + IIPT	NAEM Deriv. +ETD/IIPT	ETD + IIPT	Composite c-ions	Composite z•-ions	ETD + IIPT	NAEM Deriv. +ETD/IIPT	ETD Parking + IIPT	#
1				D					219
2				V					218
3				L					217
4				M					216
5				T					215
6				Q					214
7				T					213
8				P					212
9				L					211
10				S					210
11				L					209
12				P					208
13				V					207
14				S					206
15				L					205
16				G					204
17				D					203
18				Q					202
19				A					201
20				S					200
21				I					199
22				S					198
23				**C**					197
24				R					196
25				S					195
26				S					194
27				Q					193
28				Y					192
29				I					191
30				V					190
31				H					189
32				S					188
33				N					187
34				G					186
35				N					185
36				T					184
37				Y					183
38				L					182
39				E					181
40				W					180
41				Y					179
42				L					178
43				Q					177
44				K					176
45				P					175
46				G					174
47				Q					173
48				S					172

	ETD Parking + IIPT	NAEM Deriv. +ETD/IIPT	ETD + IIPT	Composite c-ions	Composite z•-ions	ETD + IIPT	NAEM Deriv. +ETD/IIPT	ETD Parking + IIPT	
49	■	■	■	P					171
50	■	■	■	K					170
51	■	■	■	L					169
52	■	■	■	L					168
53	■	■	■	I					167
54	■	■	■	Y					166
55		■	■	K					165
56				V					164
57	■	■	■	S					163
58	■	■	■	N					162
59	■	■	■	R					161
60	■	■	■	F					160
61	■	■	■	S					159
62	■	■	■	G					158
63	■	■	■	V					157
64	■	■	■	P					156
65	■	■	■	D					155
66	■	■	■	R					154
67		■	■	F					153
68	■	■	■	S					152
69	■	■	■	G					151
70	■	■	■	S					150
71	■	■	■	G					149
72	■	■	■	S					148
73		■	■	G					147
74	■	■	■	T					146
75	■	■	■	D					145
76		■	■	F					144
77	■	■	■	T					143
78		■	■	L					142
79				K					141
80		■	■	I					140
81	■	■	■	S					139
82	■	■	■	R					138
83	■	■	■	V					137
84		■	■	E					136
85		■	■	A					135
86	■	■	■	E					134
87	■	■	■	D					133
88	■	■	■	L					132
89	■		■	G					131
90				V					130
91				Y					129
92	■		■	Y					128
93	■		■	C					127
94	■		■	F					126
95				Q					125
96				G					124

	ETD Parking + IIPT	NAEM Deriv. +ETD/IIPT	ETD + IIPT	Composite c-ions	Composite z•-ions	ETD + IIPT	NAEM Deriv. +ETD/IIPT	ETD Parking + IIPT	
97				S					123
98	blue	red	green	H					122
99	black	black	black	V	green				121
100			green	P	black				120
101			green	L					119
102	blue	red	green	T					118
103	blue	red	green	F					117
104	blue	red	green	G					116
105	blue		green	A	green				115
106	blue	red	green	G					114
107				T					113
108	blue	red	green	K					112
109	blue		green	L					111
110	blue		green	E	green				110
111			green	I					109
112		red		K					108
113	blue	red	green	R					107
114	blue	red	green	A	green				106
115				D					105
116			green	A					104
117	black	black	black	A					103
118			green	P	black				102
119		red	green	T					101
120	blue		green	V	green				100
121	blue		green	S					99
122	blue	red	green	I					98
123	black	black	black	F	green	red		blue	97
124	black			P	black				96
125				P	black				95
126				S					94
127				S					93
128	blue		green	E					92
129	blue	red	green	Q					91
130	blue		green	L	green	red		blue	90
131	blue		green	T	green			blue	89
132			green	S	green			blue	88
133				G					87
134				G					86
135	blue		green	A					85
136				S					84
137				V					83
138				V					82
139				C					81
140				F					80
141			green	L	green			blue	79
142				N					78
143			green	N	green			blue	77
144			green	F	green	red		blue	76

	ETD Parking + IIPT	NAEM Deriv. +ETD/IIPT	ETD + IIPT	Composite c-ions	Composite z•-ions	ETD + IIPT	NAEM Deriv. +ETD/IIPT	ETD Parking + IIPT	
145	■	■	■	Y		blue			75
146				P	■	■	■	■	74
147			green	K		blue			73
148			green	D		blue			72
149			green	I					71
150			green	N	red	blue			70
151			green	V	red	blue			69
152				K	red	blue			68
153			green	W					67
154			green	K	red	blue			66
155			green	I	red	blue			65
156				D	red				64
157				G					63
158				S					62
159				E	red	blue			61
160			green	R	red	blue			60
161			green	Q	red	blue			59
162				N	red				58
163				G	red				57
164				V	red				56
165			green	L					55
166			green	N					54
167			green	S	red	blue			53
168			green	W	red	blue			52
169				T	red	blue			51
170			green	D	red	blue			50
171			green	Q	red	blue			49
172				D	red	blue			48
173				S	red	blue			47
174				K	red				46
175				D	red	blue			45
176			green	S	red	blue			44
177			green	T	red	blue			43
178			green	Y	red	blue			42
179			green	S	red	blue			41
180				M	red	blue			40
181				S	red	blue			39
182				S	red	blue			38
183				T	red	blue			37
184				L	red	blue			36
185				T	red	blue			35
186				L	red	blue			34
187				T	red	blue			33
188				K	red	blue			32
189				D	red	blue			31
190			green	E	red	blue			30
191				Y	red	blue			29
192				E	red	blue			28

ETD Parking + IIPT	NAEM Deriv. +ETD/IIPT	ETD + IIPT	Composite c-ions		Composite z•-ions	ETD + IIPT	NAEM Deriv. +ETD/IIPT	ETD Parking + IIPT
193				R				27
194				H				26
195				N				25
196				S				24
197				Y				23
198				T				22
199				C				21
200				E				20
201				A				19
202				T				18
203				H				17
204				K				16
205				T				15
206				S				14
207				T				13
208				S				12
209				P				11
210				I				10
211				V				9
212				K				8
213				S				7
214				F				6
215				N				5
216				R				4
217				N				3
218				E				2
219				C				1

2.7 Conclusions and Future Directions

The results presented here demonstrate that large fragment ions can indeed be preserved using parallel ion parking. ETD on apomyoglobin while parking preserved products that had only fragmented once as evidenced by c_{152} and $z•_{152}$. Each of these ions represented precursor ions that lost only a single amino acid. In fact, the nine largest c-ions and 13 largest z•-ions were observed, providing complete sequence coverage (backward and forward) of both termini. Further, we observed 85.8% of the possible

fragments produced by ETD of intact apomyoglobin. With no parking techniques employed, fragment ion coverage was only 59.5%. This represents the most comprehensive coverage of a protein the size of apomyoglobin to date. However, the technique used requires a highly purified sample as it was achieved using an experiment lasting more than two hours. The method is therefore not practical for LC-MS/MS analysis.

When applied to the light chain of a monoclonal antibody, our novel method resulted in the observation of fragments containing more than 200 amino acids. Experiments in which parking was not employed yielded fragment ions with up to 129 amino acids. We thus achieved our goal of increasing the size of fragment ions produced by ETD. Using parking techniques, the fragment ion coverage we were able to achieve was also increased to 45.9% from 30.9%. While these results show promise toward achieving complete fragment ion coverage and the ability to elicit the sequence of a protein forward and backward, several questions remain to be answered.

First, can the gap in large $z\bullet$-ion coverage seen in apomyoglobin be filled? Several observations lead us to believe that the ions in question should be observed. Ions are observed that are larger than those that are missing. Therefore, the problem is not due to the size of the fragment ions. Additionally, c-ions that are complimentary to the missing $z\bullet$-ions are observed. Thus, it can be concluded that fragmentation is indeed occurring at the relevant N-$C\alpha$ bonds. Finally, ions of a similar m/z and frequency to

those missing are observed. Therefore, the missing ions are not continuing to react due to inefficient ion parking. It is possible that the lack of a basic residue in this region is causing the lack of fragment ion coverage that we observe. Studies are ongoing to determine the reason for lack of coverage in this region.

Finally, will a more densely charged protein provide a higher degree of fragment ion coverage? The results of online parallel ion parking during ETD of the light chain of IgG were promising with regard to the size of the fragments observed. However, large gaps in sequence and fragment ion coverage in some regions of the protein were most likely due to the lack of basic residues. Presumably, a protein that is similar in size with a higher density of basic residues would provide a more impressive degree of fragment ion coverage. Attempts at enhancing the charge of the light chain showed limited increases to fragment ion coverage in areas where the additional charge was deposited. Currently, work is being conducted to enhance the charge states exhibited in ESI of proteins. This includes derivatization techniques which add charge to certain amino acid side chains and "supercharging" techniques which drive charge states higher by increasing the basicity of the peptide backbone.

2.8 References

1. Bailey AO, Panchenko T Shabanowitz J, Lehman SM, Bai DL, Hunt DF, Black BE, Foltz DR. Identification of post-translational modifications in centromeric chromatin. *Mol Cell Proteomics.* **2016,** *15,* 918-931.

2. Richards AL, Hebert AS, Ulbrich A, Bailey DJ, Coughlin EE, Westphall MS, Coon JJ. One-hour proteome analysis in yeast. *Nat Protoc.* **2015,** *10*, 701-714.

3. Gregorich ZR, Chang YH, Ge Y (2014) Proteomics in heart failure: top-down or bottom-up? *Pflugers Arch – Eur J Physiol* 466: 1199-1209.

4. Smith LM, Kelleher NL. Proteoform: a single term describing protein complexity. *Nat Methods.* **2013,** *10,* 186-187.

5. Fornelli L, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, Tsybin YO. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol Cell Proteomics.* **2012,** *11,* 1758-1767.

6. Anderson LC, Karch KR, Ugrin SA, Coradin M English AM, Sidoli S, Shabanowitz J, Garcia BA, Hunt DF. Analyses of histone proteoforms using front-end electron transfer dissociation-enabled Orbitrap instruments. *Mol Cell Proteomics*. **2016,** *15*, 975-988.

7. Shaw JB, Li W, Holden DD, Zhang Y, Griep-Raming J, Fellers RT, Early BP, Thomas PM, Kelleher NL, Brodbelt JS. Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J Am Chem Soc.* **2013,** *135,* 12646-12651.

8. Coon JJ, Ueberheide B, Syka JEP, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci USA.* **2005,** *102,* 9463-9468.

9. Anderson LC, English AM, Wang W, Bai DL, Shabanowitz J, Hunt DF. Protein derivatization and sequential ion/ion reactions to enhance sequence coverage produced by electron transfer dissociation mass spectrometry. *Int J Mass Spectrom.* **2015,** *377,* 617-624.

10. Earley L, Anderson LC, Bai DL, Mullen C, Syka JEP, English AM, Dunyach JJ, Stafford GC, Jr., Shabanowitz J, Hunt DF, Compton PD. Front-end electron transfer dissociation: a new ionization source. *Anal Chem*. **2013,** *85*, 8385-8390.

11. Coon JJ, Ueberheide B, Syka JEP, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci*. **2005,** *102*, 9463-9468.

12. Chrisman PA, Pitteri SJ, McLuckey SA. Parallel ion parking: improving conversion of parents to first-generation products in electron transfer dissociation. *Anal Chem*. **2005,** *77*, 3411-3414.

13. Schwartz JC, Senko MW, Syka JE. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom*. **2002,** *13*, 659-669.

14. March RE. An introduction to quadrupole ion trap mass spectrometry. *J Mass Spec*. **1997,** *32*, 351-369.

15. Kaiser RE, Jr., Cooks RG, Syka JEP, Stafford GC, Jr.. Collisionally activated dissociation of peptides using a quadrupole ion trap mass spectrometer. *Rapid Commun Mass Spectrom*. **1990,** *4*, 30-33.

16. Dass C. *Principles and Practices of Biological Mass Spectrometry*. **2001,** John Wiley & Sons, Inc.

17. Wells JM, Chrisman PA, McLuckey SA. Formation and characterization of protein-protein complexes in vacuo. *J Am Chem Soc*. **2003,** *24*, 7238-7249.

18. McLuckey SA, Reid GE, Wells JM. Ion parking during ion/ion reactions in electrodynamic ion traps. *Anal Chem*. **2002,** *74*, 336-346.

19. Beck A, Sanglier-Cianferani S, Van Dorsselaer A. Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry. *Anal Chem.* **2012,** *84,* 4637-4646.

20. Fernandez L, Kalume D, Calvo L, Mallo M, Vallin A, Roepstorff P. Characterization of a recombinant monoclonal antibody by mass spectrometry combined with liquid chromatography. *J Chromatogr B.* **2001,** *752,* 247-261.

21. Robert GD, Johnson WP, Burman S, Anumula KR, Carr SA. An integrated strategy for structural characterization of the protein and carbohydrate components of monoclonal antibodies: applications to anti-respiratory syncytial virus mab. *Anal Chem.* **1995,** *67,* 3613-3625.

22. Zhang L, English AM, Bai DL, Ugrin SA, Shabanowitz J, Ross MM, Hunt DF, Wang W. Analysis of monoclonal antibody sequence and post-translational modifications by time-controlled proteolysis and tandem mass spectrometry. *Mol Cell Proteomics.* **2016,** *15,* 1479-1488.

23. Liu H, May K. Disulfide bond structures of IgG molecules structural variations, chemical modifications and possible impacts to stability and biological function. *Mabs.* **2012,** *4,* 17-23.

24. Ruegg UT, Rudinger J. Reductive cleavage of cystine disulfides with tributylphosphine. *Methods Enzymol.* **1977,** *47,* 111-116.

25. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein

sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA*. **2004,** *101*, 9528-9533.

26. Smythe CV. The reaction of iodoacetate and iodoacetamide with various sulfhydryl groups, with urease, and with yeast preparations. *J Biol Chem.* **1936,** *114,* 601-612.

Chapter 3 IIPT and Parallel Ion Parking for Identifications of Proteins in a Mixture

3.1 Abstract

In the previous chapter, we demonstrated the ability of parallel ion parking to slow ion-ion reaction rates and preserve large ETD fragments, enhancing protein sequence coverage. This chapter will focus on an alternative application of this technology; the identification of intact proteins in complex biological mixtures by combining the signal from protein cation into a single charge state using parallel ion parking. Identification of intact proteins in biological mixtures is made difficult by the wide range of charge states produced by electrospray ionization (ESI). This limits the signal-to-noise ratio of each protein's signal as the ion current that can be attributed to each species is split amongst all charge states. In addition, data-dependent selection of proteins for fragmentation is based on peak height in the survey scan. This typically means that the same protein is selected over and over again at different charge states. Concentrating each protein's signal into a single charge state increases both the S/N of each protein and the likelihood that all proteins will be selected for fragmentation, allowing them to be positively identified.

Here we describe the implementation of parallel ion parking during ion-ion proton transfer (IIPT) reactions. Parallel ion parking during IIPT enables us to slow the reaction rate of a specific charge state of a protein, allowing higher charge state products to continue to react. We first identify a high mass IIPT reagent to increase parking efficiency and implement reagent activation to more effectively park product ions.

Additionally, we introduce wide isolation windows and wideband excitation regions to enable the parking and identification of unknown species. Ultimately, we demonstrate the utility of our method in providing the accurate mass measurement of intact ribosomal proteins from human and *Escherichia coli* cells on a chromatographic time scale. Subsequent targeted ETD of human ribosomal proteins provided sufficient fragment ion coverage of several species, allowing for identification. Alternatively, data dependent fragmentation of the purified products by HCD allowed us to confirm the identity of each *E. coli* ribosomal protein via the amino acid sequence. The spectra obtained from the experiment were used to identify 46 of 54 ribosomal proteins by search software against the *E. coli* database. Manual data interpretation not only verifies the accuracy of the automated results, but resulted in the identification of 53 of 54 ribosomal proteins by intact mass and at least 5 sequence-determinate fragment ions. Furthermore, many of the proteins identified contained post-translational modifications, single amino acid mutations, or sequence truncations. This data was acquired over the course of a single one hour HPLC-MS experiment and required only 25 fmol of each protein. For comparison purposes, a traditional method was employed which did not include a parallel ion parking scan. This experiment resulted in only 38 successfully identified proteins, further illustrating the practicality of our improved methodology.

3.2 Introduction

In the analysis of protein mixtures, proteins are typically chemically or enzymatically digested. The resulting peptides are chromatographically separated and

analyzed by mass spectrometry. Identification of a particular peptide allows for the inferred identification of the source protein. As was discussed in **Chapter 1**, several disadvantages are associated with this type of analysis, including the inability to differentiate multiple modifications as they exist in context to one another. Analyzing proteins as intact molecules avoids several of these disadvantages by providing a more comprehensive picture of each proteoform [1, 2]. However, intact analysis inherently suffers from low signal-to-noise ratios compared to analysis of smaller digested peptides. This is primarily due to the fact that large proteins exhibit a wide range of charge states when ionized by ESI [3]. Thus, the signal of each protein is spread amongst several peaks. Recall that the 17 kDa, 153 amino acid protein apomyoglobin exhibits charge states of $[M+12H]^{12+}$ through $[M+30H]^{30+}$ by ESI (**Figure 1.2**). The most abundant charge state consists of only about 12% of the total signal attributed to the protein. In this chapter we will describe a technique which concentrates a majority of the signal from a protein into a single charge state. This increases the signal-to-noise ratio of the protein and allows for detection and fragmentation of lower abundance species.

Typically, IIPT is used to reduce the charge of product ions from ETD reactions to spread the fragment ions throughout the available m/z range [4, 5]. However, IIPT can also be used to reduce the charge state of intact species [6], making it an ideal tool for manipulating the charge state of proteins for our purposes. However, IIPT products exhibit a range of charge states similar to those seen in the ESI spectrum. This further dilutes the signal from each protein. In addition, uncontrolled IIPT will carry products

beyond the available 4,000 m/z range. Given enough reaction time, products will eventually be neutralized [7]. We therefore aim to use parallel ion parking to stop the IIPT reaction once the product reaches a specified charge state. Products that have not reached the desired m/z continue to react, resulting in a majority of the ion current being channeled into a single peak.

Several of the parking principles discussed in **Chapter 2** apply to the methods described here. Similar to ETD, IIPT is a gas-phase ion-ion reaction for which the rate is proportional to the square of the precursor charge state and inversely proportional to the cube of the relative velocity between the two reactants. Increasing the velocity of specific products will thus decrease their subsequent reaction rate [8]. Parallel ion parking can also be used to preserve products of IIPT. Because IIPT products are always higher in m/z than the precursor, only the frequency components corresponding to the reagent anion and the m/z range which includes the desired charge state are included in the waveform. Thus, products that have not undergone a sufficient number of proton transfer events to reach the target m/z are allowed to undergo further reactions. However, as the product reaches the target m/z, it stops reacting and all signal from an analyte is concentrated into a single peak.

The goal of this project is to develop a method that utilizes IIPT and parallel ion parking to increase the signal-to-noise ratio of intact proteins by combining the protein signal into a single peak. To be of practical use, the method must be applicable on a

chromatographic time scale and enable the fragmentation of the charge-reduced proteins for unambiguous identification of each proteoform. We demonstrate the power of our novel method through the identification of proteins in a biological mixture during a single one hour chromatography gradient.

3.3 Materials and Instrumentation

Agilent Technologies (Palo Alto, CA)

> 1100 Series high performance liquid chromatograph

> 1100 Series vacuum degasser

> POROSHELL 300SB-C18 resin (5-μm, 300-Å)

Branson (Danbury, CT)

> Branson 1200 ultrasonic bath

Eppendorf (Hauppauge, NY)

> 5414R Benchtop centrifuge

GTS Welco (Allentown, PA)

> 10 ppm SF_6 in nitrogen gas

Honeywell (Morristown, NJ)

> Burdick and Jackson® Acetonitrile, LC-MS grade

Labconco Corporation (Kansas City, MO)

> Centrivap centrifugal vacuum concentrator

Millipore (Billerica, MA)

> Amicon® Ultra-0.5 10K centrifugal filter

Molex (Lisle, IL)

Polymicro Technologies™ polyimide coated fused silica capillary

Sizes: 360 μm o.d. x 25, 50, 75 & 150 μm i.d.

New England Biolabs (Ipswitch, MA)

Escherichia coli Ribosome (B Strain)

Oakwood Chemical (Estill, SC)

Perfluoromethyldecalin

Omega Engineering (Stamford, CT)

Temperature controllers

Parker Hannifin Corporation (Cleveland, OH)

Porter back pressure regulator

Porter mass flow controller

PCR, Inc. (Gainesville, FL)

Tris(perfluorononyl)-S-triazine

PQ Corporation (Valley Forge, PA)

Kasil – Potassium silicate solution

Santa Cruz Biotechnology (Dallas, TX)

Perfluoroadamantane

SGE Analytical Science (Melbourne, Australia)

PEEKsil tubing 1/16" o.d., 0.025 mm i.d.

Sigma Aldrich (St. Louis, MO)

2-propanol, LC-MS grade

Ammonium Acetate

Ammonium Hydroxide

Apomyoglobin from equine skeletal muscle,

protein sequencing standard, lyophilized powder

Decafluorobiphenyl

Dithiothreitol for molecular biology

Fluoranthene

Glacial acetic acid

Iodoacetamide

Perfluoro-1,3-dimethylcyclohexane

Perfluorodecalin

Trichloroacetic acid

Ubiquitin from bovine erythrocytes

Sutter Instrument Co. (Novato, CA)

P-2000 microcapillary laser puller

Swagelok (Solon, OH)

316 stainless steel plug for 1/8" Swagelok® tube fitting

Tektronix (Beaverton, OR)

MDO3014 Mixed domain oscilloscope

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

Calibration mixture

Formic acid, LC-MS grade

LTQ mass spectrometer (custom modified with back-end ETD)

Orbitrap Elite™ mass spectrometer (custom modified with front-end ETD)

Upgraded from Velos Pro™ July, 2014

Pierce® water, LC-MS grade

Urea

Zeus Industrial Products (Orangeburg, SC)

Teflon tubing, 0.012" i.d. x 0.060" o.d.

3.4 Methods

3.4.1 Instrument Modification and Operation

To enable parallel ion parking, custom instrument control (ITCL) software was written for implementation of ion parking on an FETD-enabled Orbitrap Elite™ capable of performing multiple product ion fills of the C-trap [9]. Supplemental waveforms were designed for application during the charge-sign independent trapping portion of IIPT gas phase reactions (**Figure 3.1**). Waveform frequency ranges included all frequencies within the given range at 500 Hz increments. Unless otherwise specified, the waveform included frequencies corresponding to the IIPT reagent (perfluoromethyldecalin; m/z 512) +/- 10 m/z. Also included was a 1000 m/z range within which products were parked. The actual m/z range used was selected based on the experiment and will be specified.

Several high mass chemicals were evaluated as candidates for IIPT reagents: 2,4,6-Tris(perfluorononyl)-S-triazine, Perfluoroadamantane, Decafluorobiphenyl,

Perfluoro-1,3-dimethylcyclohexane, Perfluorodecalin, and Perfluoromethyldecalin. A customized linear trap quadrupole (LTQ; Thermo Fisher Scientific, San Jose, CA) instrument was used for the purpose of screening chemicals as candidates for IIPT reagents. Modifications to the instrument have been previously reported [10]. Briefly, reagents are stored in the cavity of a stainless steel Swagelok™ cap and heated to their boiling point in the oven of a gas chromatograph. Reagent gas is ionized by chemical ionization (CI) in a dedicated vacuum chamber. Reagent anions produced by CI are isolated within the ion trap by adjusting the trapping RF amplitude to destabilize unwanted ions.
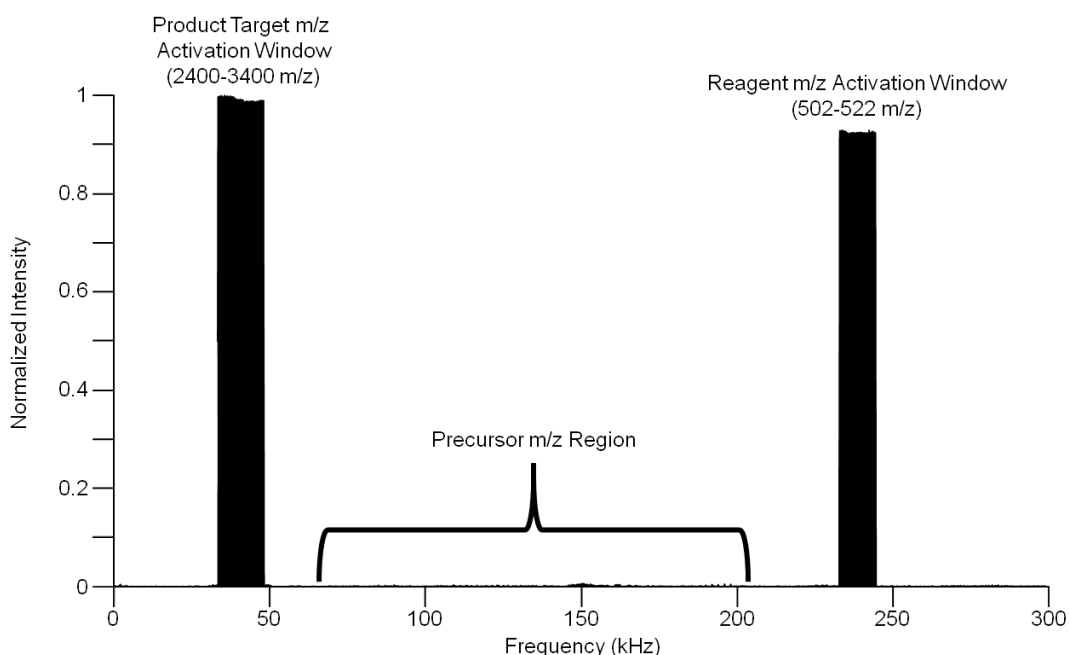
Figure 3.1. Parking waveform applied during charge-sign independent trapping portion of IIPT. Frequency range windows used for parking during IIPT include a 20 m/z window centered on the reagent anion and a 1000 m/z window within which product ions will be parked. The m/z range of the product target window is altered depending on the experiment.

Liquid reagents selected for IIPT on the customized Orbitrap Elite™ were contained in a stainless steel Swagelok™ cap and allowed to evaporate at room temperature. A two meter length of fused silica capillary (360 µm o.d. x 10 µm i.d.) carried evaporated reagent to the ionization volume of the FETD reagent ion source. Anions higher in mass than the reagent were removed by a DC offset applied to a quadrupole in the ion optics system. Anions lower in mass than the reagent were removed by adjusting the trapping RF amplitude to control their stability.

3.4.2 Direct Infusion of Standard Proteins

Ubiquitin (bovine) and apomyoglobin (equine) were purchased from Sigma-Aldrich (St. Louis, MO). One nanomole of lyophilized apomyoglobin was reconstituted in a solution containing 40% acetonitrile and 60%, 0.1% acetic acid in water to a final concentration of 2.5 pmol/µL (all percentages expressed as v/v). Alternatively, a standard protein mixture was prepared which contained ubiquitin and apomyoglobin (each 2.5 pmol/µL) in a solution containing 40% acetonitrile and 60%, 0.1% acetic acid in water. The samples were directly infused via a fused silica capillary emitter tip (360 µm o.d. x 75 µm i.d.) and ionized by micro-electrospray ionization.

MS/MS spectra were acquired using the following instrument parameters: resolution (R) = 120,000 at 400 m/z; 300–4,000 m/z scan range; 1×10^5 ion counts precursor automatic gain control (AGC) target; 3×10^5 reagent (perfluoromethyldecalin) AGC target; 1 scan/spectrum; 5 or 400 m/z precursor isolation window centered at 653

([M+16H]$^{16+}$ charge state of apomyoglobin) or 800 m/z, respectively; 25–120 ms IIPT. The parking waveform was applied during the charge-sign independent trapping portion of the IIPT reaction. The product target activation window included all frequencies corresponding to a 1,000 m/z range starting at either 1,150 or 1,450 m/z at a normalized activation amplitude of 0.090. Where specified, the reagent activation window included all frequencies corresponding to a 20 m/z window centered on 512 m/z at a normalized activation amplitude of 0.020.

3.4.3 LC-MS Identification of Ribosomal Protein Mixture

Human Ribosomal Proteins by IIPT and Wide Band ETD

Intact human ribosomal proteins were purified and provided by members of the Hart laboratory at the Johns Hopkins University School of Medicine. Human HepG2 cell pellets were lysed in TMK buffer (Tris-HCl, MgCl$_2$, and KCl, pH = 7.5) supplemented with 1% Triton-X (Sigma-Aldrich; St. Louis, MO), 1 µM 1,1,3,3-tetramethylguanidine (TMG), 100 µg/mL cycloheximide, 2 mM phenylmethanesulfonyl fluoride (PMSF), and protease and phosphatase inhibitors by pipetting up and down several times. The suspension was incubated in ice for 15 minutes and then passed through a 20 or 22G needle and incubated in ice for an additional 15 minutes. The solution was then spun at 3000g for 15 minutes. The supernatant was loaded onto 1 M sucrose cushions prepared in TMK buffer supplemented with 100 µg/mL cycloheximide and 1 µM PUGNAc (N-acetylhexosaminidase inhibitor). The samples were subjected to ultracentrifugation at 41,000 RPM for 3 hours at 4º C. The resulting ribosomal pellets were then resuspended in

200 µL TMK buffer with 1 mM dithiothreitol (DTT) to which was added 20 µL MgCl$_2$

and 400 µL acetic acid. The sample was incubated with agitation at 4º C for 30 minutes

and then spun down at 14,000 RPM for 15 minutes at 4º C. The supernatant was

transferred to a 15 mL conical vial to which 6 mL acetone was added. The sample was

then spun down for 1 minute at 14,000 RPM and washed three times with cold acetone.

We received the sample in 8 M urea and 50 mM triethylammonium bicarbonate

(TEABC) at an estimated concentration of 1.5 µg/µL. The sample (1 µL) was diluted in

7.3 µL 8 M urea, 50 mM ammonium acetate to a final concentration of approximately 4

pmol/µL total protein. The sample (2 pmol) was pressure-loaded onto a precolumn (360

µm o.d. x 150 µm i.d. fused silica capillary packed with 14 cm POROSHELL 300SB-

C18 5 µm diameter (Agilent)). The sample was desalted by flowing solvent A through

the column for one hour (Solvent A: 0.3% formic acid in water). The precolumn was then

connected to an analytical column (360 x 75 µm i.d. fused silica capillary packed with 22

cm POROSHELL 300SB-C18 5 µm diameter (Agilent)) for HPLC-MS/MS analysis. An

Agilent (Palo Alto, CA) 1100 Series binary HPLC system was interfaced with the

Orbitrap Elite mass spectrometer for online protein separation. Proteins were eluted at

100 nL/min using the following gradient: 0-30% solvent B in 10 minutes, 30-70% B in

130 minutes, and 70-100% B in 140 minutes (Solvent B: 0.3% formic acid, 72%

acetonitrile, 18% isopropanol, 9.7% water).

MS1 scans were taken in the Orbitrap mass analyzer from 300-4,000 m/z (R =

120,000 at 400 m/z; 1e6 AGC target). A single IIPT MS/MS scan was also taken in the

Orbitrap under the following conditions: R = 120,000 at 400 m/z; 480-4,000 m/z scan range; 1×10^5 precursor and 3×10^5 reagent AGC targets; 10 product ion fills; 1 microscan; 400 m/z isolation window centered at 800 m/z; 150 ms IIPT. The parking waveform was applied during IIPT MS/MS and included all frequencies corresponding to the m/z ranges 502-522 and 1,150-2,150. A second isolation-only MS/MS scan was taken in the Orbitrap for comparison purposes. An ETD, IIPT MS/MS scan was taken the Orbitrap under the following conditions: R = 120,000 at 400 m/z; 200-4,000 m/z scan range; 10 product ion fills; 1 microscan; 400 m/z isolation window centered at 800 m/z; 20 ms ETD, 40 ms IIPT. The ETD reagent was fluoranthene and the IIPT reagent was perfluoromethyldecalin (PFMD).

Escherichia Coli Ribosomal Proteins by IIPT and Data-Dependent HCD

Alternatively, Ribosomes from *Escherichia coli* (B Strain) were purchased from New England Biolabs (Ipswitch, MA). Intact ribosomes were reduced and alkylated using dithiothreitol and iodoacetamide, respectively. For LC-MS analysis, the protein mixture (1.33 pmol total protein) was pressure-loaded onto a precolumn (360 μm o.d. x 150 μm i.d. fused silica capillary packed with 14 cm POROSHELL 300SB-C18 5 μm diameter). The sample was desalted by flowing solvent A through the column for one hour (Solvent A: 0.3% formic acid in water). After desalting, the precolumn was connected to an analytical column (360 μm o.d. x 75 μm i.d. fused silica capillary packed with 22 cm POROSHELL 300SB-C18 5 μm diameter) which was equipped with a laser-pulled nanoelectrospray emitter tip. Proteins were eluted at 100 nL/min using the

following gradient: 0-30% solvent B in 10 minutes, 30-70% B in 50 minutes, and 70-100% B in 60 minutes (Solvent B: 0.3% formic acid, 72% acetonitrile, 18% isopropanol, 9.7% water).

MS1 scans were taken in the Orbitrap mass analyzer from 300-2000 m/z (R = 120,000 at 400 m/z; 1e6 AGC target). A single IIPT MS/MS scan, upon which data-dependent selection was based, was also taken in the Orbitrap under the following conditions: R = 120,000 at 400 m/z; 300-2,000 m/z scan range; 1×10^5 precursor and 3×10^5 reagent AGC targets; 10 product ion fills; 1 microscan; 400 m/z isolation window centered at 800 m/z; 100 ms IIPT. The parking waveform was applied during IIPT MS/MS and included all frequencies corresponding to the m/z ranges 502-522 and 1,150-2,150. A second isolation-only MS/MS scan was taken in the Orbitrap for comparison purposes. A higher energy collisional dissociation (HCD) MS3 scan was taken in a data-dependent manner of the most intense species in the IIPT parked MS/MS spectrum under the following conditions: R = 120,000 at 400 m/z; 200-2,000 m/z scan range; 10 product ion fills; 1 microscan; 5 m/z isolation window; 35% normalized collision energy.

3.4.4 Data Analysis

Human ribosomal protein high resolution full MS and MS/MS data were manually inspected using Qual Browser software (Thermo Fisher Scientific). Interpretation of ETD/IIPT MS/MS data was performed manually on the unprocessed raw spectra. In addition, MS/MS spectra were searched using an in-house developed

software program that compared the m/z of peaks in the spectra to the peaks predicted by *in silico* ETD fragmentation of a ribosomal protein database. Results of this program were manually validated. Theoretical fragment ion masses were calculated using in-house developed fragment mass calculator software. Theoretical isotopic distributions were modeled using Isotope Pattern Calculator (IPC) software (Pacific Northwest National Laboratories; http://omics.pnl.gov/). Mass difference in parts per million (ppm) was calculated by comparing the observed mass of the most abundant isotope to the theoretical mass of the most abundant isotope predicted by IPC. The signal-to-noise ratio was calculated by dividing the normalized level (NL) signal of the most abundant charge state of each protein by the noise level reported by the instrument software.

Escherichia coli ribosomal protein mixture data was searched via ProSight PD™ against the *E. coli* intact database (uniprot; ecoli_swissprot_032416). A three-step search was performed: an absolute mass search with narrow parameters (100 Da precursor mass tolerance; 10 ppm fragment mass tolerance; delta m mode off), a biomarker search (10 ppm precursor mass tolerance; 10 ppm fragment mass tolerance), and finally a second absolute mass search with wider search parameters (1,000 Da precursor mass tolerance; 10 ppm fragment mass tolerance, delta m mode on). Spectra that were not identified with high confidence in one search were passed to the next search for further analysis. For all searches, a precursor S/N threshold of 3 and fragment ion S/N threshold of 1.5, and static modification of carbamidomethyl cysteine were used. All data were manually validated to confirm ProSight™ results. Theoretical fragment ion masses were calculated using in-

house developed fragment mass calculator software. Theoretical isotopic distributions were modeled using Isotope Pattern Calculator (IPC) software (Pacific Northwest National Laboratories). Mass difference in parts per million (ppm) was calculated by comparing the observed mass of the most abundant isotope to the theoretical mass of the most abundant isotope predicted by IPC.

3.5 Results and Discussion

3.5.1 High Mass IIPT Reagent

During early experiments, we experienced difficulty parking products above 1,500 m/z. At low amplitudes the products would continue to react beyond the targeted m/z range. When the amplitude of the parking window was increased, product ions were ejected before they were effectively parked. Recall from Chapter 1 that the stability of each trapped ion is defined by its Mathieu q-value. During IIPT, the q-value of the reagent m/z is set to a value of 0.55. This value was selected to avoid the phenomenon of parametric resonance that occurs when trapped ions fall into resonance with the end lens RF applied during charge-sign independent trapping [11]. Parametric resonance occurs at a q-value of approximately 0.62. All products of IIPT will have a q-value below that set for the reagent (0.55) and will therefore not be affected by parametric resonance. However, as the m/z of an ion gets higher (and its q value gets lower), it is less stable in the ion trap. Thus, even the small amount of activation used for parking was causing the higher m/z ions to be ejected rather than parked.

In order to increase the stability of the high m/z ions, we sought to increase their

q-value by evaluating several higher mass reagents for the ability to perform IIPT. Setting

the same q-value to a higher mass reagent allows us to increase the q-value for each m/z

in the trap without the need to change the q-value at which the reagent anion is held.

Several higher mass reagents were selected for evaluation. These experiments were

performed on an LTQ instrument that was modified to introduce ETD and IIPT reagents.

Reagents were stored in a stainless steel reservoir and heated in the oven of a gas

chromatograph to ensure vaporization. An ion-ion reaction between a reagent anion and a

protein cation can either result in the transfer of a radical electron from the reagent to the

cation or the transfer of a proton from the cation to the reagent anion [12]. Reagent

anions were evaluated by allowing them to react with the $[M+13H]^{13+}$ charge state of

ubiquitin for 20 ms. The resulting spectra were then evaluated qualitatively to answer

three questions: 1) Does the reagent ionize? 2) Does proton transfer occur? 3) Is there any

evidence of electron transfer?

Figure 3.2A shows the reagent ion scan of 2,4,6-Tris(Perfluorononyl)-S-Triazine

(m/z 1,485.25). Several peaks were observed in the spectrum, resulting from the loss of

multiple fluorine atoms. Fragmentation likely occurred upon injection and trapping of the

reagent ions in the LTQ via collisions with the helium bath gas. The MS/MS spectrum

resulting from 20 ms reaction of triazine anions with the $[M+13H]^{13+}$ charge state of

ubiquitin is shown in **Figure 3.2B**. A majority of the product ions were the result of

charge-reduction by IIPT. However, evidence of a small amount of ETD was observed at

approximately 2% relative abundance, as is shown in the inset. We disqualified the

reagent because it did not produce exclusively IIPT products. In addition, the ease with

which fragmentation occurred was concerning. As we will discuss in a later section, we

determined that activation of the reagent was necessary for efficient parking. A labile

reagent was therefore undesirable as it is likely that the reagent would fragment during

the parking process and negatively affect the ion-ion kinetics.
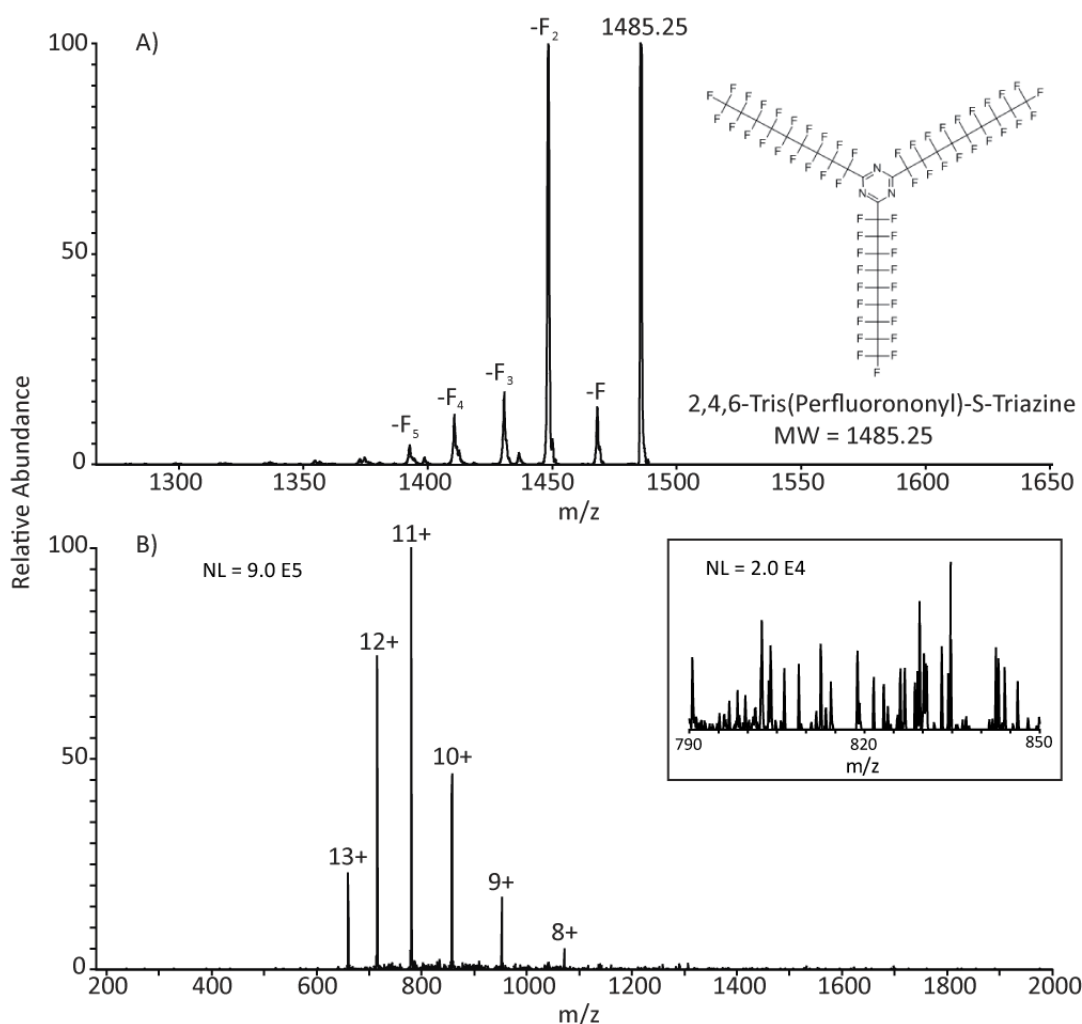
Figure 3.2. IIPT reagent screen of 2,4,6-Tris(perfluorononyl)-S-triazine. Reagent anion spectrum resulting from chemical ionization and ion trap injection of 2,4,6-Tris(perfluorononyl)-S-triazine (**A**). Result of 20 ms reaction between 2,4,6-Tris(perfluorononyl)-S-triazine and [M+13H]$^{13+}$ charge state of ubiquitin (**B**). Inset of (**B**) shows ion signal between reduced charge products.
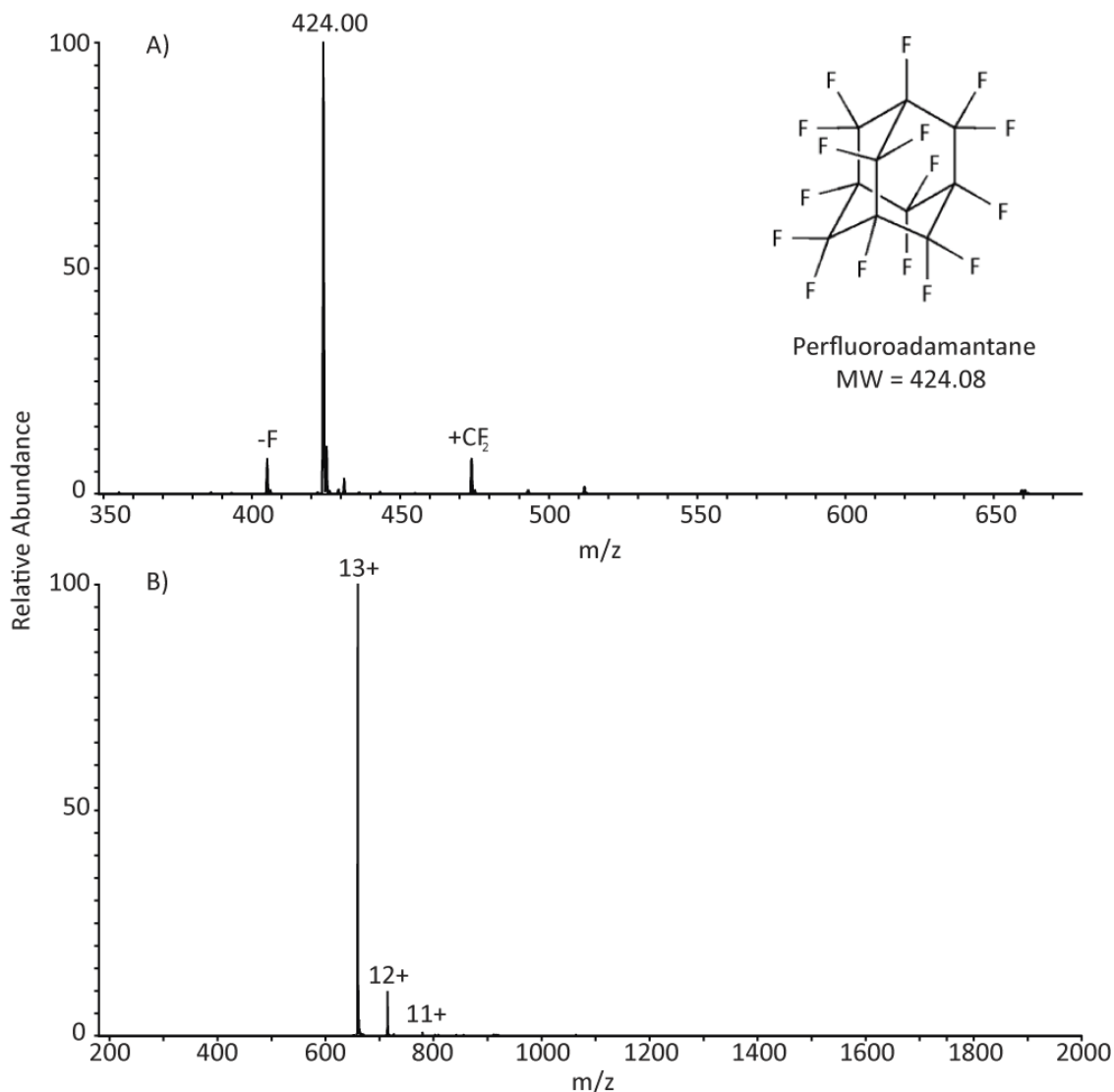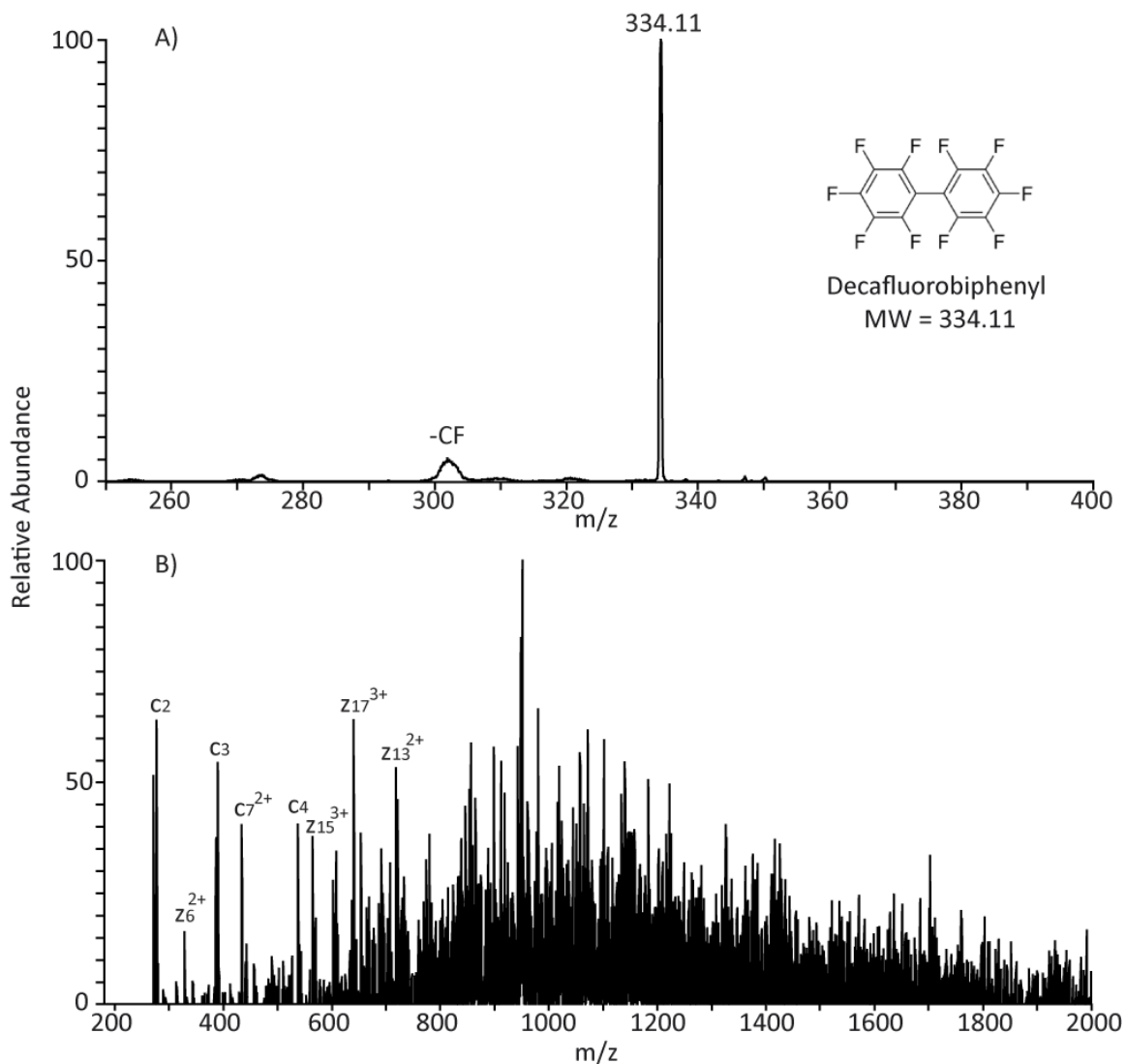
Figure 3.3. IIPT reagent screen of perfluoroadamantane. Reagent anion spectrum resulting from chemical ionization and ion trap injection of perfluoroadamantane (**A**). Result of 20 ms reaction between perfluoroadamantane and [M+13H][13+] charge state of ubiquitin (B)

Figure 3.3A shows the reagent ion spectrum of perfluoroadamantane (m/z

424.08). The reagent spectrum exhibited two contamination peaks, each at approximately

10% relative abundance. One peak corresponded to the loss of a fluorine atom from the

reagent anion while the other peak corresponded to the addition of one carbon atom and

two fluorine atoms. Reagent documentation suggested that perfluoromethyladamantane

contamination was expected. However, the loss of fluorine was likely the result of reagent anion fragmentation. **Figure 3.3B** shows the result of 20 ms reaction between perfluoroadamantane and ubiquitin. Interestingly, very few products were observed. Those that were observed (approximately 10% relative abundance) were exclusively the result of proton transfer. We disqualified perfluoroadamantane because it did not demonstrate a high enough IIPT rate to be useful on a chromatographic timescale.

     Figure 3.4A shows the reagent ion scan of decafluorobiphenyl (m/z 334.11). Of the candidate reagents screened, decafluorobiphenyl produced the most unexpected results. **Figure 3.4B** shows the MS/MS spectrum resulting from the reaction of decafluorobiphenyl reagent anions and ubiquitin $[M+13H]^{13+}$ cations for 10 ms. Rather than producing charge-reduced proton transfer products, the primary ions in the MS/MS spectrum looked to be fragment ions resulting from ETD, although only low mass ions could be positively identified due to the low resolution of the spectra. Though there is currently no need for an alternative ETD reagent of higher mass, should future experiments call for one, decafluorobiphenyl appears to be a good candidate.
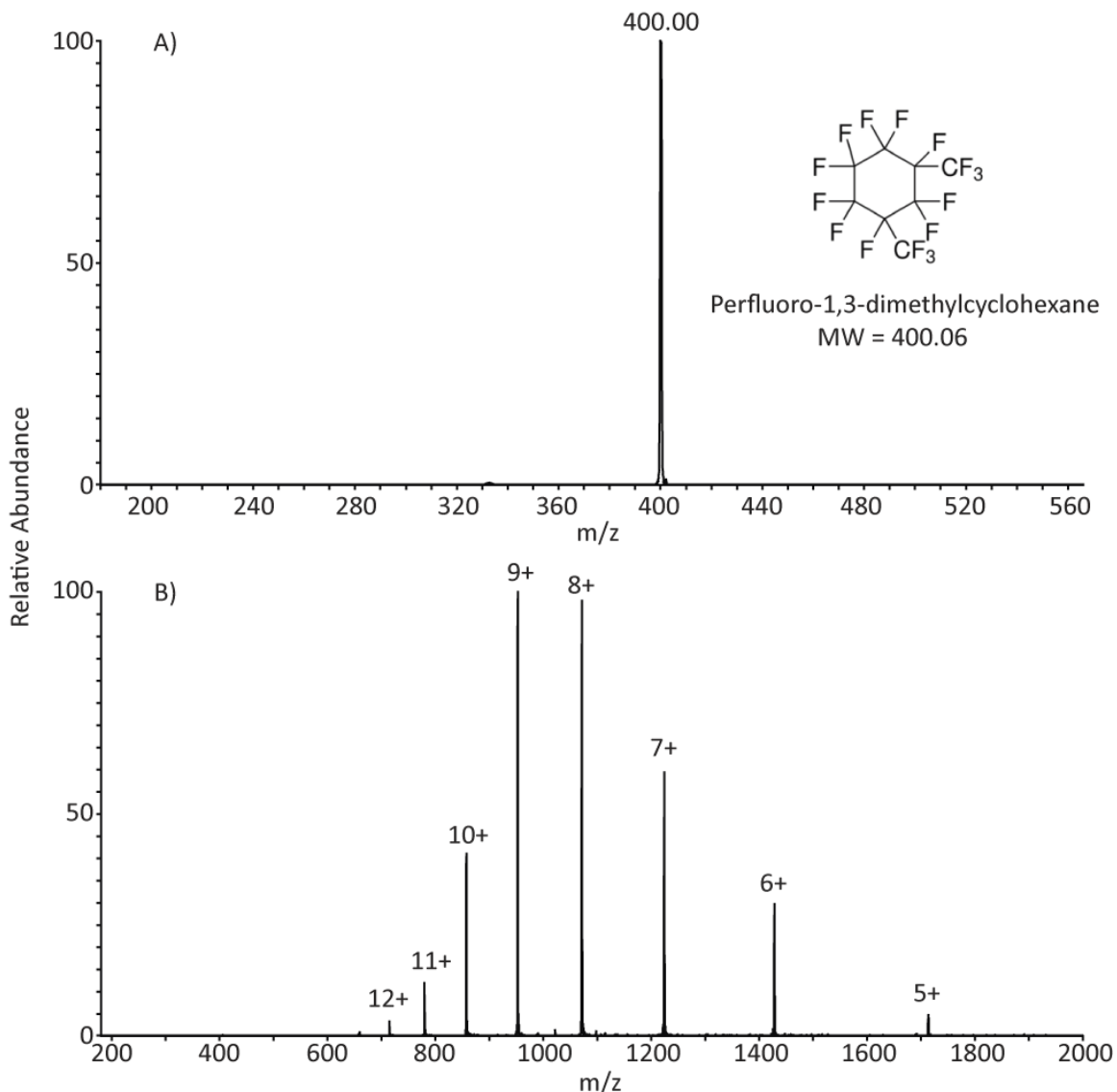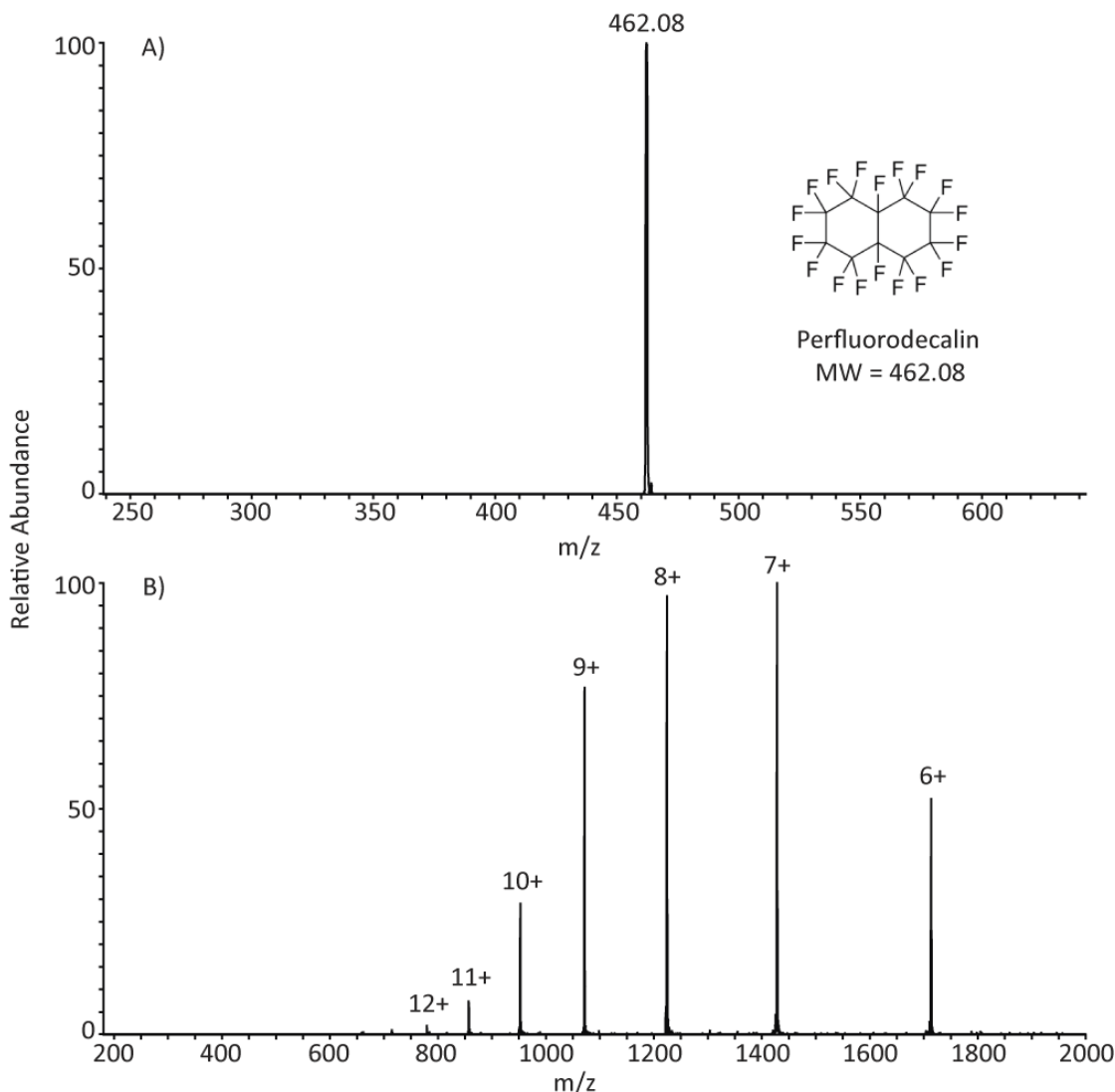
Figure 3.4. IIPT reagent screen of decafluorobiphenyl. Reagent anion spectrum resulting from chemical ionization and ion trap injection of decafluorobiphenyl (**A**). Result of 20 ms reaction between decafluorobiphenyl and $[M+13H]^{13+}$ charge state of ubiquitin (B)

Figure 3.5A shows the reagent ion scan of perfluoro-1,3-dimethylcyclohexane (PDCH; m/z 400.06). The reagent ion scan was very clean, with no peaks attributed to contaminants. **Figure 3.5B** shows the spectrum resulting from 20 ms reaction between PDCH and ubiquitin. PDCH has previously been identified as a compound whose radical anion will result in proton transfer from protein cations [13]. Indeed, the products

observed were exclusively reduced charge precursor ions, suggesting that PDCH

performs IIPT exclusively.

Figure 3.5. IIPT reagent screen of perfluoro-1,3-dimethylcyclohexane. Reagent anion spectrum resulting from chemical ionization and ion trap injection of perfluoro-1,3-dimethylcyclohexane (**A**). Result of 20 ms reaction between PDCH and $[M+13H]^{13+}$ charge state of ubiquitin (B)

Figure 3.6A shows the reagent ion scan of perfluorodecalin (PFD; m/z 462.08).

Similar to PDCH, PFD exhibited no contaminant peaks when ionized and injected into

the linear ion trap. The products of 20 ms reaction between PFD and ubiquitin, shown in

Figure 3.6B, were exclusively the result of proton transfer. It is worthwhile to note that

the reaction rate of perfluorodecalin was slightly faster than that of PDCH.

Figure 3.6. IIPT reagent screen of perfluorodecalin. Reagent anion spectrum resulting from chemical
ionization and ion trap injection of perfluorodecalin (**A**). Result of 20 ms reaction between PFD and
$[M+13H]^{13+}$ charge state of ubiquitin (B)

Perfluorodecalin and perfluoro-1,3-dimethylcyclohexane both performed IIPT

exclusively. **Figure 3.7** shows the relationship between m/z and Mathieu q for each of the

candidate reagents and the currently employed reagent, sulfur hexafluoride. Using our

experimentally observed limit of m/z 1,500, we determined that ions were expected to be

stable while parking if their q was above 0.053. Both PDCH and PFD offered parking

stability throughout the mass range of the instrument. However, the plot for PDCH

approaches the threshold near 4,000 m/z. Therefore, PFD was selected as it offers

increased parking stability at high m/z values.
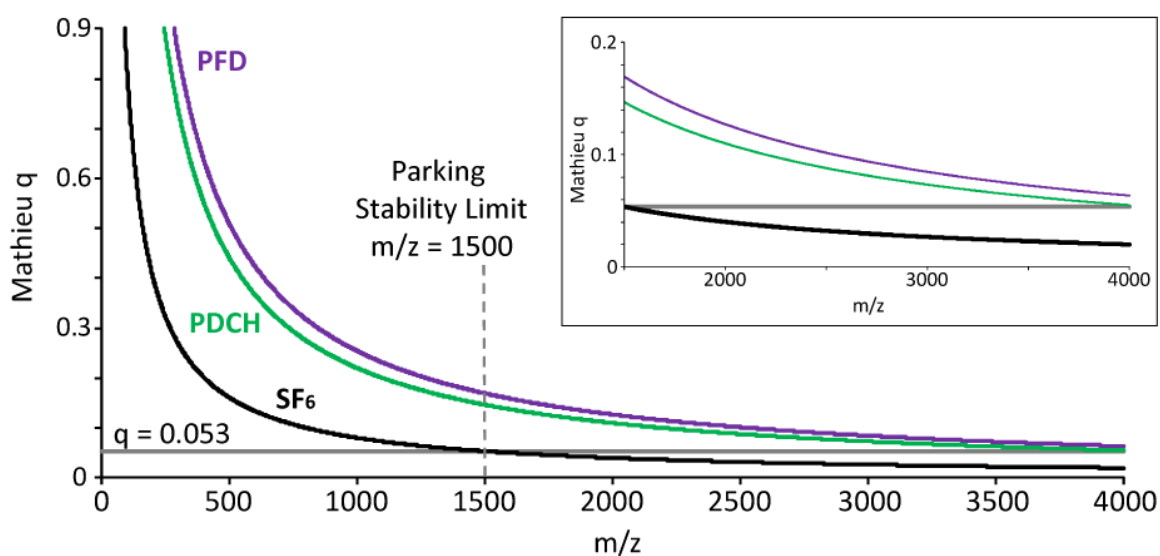
Figure 3.7. Mathieu q value as a function of m/z for candidate reagents. Theoretical relationship
between m/z and Mathieu q value for SF_6 (Black), PDCH (Green), and PFD (Purple). Assumes reagent
q value of 0.55. Both PFD and PDCH ensure all m/z values remain above the experimentally observed
limit (Grey) throughout the available mass range.

Twenty microliters of the reagent was stored in a stainless steel reservoir and fed

into the main reagent line through a 10 µm inner diameter, 2 meter length of fused silica.

ETD and IIPT reagents were ionized in the same reagent ion source and then optionally

isolated from one another by a resolving quadrupole in the ion optics system. We quickly

determined that too much PFD was reaching the reagent ion source as the ETD reagent

signal was lower by approximately an order of magnitude. We initially attempted to slow

the rate at which PFD was introduced by attempting a longer restriction length but the setup quickly became complicated and unreliable. Thus, we again turned our attention to alternative reagents.

PFD is liquid at room temperature with a vapor pressure of 6.60 torr at 25ºC [14]. The obvious alternative candidate, PDCH, has a vapor pressure of 36.0 torr and would have increased the amount of reagent reaching the source, exacerbating the problem [15]. Alternatively, we evaluated a methylated $(+CF_3)$ form of perfluorodecalin. Theoretically, perfluoromethyldecalin (PFMD) should have the similar characteristics to perfluorodecalin with a lower vapor pressure. Indeed, the reported vapor pressure for PFMD is 2.17 torr [16]. Introduction of PFMD in the same manner as previously stated gave a reagent ion signal similar to that of sulfur hexafluoride. In addition, ionization of PFMD did not quench the ionization of fluoranthene. Finally, PFMD performed IIPT at a rate and purity similar to PFD.

3.5.2 Reagent Activation

Shown in **Figure 3.8A** is the IIPT MS/MS spectrum resulting from apomyoglobin $[M+26H]^{26+}$ (m/z = 653) reacting with perfluoromethyldecalin radical anions for 25 ms. Concurrently, a waveform including all frequencies corresponding to m/z 1,450-2,450 was applied to the x-rods of the linear ion trap. Apomyoglobin's $[M+11H]^{11+}$ charge state has an m/z of 1,542 and will thus fall within the product target window. After a sufficient number of proton transfers have occurred, that product becomes excited and unreactive.

However, excitation of only the product window exhibits limited effectiveness, with a majority of the product ions continuing to react past the targeted $[M+11H]^{11+}$ product. In order to retain the ions in a single charge state, it is necessary to further slow the IIPT reaction.
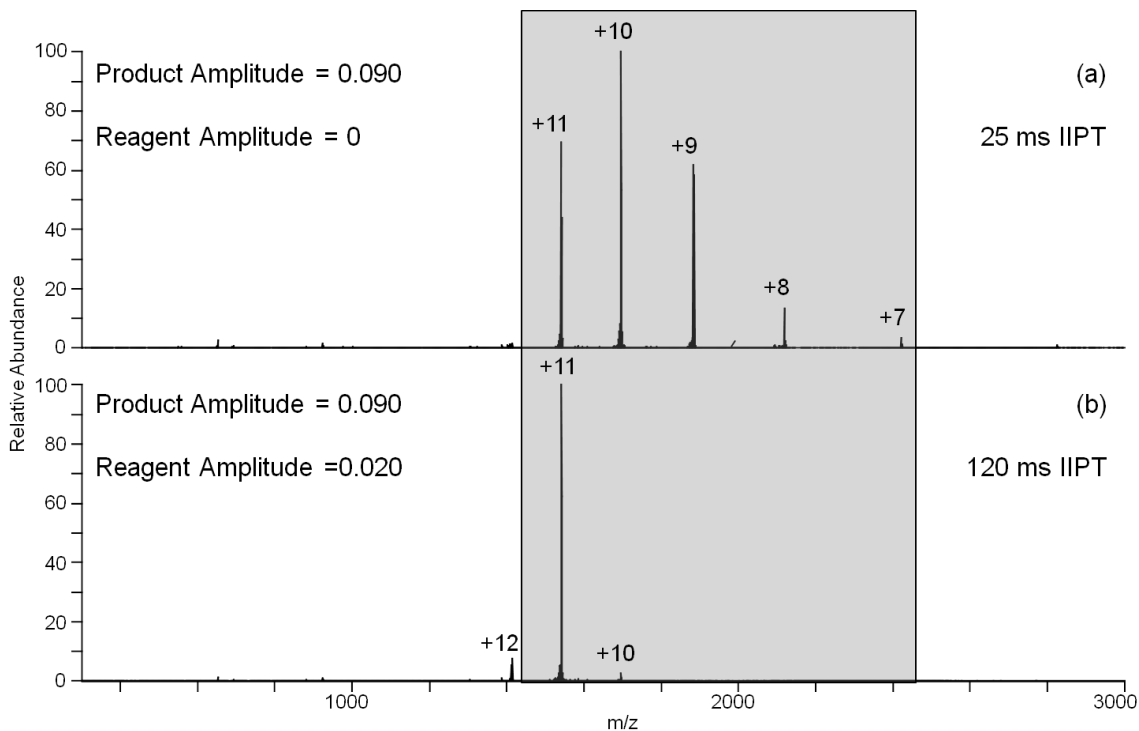

Figure 3.8. IIPT mass spectrum of apomyoglobin from the $[M+26H]^{26+}$ precursor. Excitation of only the target range (a) results in continued charge reduction past the selected $[M+11H]^{11+}$ product. Concurrent excitation of the reagent anion (b) provides additional reduction of reaction rate, allowing for efficient parking of >90% of the product signal in the $[M+11H]^{11+}$ charge state.

The most obvious solution was to apply a higher amplitude parking voltage to the target m/z range. However, the mechanism by which parking is accomplished is similar to that of collisionally activated dissociation (CAD). As such, increasing the amplitude of the parking voltage resulted in fragmentation of the intact protein. Therefore, an alternative method was devised to increase the parking efficiency.
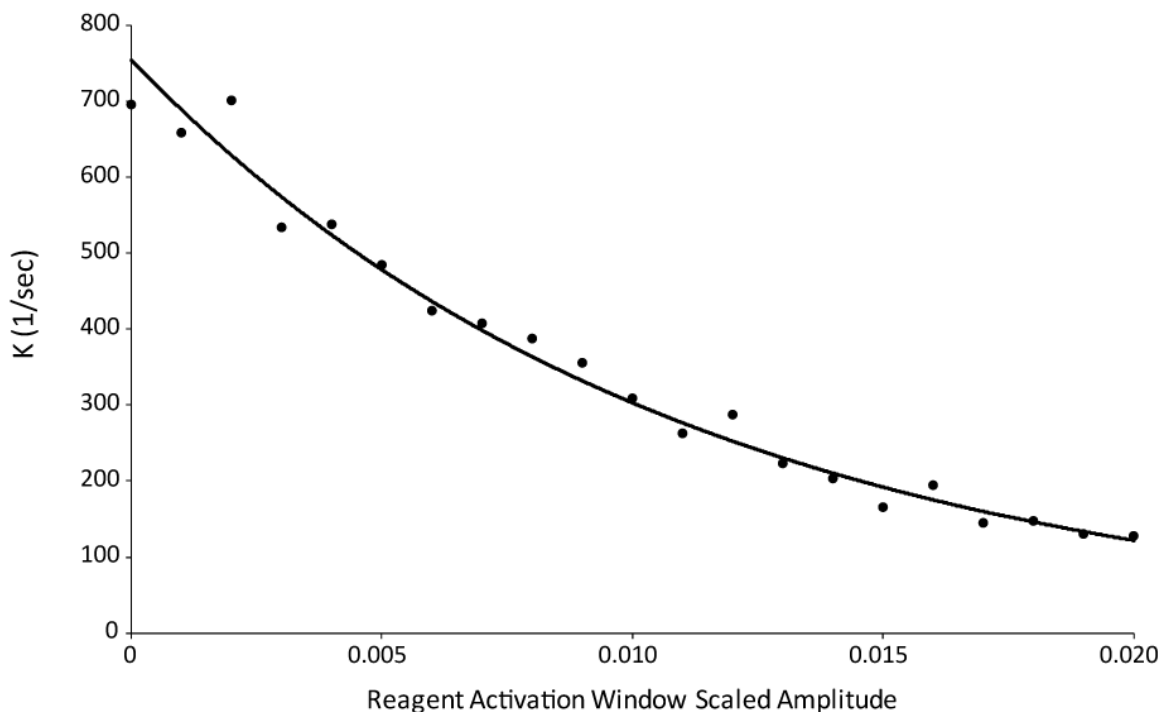
Figure 3.9. Effect of reagent activation amplitude on experimental ion-ion reaction rate constant. Experimentally determined rate constant of the IIPT reaction of the $[M+13H]^{13+}$ charge state of Ubiquitin (m/z 660) with perfluoromethyldecalin reagent anions as a function of amplitude (arbitrary units) of the frequency components for kinetic excitation of the reagent. The Mathieu q of the reagent (m/z 512) was set to 0.55.

While excitation of the target window slows the reaction for only those ions in the selected m/z range, excitation of the IIPT reagent will slow the overall reaction. This can be observed by plotting the rate constant (k) of ubiquitin $[M+13H]^{13+}$ (m/z 660) as a function of the reagent excitation amplitude (**Figure 3.9**). With minimal excitation of the reagent, we can reduce the reaction rate by approximately an order of magnitude. Shown in **Figure 3.8B** is the resulting spectrum after 120 ms IIPT of the $[M+26H]^{26+}$ apomyoglobin precursor while exciting ions with an m/z between 1,450 and 2,450 and the reagent anion (512 +/- 20 m/z; perfluoromethyldecalin). When parking during ETD, we determined that it was necessary to reduce the rate of the reaction in order to effectively preserve the desired reaction products. Similarly, slowing the reaction by

activating the reagent resulted in much more effective concentration of signal into a single peak. The time necessary to reach the desired product was increased five-fold due to the reagent activation slowing the IIPT kinetics. However, more than 90% of the product ion signal was concentrated into a single charge state.

3.5.3 Wideband Isolation and Product Target Window

For our method to be effective at isolating and parking proteins across a wide range of unknown masses, it was necessary to ensure that a majority of ions would be selected for IIPT. On average, one of every seven residues is basic and the average residue mass is 110 Da (average m/z ≈ 770). Therefore, to ensure we selected the majority of species eluting at any given time, we used a wide isolation window (400 m/z) centered at 800 m/z. Shown in **Figure 3.10A** is the ESI mass spectrum of ubiquitin (7.6 kDa) and apomyoglobin (17 kDa) at equal concentrations. Note that even at the same concentration, the most abundant charge state of the larger protein (apomyoglobin) is at 30% relative abundance to that of ubiquitin. A majority of the ion current from both proteins was within the selected isolation window (m/z 600-1,000). Isolating a wide m/z range is especially important for the analysis of mixtures of unknown proteins that differ in size as it maximizes the probability that at least one charge state of a protein will be included in analysis.

In addition to the wide isolation window, a wideband waveform that included a range of frequencies was used. This was necessary to park multiple species in a single

experiment. Additionally, it was not necessary to know the precursor identity *a priori*, so

long as one charge state product fell somewhere between 1150 and 2150 m/z. **Figure**

3.10B shows the MS/MS spectrum resulting from 80 ms IIPT of the isolated charge states

from **Figure 3.10A**. Note that although the parked products differ in m/z, both ubiquitin

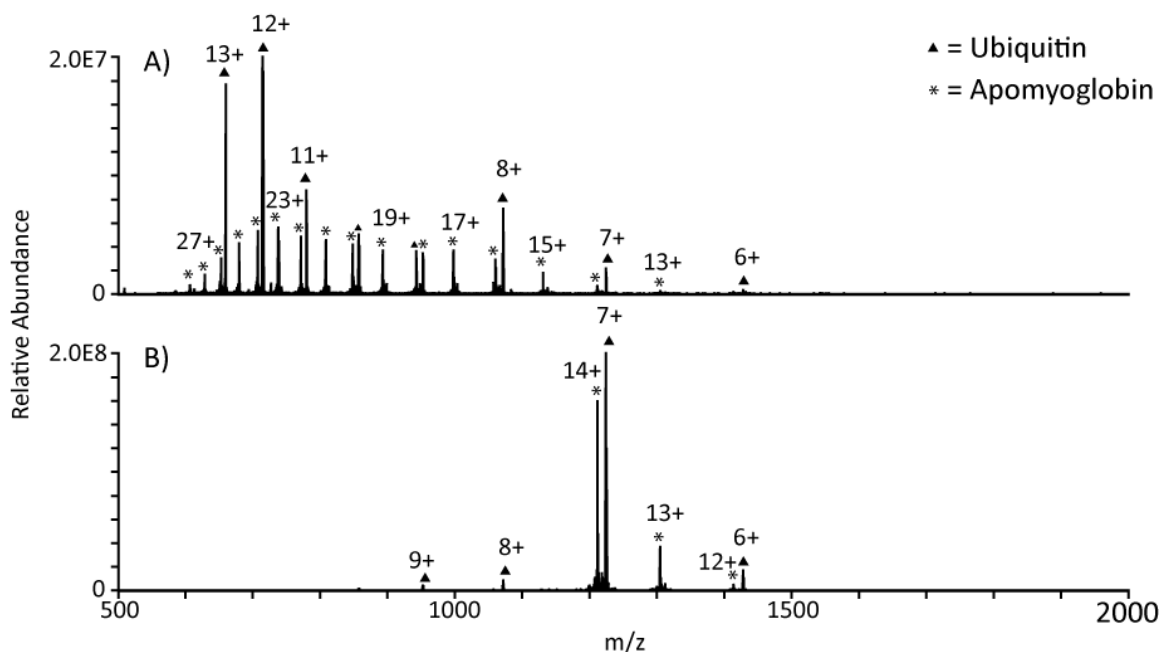(▲) and apomyoglobin (*) are concurrently parked during the experiment.

Figure 3.10. Concurrent isolation, ion-ion proton transfer, and parking of apomyoglobin and ubiquitin. Mass spectra of ubiquitin (▲) and apomyoglobin (*) directly infused at equal concentrations. Electrospray of the solution results in a range of charge states for each species (a). A wide isolation window (600-1000 m/z) includes charge states from both proteins. IIPT for 80 ms while activating products within the shaded area produces primarily one charge state for each protein (b). The result is an order of magnitude increase in the base peak signal.

3.5.4 Ribosomal Proteins Background

Ribosomes are essential for cellular function as they are responsible for

synthesizing proteins from messenger RNA (mRNA) during translation [17]. Intact

ribosomes and their component subunits are differentiated by their sedimentation rate, in

svedburg units (S), by ultracentrifugation. **Figure 3.11** shows the differentiation and makeup of the 80S and 70S ribosomes [18]. Eukaryotic cells contain the 80S ribosome which consists of the large 60S subunit and small 40S subunit, while prokaryotic cells contain the 70S ribosome which consists of the large 50S subunit and small 30S subunit. A majority of the work presented here focuses on 70S ribosomes from *E. coli.* However, some experiments were conducted with human 80S ribosomes. The 50S and 30S subunits are made up of 33 and 22 proteins, respectively. Proteins in the large subunit are labeled L1 to L36, while small subunit proteins are named S1 to S22. Each protein is different with the exception of S20 and L26 which are identical in sequence.

As with most proteins, alterations in ribosomal function may occur due to post-translational modifications (PTMs), changes in the amino acid sequence, or truncations. It is therefore important to identify and, if possible, site-localize modifications. *E. coli* ribosomes contain multiple modifications that are well documented [19]. Six ribosomal proteins are methylated (S11, L3, L11, L7/L12, L16, and L33), three are acetylated (S5, S18, and L7), and one is methylthiolated (S12). Ribosomal protein S6 is polyglutamylated on the C-terminus. Additionally, ribosomal protein S7 is truncated in B strain *E. coli*. In some cases, the absence of the specified modification has no effect. For example, ribosomes with non-acetylated S18 are just as viable as the acetylated form [20]. However, this is not always the case. Ribosomal proteins L7 and L12 have identical amino acid sequences. The only difference between the two species is the presence of an acetyl group on the α-N atom of Ser1. The non-acetylated form is referred to as L12,

while the acetylated species is L7, hence the nomenclature L7/L12. Whereas other

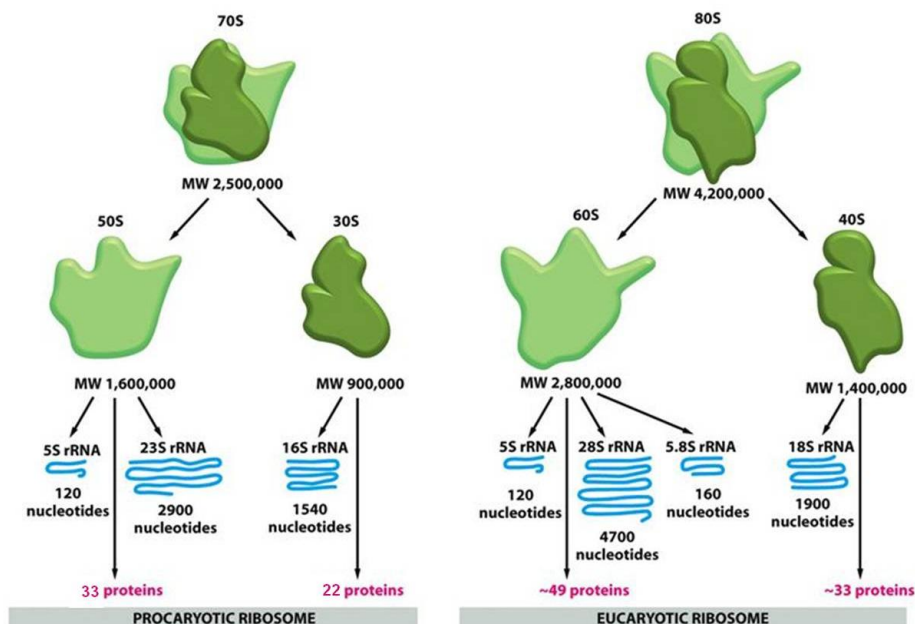ribosomal protein modifications are always observed, L12 is partially acetylated.

Figure 3.11. Structure of Prokaryotic (70S) and Eukaryotic (80S) Ribosomes. The 70S ribosome is made up of one small 30S subunit, comprised of 22 proteins, and one large 50S subunit, comprised of 33 proteins. The 80S ribosome is made up of one small 40S subunit, comprised of 33 proteins, and one large 60S subunit, comprised of 49 proteins. Prokaryotic ribosomes are more conserved than their eukaryotic counterparts. Ribosomal RNA (rRNA) is also present in both 70S and 80S ribosomes. This image was modified from [18].

3.5.5 Identification of Human Ribosomal Protein Mixture Using IIPT and ETD

To demonstrate the utility of parking during IIPT in simplifying the process of

identifying proteins in a complex biological mixture we investigated a mixture of

ribosomal proteins purified from human HepG2 cells. Ribosomal proteins provided a

model sample that was complex enough to illustrate the advantages of our method, yet

simple enough to enable us to monitor the accuracy of positive identifications. Two pmol

total protein was loaded onto a reverse-phase HPLC column and separated over a two and

a half hour gradient. After the initial MS scan (Scan 1), a 400 m/z window centered on

800 m/z was isolated (Scan 2) and 150 ms IIPT was performed (Scan 3). During the IIPT

reaction, a waveform was applied to park all products above 1,150 m/z. The same 400

m/z window centered on 800 m/z was again isolated and 20 ms ETD followed by 40 ms

IIPT with no parking waveform was performed (Scan 4). Because the same window was

isolated for scans 2, 3, and 4 the same proteins were presumably present in all scans. In

scan 2, the proteins were intact and in multiple charge states as produced by ESI of the

eluting species. Scan 3 contained the same intact proteins that had been charge-reduced

by IIPT and combined into a single charge state by parallel ion parking. Scan 4 contained

fragment ions resulting from ETD of all proteins in the window. The ETD fragments

were then charge-reduced by IIPT to their 1+ charge states

The benefits offered by parallel ion parking are easily seen when comparing MS1

spectra to parked spectra. **Figure 3.12A and B** show single scan spectra of scan 1 and

scan 3 respectively as two proteins, 40S S15 (★) and 60S L23a (●), coelute. The

simplicity of the parked spectrum is readily apparent, as the ion current from each protein

has been combined into a single dominant peak. A small amount of the ions continued to

react with the IIPT reagent to produce species that were higher in m/z. This phenomenon

was observed throughout our experiments, especially for highly charged (>10+) proteins,

and the negative effect was minimal. Although parallel ion parking dramatically

decreases the likelihood that a reaction will occur, the reagent and product ion clouds will

still pass through one another. For this reason, we believe some degree of undesired

reaction is unavoidable. An additional benefit of our method is illustrated in panels **3.12C**

and **D**. These are zoomed in portions of the respective spectra to show the dominant peak

in each. Both are charge states of ribosomal protein 40S S15. After IIPT and parking, the

signal-to-noise ratio of the protein signal was increased by a factor of 20. In addition, the

isotopic envelope of the peak in panel **D** is much easier to identify. This is important

when determining the exact intact mass of a protein and will be discussed in more detail
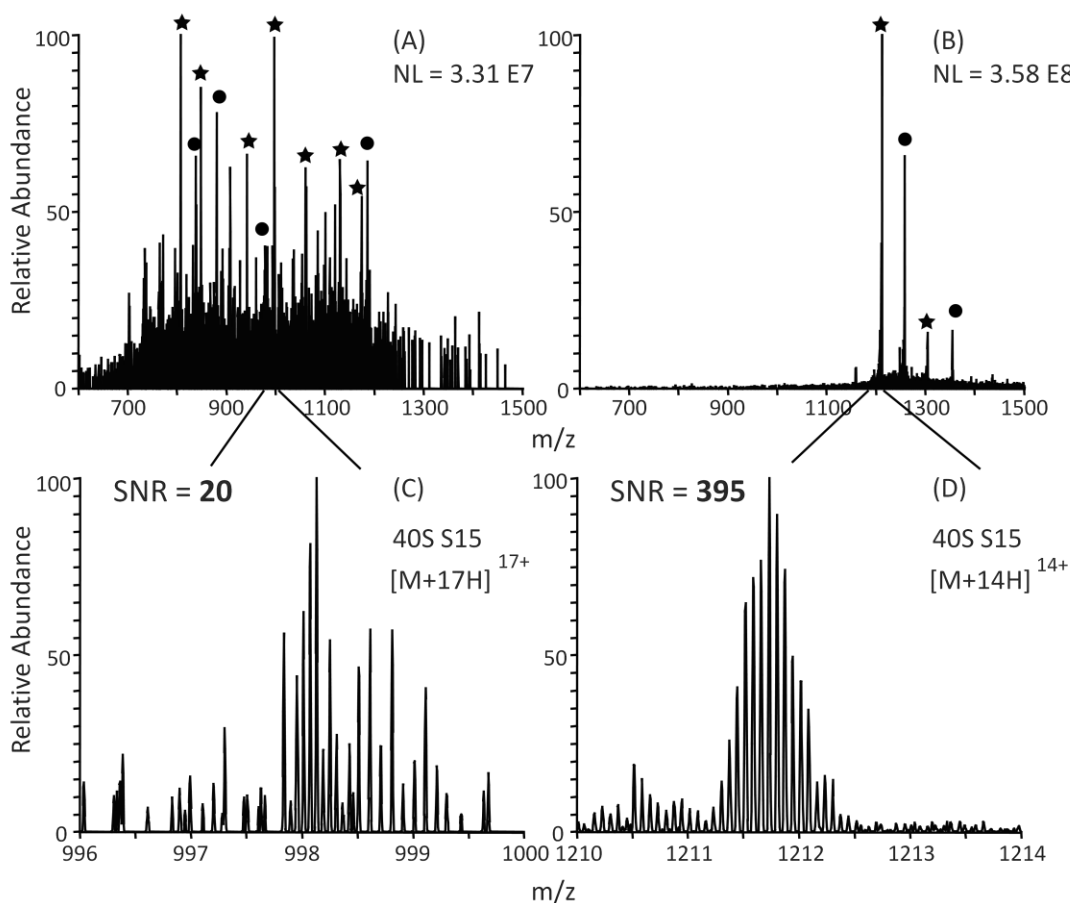
in a later section.

Figure 3.12. Online IIPT and parallel ion parking of ribosomal proteins 40S S15 and 60S L23a.
The single scan ESI spectrum (A) of 40S S15 (★) and 60S L23a(●) is simplified to a few charge state
peaks after 100 ms IIPT with the parking waveform applied (B). The signal-to-noise ratio of the most
abundant charge state of ribosomal protein 40S S15 (C and D) is increased by a factor of almost 20.

Previous efforts by our laboratory to identify intact ribosomal proteins by HPLC-MS/MS yielded little success because data were acquired on an instrument that lacked high-resolution mass analysis capabilities [21]. It was not possible to determine the charge state of fragment ions produced by ETD. To address this problem, IIPT reaction times were lengthened to force all observed fragments into their 1+ charge states. Given the 2,000 m/z limit of the instrument and the average amino acid mass of 110 Da, sequence ion coverage was limited to the first approximately 17 amino acids on each end of the protein. This was sufficient to confidently identify the protein but did not allow for site-localization of any PTMs beyond those residues.

We attempted a similar approach but did so for different reasons. The current work was conducted on an instrument capable of high resolution scans, which enabled the differentiation of fragment charge states. However, initial results showed limited fragmentation of charge-reduced proteins by ETD. A majority of the ion signal in the ETD spectra were products of ETnoD, providing no sequence informative fragments for identification. Several hypotheses have been proposed to explain the lack of fragmentation of low charge-density precursors and work is still ongoing to deduce the mechanism of ETnoD [22]. To maximize the charge of precursors fragmented by ETD, the same window (400 m/z centered on 800 m/z) was isolated for fragmentation. The resulting spectra contained fragments from all proteins eluting at that time. To reduce the complexity of the fragment ion spectra, IIPT was employed to charge-reduce the fragments to the 1+ charge states. As in the work discussed above, sequence ion coverage

was therefore limited to fragments near the N- and C-termini. However, the extended

high mass range of our instrument (4,000 m/z) allowed us to extend our theoretical

coverage limit from approximately 17 residues to approximately 34 residues. Shown in

Figure 3.13 is the MS/MS spectrum resulting from ETD/IIPT of all species shown in

Figure 3.12A. The spectrum contains fragments from both 40S S15 and 60S L23a,

primarily in the 1+ charge state. Confident identification of each protein is easily

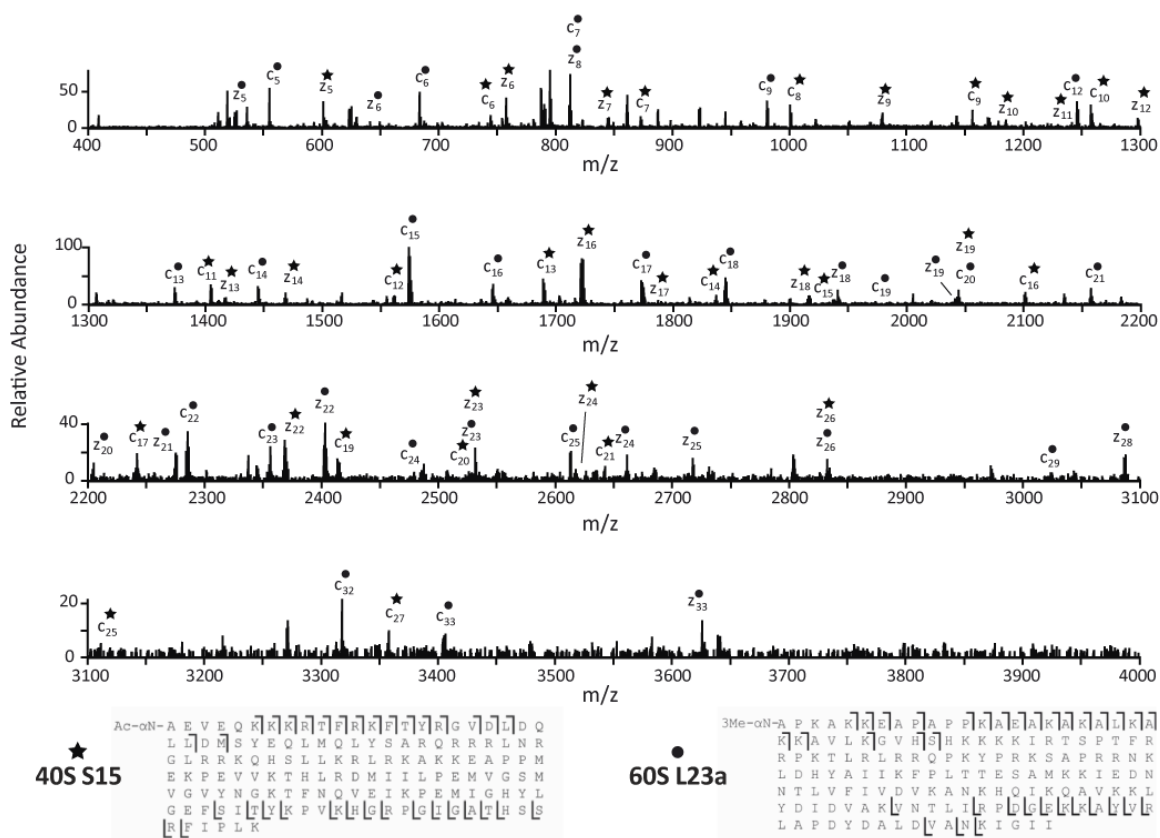achieved as extensive sequence coverage is obtained near the protein termini.

Figure 3.13. Mixed ETD/IIPT Spectrum of 40S S15 and 60S L23a. MS/MS spectrum exhibiting fragment ions from each of two ribosomal proteins as they co-elute with one another. Fragment ions observed are sufficient to identify both proteins by the first approximately 34 amino acids.

Conclusions

While this method appeared to be successful in allowing the identification of co-eluting proteins that were present at relatively similar abundances, only 33 ribosomal proteins could be identified out of the 81 that were expected to be in the sample. Several factors led us to reassess the sample being analyzed and the method of precursor selection and fragmentation. First, the sample was prepared for alternative reasons and had been stored for an extended period of time in 8 M urea. Although the sample was frozen, several species were observed to be multiply carbamylated, a chemical modification known to be introduced by urea storage [23]. This split the abundance of each protein between several modified forms. It is possible that this was the reason a majority of expected species were not observed. For future experiments, *Escherichia coli* ribosomes were purchased so that the preparation and storage of the sample could be controlled. Second, all commercial search programs, including ProSightPC™ require a precursor mass associated with each fragment ion spectrum. Because the selected isolation window for all MS/MS spectra was centered on 800 m/z, this was the precursor m/z used by the search algorithm to determine the protein's identity. A custom program was designed in-house that searched spectra for fragment ion masses independent of intact protein mass. This program was successful in identifying several previously unobserved species, but it was not automated and still required a large time commitment to manually verify results. This was partially due to the fact that a majority of the results returned by the program were false positives. Third, fragments from more highly abundant species dominated the fragment ion spectra, meaning that fragments from proteins that were lower in relative

abundance than approximately 50% were not observed. Future experiments utilized data-dependent selection of each individual protein after IIPT to ensure the observation of relevant fragments. Finally, limiting the sequence informative ion coverage to the N- and C-termini of the protein limited the number of post-translational modifications that could be identified and site-localized. This problem was partially solved by individual selection of proteins for fragmentation. However, as discussed above, ETD of precursors with a low charge density does not result in extensive fragmentation and was therefore not explored. Instead, higher energy collisional dissociation (HCD) was used because the effectiveness of fragmentation is largely independent of the precursor's charge when using this method.

3.5.6 Identification of Escherichia Coli Ribosomal Protein Mixture Using IIPT Parking and Higher-energy Colissional Dissociation (HCD)

Further studies of IIPT and parallel ion parking for the identification of proteins in a complex mixture utilized *E. coli* ribosomes as the model sample. The component proteins of the 70S ribosome provide a sample mixture of 54 proteins that range in mass from 4.5 to 61 kDa. This was ideal in that the sample was sufficiently complex, allowing us to demonstrate the utility of this method on a wide range of protein masses. HPLC-MS/MS analysis was performed over a one hour gradient using only 25 femtomoles of each protein.
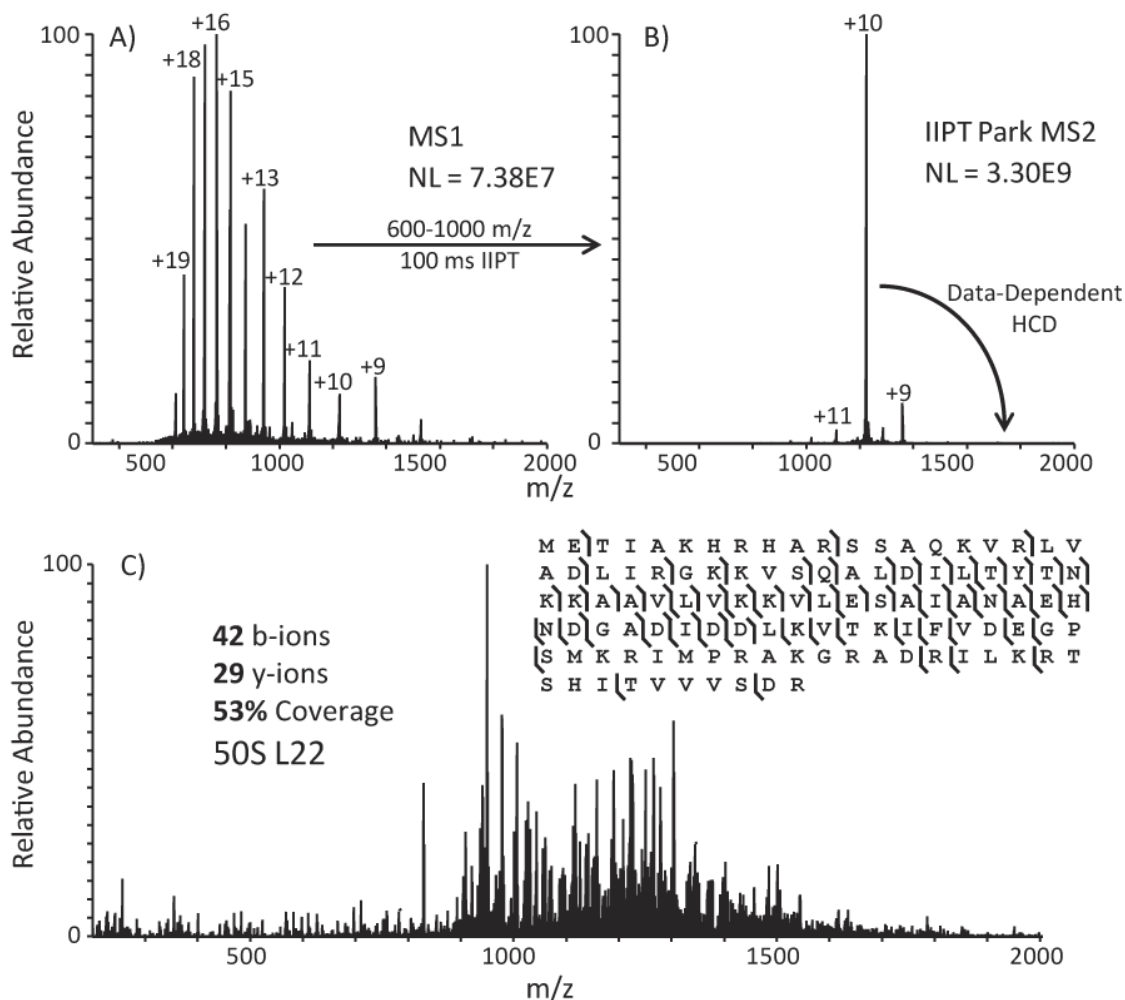
Figure 3.14. Automated IIPT and parallel ion parking followed by data-dependent HCD of ribosomal protein 50S L22. Ribosomal protein 50S L22 MS1 spectrum (A) is simplified and signal is maximized by IIPT and parallel ion parking to a primary charge state (B). From the simplified MS2, data-dependent HCD fragmentation provides sequence information (C) for unambiguous identification of proteins in the mixture.

Shown in **Figure 3.14A** is the MS1 spectrum resulting from ESI of a single protein eluting during the LC-MS gradient. After wideband isolation and IIPT with parallel ion parking (**Figure 3.14B**), the protein signal is enhanced. The simplified spectrum served as the survey scan from which proteins were selected for fragmentation by HCD. **Figure 3.14C** shows the fragment ion spectrum produced by HCD of the

[M+10H]$^{10+}$ charge state of 50S L22. The sequence coverage obtained is sufficient to identify the protein. In addition, sequence coverage extends beyond the limits of the previous experiment, providing extensive overlap near the middle of the protein.

An example which highlights the true benefit of our method is demonstrated in **Figure 3.15**. At least two different proteins are observed in the ESI spectrum. However, once the spectrum is simplified by IIPT and ion parking (**Figures 3.15B and 3.15C**), we see that there are actually four proteins eluting. The two observed in **Figure 3.15A** are identified as 30S S13 (●) and 50S L29 (▲). A third protein, 50S L18 (■), is near the end of its elution profile and can be seen as the major peak eluting earlier in the gradient. However, the fourth protein, 30S S3 (★), is eluting at the same time as S13 and L29.

The inset of **Figure 3.15A** shows an isotopic envelope from the ESI mass spectrum with an m/z that corresponds to the [M+26H]$^{26+}$ charge state of S3. However, the isotopic distribution is very poor, making estimation of the m/z of the most abundant isotope nearly impossible. After IIPT with parking, the signal attributed to S3 has been combined into just two charge states. The inset of **Figure 3.15B** shows the [M+22H]$^{22+}$ charge state of S3 after this process. The S/N of the protein increased almost 30 fold from 7.5 to 211.8. The isotopic distribution has also dramatically improved, making intact mass determination much more accurate. Further, the protein is now the fourth most abundant species in the spectrum from which data-dependent fragmentation is determined, increasing the likelihood that it will be selected for sequence interrogation.
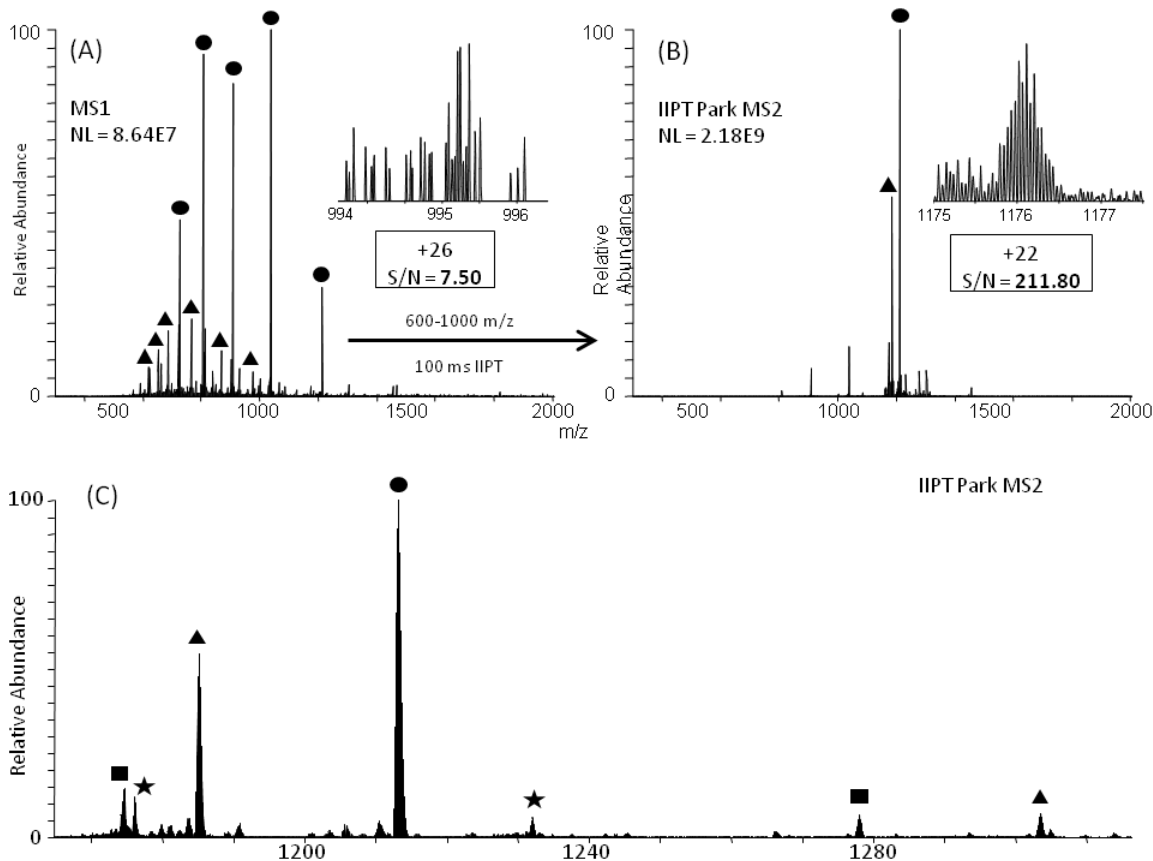
Figure 3.15. IIPT with parallel ion parking MS/MS spectrum of several ribosomal proteins. Panel (A) shows a single scan MS1 spectrum in which two proteins are readily visible. Panel (B) is the MS/MS spectrum resulting from isolating a 400 m/z window centered at 800 m/z and performing 100 ms IIPT while parking from m/z 1150-2150. Panel (C) is a zoomed in portion of the IIPT products showing four identifiable proteins (● = 30S S13, ▲ = 50S L29, ■ = 50S L18, ★ = 30S S3). Insets in panels (A) and (B) are the most abundant charge states of 30S S3 in the appropriate scan and illustrate an almost 30-fold increase in S/N realized by parking and IIPT.

As discussed earlier, in order to confidently determine the identity of a protein, we need two pieces of information: The mass of the intact protein and several unique fragment ions resulting from fragmentation of the protein. The monoisotopic mass of peptides is given as the mass of the ^{12}C isotope. That is, the mass of molecules whose atoms are all stable isotopes that are of the lowest mass (^{12}C, ^{16}O, ^{14}N, and ^{32}S). However, as proteins get larger, the probability increases that the molecule will contain at least one

higher mass isotope. As a rule of thumb, the ^{13}C peak of proteins and peptides becomes the most abundant peak at about 1,600 Da, based on the natural abundance of stable isotopes of organic elements. At very high masses, the ^{12}C peak may not be seen at all. Therefore, it was more practical to use the most abundant isotope of the protein as a means for determining intact mass. Isotopic modeling using Isotope Pattern Calculator™ (IPC) software from Pacific Northwest National Laboratories (PNNL) was used to determine the expected m/z and relative abundance of each isotope, given the protein sequence and desired charge state. We then compared the observed m/z of the most abundant isotope of the intact protein to that predicted by IPC. For example, ribosomal protein 50S L18 had a theoretical monoisotopic mass of 12,762.9393. However, the ^{12}C peak was expected to be present at less than 1% relative abundance. At the 10+ charge state the most abundant isotope was expected to have an m/z of 1,277.9023. The observed m/z of the most abundant isotope was 1277.9082, a mass difference of 4.62 ppm. The observation of 21 b-ions and 29 y-ions confirmed the identity of the protein as ribosomal protein 50S L18

The IIPT parked spectrum of ribosomal proteins L7 and L12 clearly demonstrates the advantages offered by analyzing intact proteins. L7 and L12 are encoded by the same gene. However, L7 is post-translationally acetylated on the N-terminus while L12 is not. In addition, ribosomal protein L7/L12 has been previously reported to be 50% monomethylated at K82. Further complicating matters, a modification that constituted a mass shift of 62 Da was observed. All of these modifications existed by themselves as

well as in combination with one another, resulting in eight separate peaks as seen in

Figure 3.16. Digesting the proteoforms would have likely resulted in the observation of

each of the modifications on separate peptides. However, we would not be able to

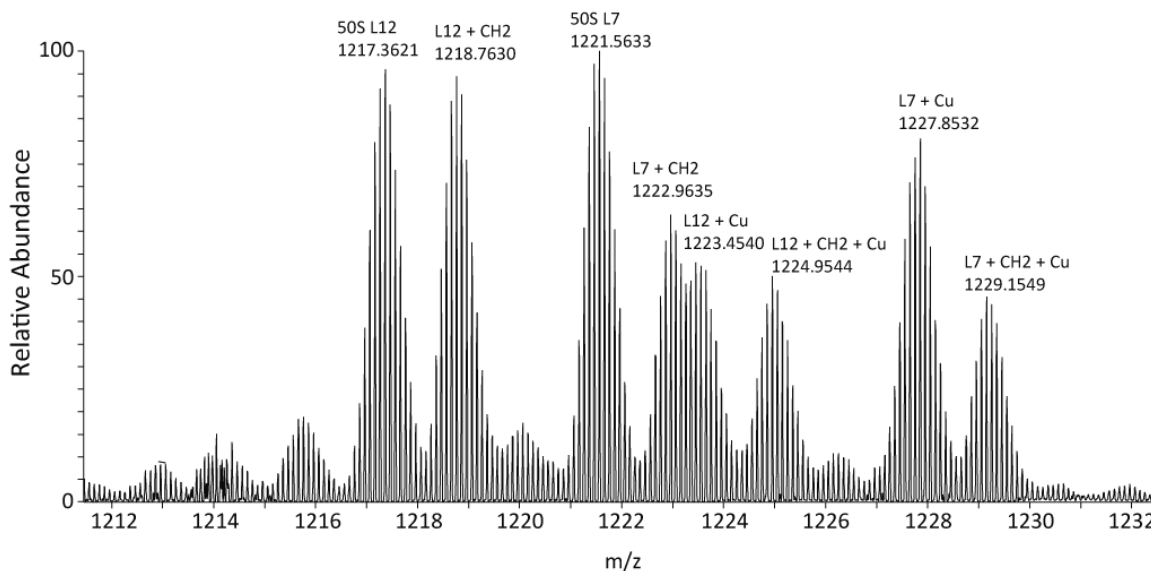conclude which modification existed in context to one another on the protein.

Figure 3.16. Intact proteoforms of ribosomal protein 50S L7/L12. MS/MS spectrum of all
proteoforms of ribosomal protein L7/L12 after 100 ms IIPT with the parking waveform applied. This
spectrum is the result of averaging 50 MS/MS spectra across the elution profile of all eight proteoforms.

Investigation of the fragment ion spectrum resulting from HCD of

monomethylated ribosomal protein L12 provided several sequence determinate ions

containing a mass shift of 14.016 Da. However, a 12 residue gap in sequence coverage

left several candidates as the source of the mass shift. In addition to a methylation, a shift

of 14.016 Da can be indicative of several amino acid substitutions, two of which are

possibilities within the 12 amino acid gap in coverage. There are therefore seven

candidates as the source of the 14.016 Da mass shift: V72X, G74A, G77A, G79A, or

methylation of K71, R74, or K82. Prior reports of a highly conserved methylation of K82

make that the most likely source of the 14 Da shift, although there is not enough evidence in this experiment to reach that conclusion [24].

We next looked to confirm the presence of an acetylation on the N-terminus of the protein. The peak at 1,221.5633 m/z exhibited a mass shift of +42.01 at the most abundant isotope. As discussed above, it is difficult to determine the precise mass difference introduced by chemical modifications by looking at only the intact mass. In addition to an acetyl group a mass shift of approximately 42 Da can instead be indicative of a trimethylation. These two modifications differ in mass by only 0.0364 Da (Acetylation = +42.0106 Da; Trimethylation = +42.0470 Da). At the 10+ charge state observed in this experiment, a mass resolution of approximately 330,000 is necessary to observe this difference. This experiment utilized a resolution of 120,000 at 400 m/z. At 1200 m/z the theoretical resolution is approximately 70,000. It was therefore necessary to look at the fragment ions to confirm the identity and site of the modification. Several N-terminal fragments were observed that contained a mass shift of 42 Da. The b_1 ion fell below our low mass limit of 200 m/z and was therefore not observed. We looked at the b_2 ion to determine the exact mass of the modification. The theoretical mass of the b_2 ion with acetylation or trimethylation is 243.1340 and 243.1705 respectively. The observed mass of the b_2 ion was 243.1342 (ppm error, acetylation = +0.8 ppm; trimethylation = -150 ppm), allowing us to conclude that the N-terminus was post-translationally acetylated.

The final modification of L7/L12 was the addition of approximately 62 Da. Inspection of the fragment ion spectra allowed us to site-localize the modification to Asp5. The fragment ions containing the modification enabled us to determine that the exact mass of the shift was 61.92 Da. This corresponds to the replacement of a proton with a copper atom. Copper differs from common biological elements in that its two stable isotopes differ in mass by 2 Da. Additionally, the natural abundance of ^{65}Cu is high when compared to other elements, making up 31% of all copper atoms. This helped us confirm the presence of copper in mass spectrometry in that we expect to see a shift in the relative abundance of isotopes for fragments that contain copper. Ribosomal protein L7/L12 has not previously been reported to coordinate with copper. However, the biological implications of this finding are beyond the scope of this project.

Intact ribosomal protein L15 was concentrated into the $[M+12H]^{12+}$ charge state after IIPT and parking. The m/z of the intact form was 1,248.2067, a mass shift of approximately -13.95 Da from the theoretical m/z of 1,249.3696. This mass corresponds to a single amino acid mutation of aspartic acid to threonine. The b_{76}^{7+} fragment ion was observed with no mass shift. Fragment ions b_{81}^{8+} through b_{90}^{8+} were also observed, each containing a mass shift of approximately -13.98 Da. We were therefore able to conclude that a single amino acid mutation, D81T, was present in ribosomal protein 50S L15. A search of the literature showed no previous reports of this or any mutation of L15.

The theoretical monoisotopic mass of 30S S18 is 8,850.1835 Da. The amino acid

sequence includes one cysteine, which was carbamidomethylated (+57.6515 Da), meaning the expected theoretical monoisotopic mass was 8,907.8350 Da. The m/z range selected for this experiment was 1,150-2,150 m/z. This means that the primary intact IIPT product that we searched for was expected to be just above 1,150 m/z. In the case of 30S S18, the intact protein with 8 protons remaining was expected to have an m/z of approximately 1,114. This is below the product window, meaning that this product will continue to react. With 7 protons, the theoretical most abundant isotope was 1,273.1274, which fell within the parking product window. However, this m/z was not observed, suggesting that the protein may be modified and differ in mass. Therefore, an alternative method of finding the protein was employed.

MS/MS spectra were searched for characteristic fragment ions from each end of the protein. A search for the first five b-type and y-type fragment ions showed a candidate protein eluting at 52 minutes. The precursor m/z was 1,280.1337, a mass difference of 42.0441 Da, suggesting that 30S S18 is acetylated. Further investigation of the MS/MS spectrum and the observation of mass shifts in n-terminal fragment ions from b_2 confirmed that the acetylation occurred on the n-terminus of the protein. While N-terminal acetylation is common in eukaryotes [25], only four *E. coli* proteins are known to be n-terminally acetylated. Three of these are ribosomal proteins: S18, S5 and L7.

Ribosomal protein 50S L11 was more difficult to identify. Modifications near the N-terminus meant a search for the first five fragment ions was unsuccessful. In addition,

the C-terminus of the protein did not contain a basic residue until Arg134. This meant

that smaller fragments near the C-terminus held no charge and were unobserved.

Therefore, identifying 50S L11 in the elution gradient required a search for fragment ions

y_{10}, y_{11}, y_{12}, and y_{13}. The candidate protein exhibited a mass shift of +126.21 Da from the

theoretical mass. While this mass did not correspond to any known modification, any

combination of three acetyl (42.0367 Da) or trimethyl (42.0797 Da) groups could have

accounted for the mass shift.

Observation of the b_2 ion at 242.1870 narrowed down the site of one acetylation

or trimethylation to the α-N-terminus of Ala1 or the side-chain of Lys2. Calculation of

the mass error for each of the fragments containing the modification resulted in the

following values: b_2 + acetyl = +152 ppm; b_2 + trimethyl = +2.06 ppm. We therefore

concluded that ribosomal protein 50S L11 is trimethylated at either the N-terminus or at

Lys2. Dognin and Wittman-Liebold reported α-N-terminal trimethylation of L11 by

edman degredation in 1980 [26]. However this experiment did not yield sufficient

evidence to confirm the site of modification. The second site of trimethylation, Lys3, was

confirmed by the observation of the b_8^{2+} ion at 486.8164 m/z. The calculated mass error

of this fragment is +1.23 ppm. The third site of modification can be narrowed to one of

three lysine residues: Lys39, Lys44, or Lys50. No fragments are observed between b_{25}

and b_{50} to narrow down the trimethylated amino acid any further.

Intact 50S L33 exhibited a mass shift of +14.04 from the theoretical mass at the

$[M+5H]^{5+}$ charge state, suggesting the presence of a methyl group. The presence of the b_2 ion at 214.1557 m/z narrowed down the possible site of modification to either the N-terminus or Lys2. Interestingly, the first two N-terminal residues of L33 (AK) are the same as those for L11, which was also determined to be trimethylated at either the N-terminus or Lys2.

Intact 30S S11 exhibited a mass shift of +14.06 from the theoretical mass at the $[M+11H]^{11+}$ charge state, suggesting the presence of a methyl group. Once again, the presence of the b_2 ion at 214.1561 m/z narrowed down the possible site of modification to either the N-terminus or Lys2. S11 again follows the pattern observed in L33 and L11, in which the first two amino acids are Ala, Lys, suggesting this may be a target motif for the enzyme responsible for methylation of ribosomal proteins. However, L1, L30, S14 and L5 all contain the same motif, yet no methylation is observed in these cases. L5 has been reported to be acetylated at Lys2 [27], but the data used to reach this conclusion were collected on a low resolution instrument and it was impossible to differentiate trimethylation from acetylation. Thus, it is possible that the modification that was observed was a trimethylation, rather than an acetylation.

Intact 30S S12 exhibited a mass shift of +46.05 from the theoretical mass at the $[M+11H]^{11+}$ charge state. This corresponds to the mass of a methylthiol group (Theoretical mass = 46.09). The fragment ion spectrum was examined to more specifically determine the identity and site-localization of the mass shift. Observation of

N-terminus containing b-ion fragments b_{87}^{8+} and b_{88}^{8+} confirmed the presence of a methylthiol group at Asp88. No y-ions were observed that confirmed this conclusion. A highly abundant fragment was also observed that was determined to be b_{88}^{7+} without a methylthiolation. This suggests that the methylthiol group may be weakly bound to its substrate and was cleaved upon fragmentation. β-methylthiolation of Asp88 in ribosomal protein 30S S12 is a post-translational modification that is highly conserved across many bacterial species. Evidence suggests that methylthiolation of Asp88 in ribosomal protein S12 is catalyzed by the RimO enzyme [28].

Intact 50S L3 exhibited a mass shift of +14.10 from the theoretical mass at the $[M+18H]^{18+}$ charge state, suggesting the addition of a methyl group. Previous literature has reported the observation of a methylated glutamine at Gln150 [29]. However, HCD of the intact precursor only revealed enough information to identify the protein. No fragments large enough to identify the site of methylation present were observed in the current experiment.

Intact 50S L16 exhibited a mass shift of +30.07 from the theoretical mass at the $[M+13H]^{13+}$ charge state. This mass does not correspond to any known chemical modification, but could be indicative of two modifications, a methylation and an oxidation. Examination of the MS/MS spectra proved difficult, as 50S L16 eluted at the same time as 50S L6. The predominant charge state of intact L6 within the parking window was $[M+16H]^{16+}$ with an m/z of 1,177.7736. The predominant charge state of

intact L16 was $[M+13H]^{13+}$ with an m/z of 1,178.7423. An isolation window of 5 m/z was used during this experiment, meaning that when selecting L16 for fragmentation, Intact L6 was selected as well. This is illustrated in **Figure 3.17**. Therefore, fragments from both proteins were present in the fragment ion spectrum. This limited the degree of sequence coverage possible for 50S L16. Ribosomal protein 50S L6 eluted over a longer period of time, and MS/MS spectra from later in the gradient provided clean product ion spectra from which a greater degree of coverage was obtained. However, sequence coverage of L16 was limited to only nine observed fragments. Observation of a b_{10}^{2+} at m/z 650.8874 allowed us to narrow down localization of a methylation to the first ten residues. Previous literature has reported a methylated N-terminal methionine [30]. However, there are two lysine and two arginine residues within the first ten residues which can also be targets for methylation. Literature also reports the hydroxylation of Arg81 [31] which could account for the remaining mass shift of 16 Da, however no fragment ions were observed that could confirm this observation.
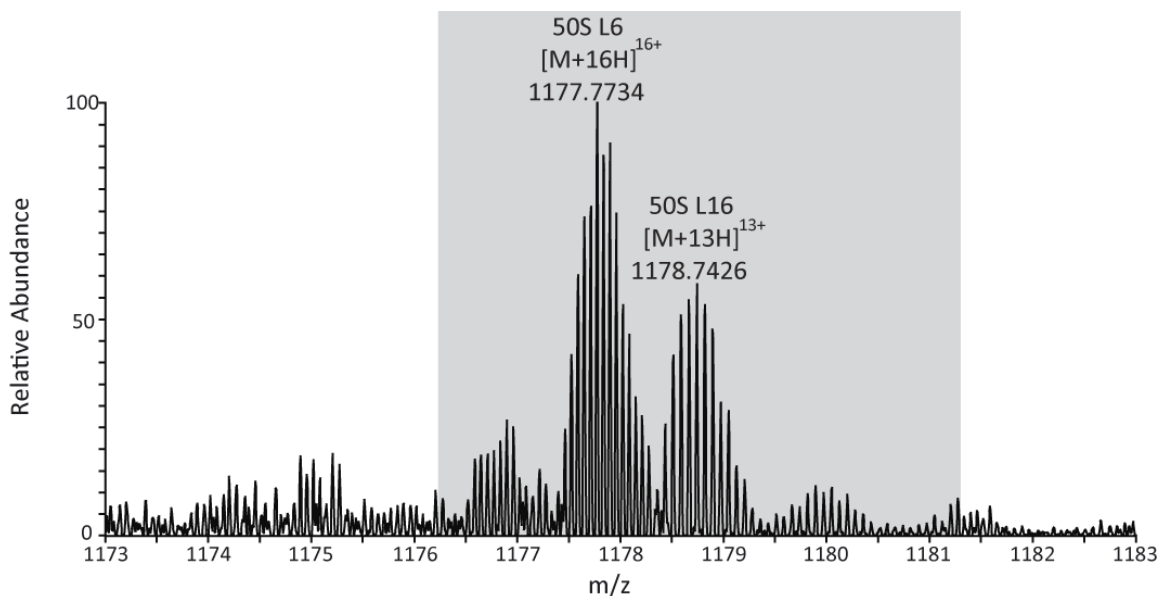
Figure 3.17. Isolation window for fragmentation of 50S L16. The predominant charge state of ribosomal protein 50S L16 produced by IIPT with the parking waveform applied is similar in m/z to that of 50S L6. Isolating 50S L16 with a 5 m/z window (Gray shaded area) does not preclude 50S L6 from fragmentation. The resulting fragment ion spectrum includes fragments from both proteins, reducing the degree of sequence coverage for each protein.

The gene for ribosomal protein 30S S6 encodes for two glutamic acid residues at the C-terminus. The RimK enzyme adds between one and four glutamic acids post-translationally to the protein C-terminus [32]. Identification of 30S S6 was made more difficult by a mass shift of -14.01 in addition to multiple polyglutamylated species. More specifically, four separate proteoforms of S6 were observed, as can be seen in **Figure 3.18**, post-translationally modified with 0, 1, 2, or 3 glutamic acids. As mentioned previously, each proteoform also included a loss of 14 Da. Several single amino acid mutations can be responsible for a mass shift of 14 Da. Therefore, further investigation of the fragment ion spectra was necessary to attempt to confirm the identity of the substitution. Unfortunately, the first basic residue on the C-terminus of the protein is not present until Arg20, and the first C-terminus containing fragment observed is y_{24}^{3+}. This

fragment does contain the loss of 14 Da, meaning the amino acid substitution is present within the first 24 residues at the C-terminus of the protein. The candidate identities and sites of substitution are: E111D, A117G, E119D, T120S, A121G, A124G, E125D, A126G, E131D, or E132D. None of these substitutions have been reported in the literature. Further work is necessary to localize the site and identity of the substitution but the acidic nature of the c-terminus of 30S S6 will make this difficult.
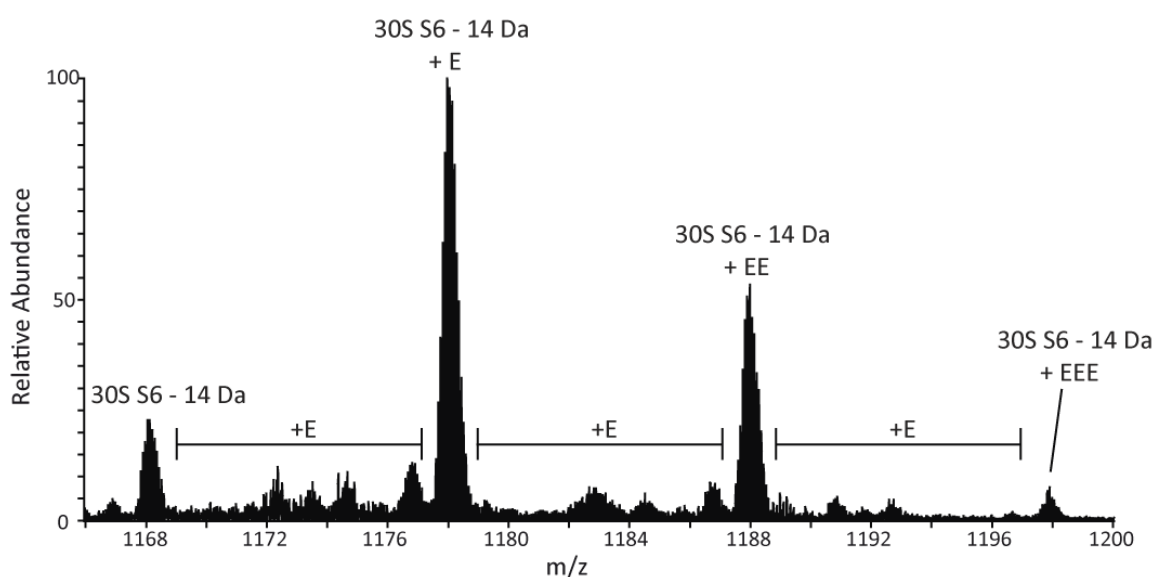
Figure 3.18. Polyglutamylation of 30S S6 C-terminus. MS/MS spectrum of 30S S6 resulting from 100 ms IIPT with the parking waveform applied from m/z 1150-2150. Four forms are present, each differing in mass by one glutamic acid residue. All four forms are 14 Da lower in mass than expected.

Ribosomal protein 30S S7 was only identified by the observation of N-terminal fragments in its MS/MS spectrum. This is because the mass shift exhibited by the intact protein was beyond the 1,000 Da threshold used in the search program. Intact 30S S7 exhibited a mass shift of -2,415 Da. Examination of the literature provided an explanation for this difference. The most common strain of *Escherichia coli* examined for proteomic study is strain K12. Therefore, the only database available for search algorithms is for

strain K12. For these experiments, we started with intact ribosomes from strain B. While a majority of the proteome is conserved between these two strains, ribosomal protein S7 in strain K12 contains a chain of 23 amino acids on the C-terminus that are not present in strain B. Removal of these amino acids from the Fragment Ion Calculator allowed for positive identification of the protein via the observation of the intact species and constituent fragments from both the C-terminus and N-terminus of the protein.

Detailed results are shown in **Table 3.1**. Overall, 53 of 54 ribosomal proteins were identified by manual analysis of intact mass and the presence at least five b or y-type fragment ions resulting from HCD. Of these, 11 contained post-translational modifications. Seven proteins contained multiple modifications or were partially modified, yielding more than one proteoforms. One protein, 30S S3 was isolated with a co-eluting species, 50S L18, yielding an HCD spectrum that included fragment ions from both species. The sole protein that was not identified by HCD fragmentation, 30S S1 (61 kDa) was present by intact mass but was not selected for fragmentation. S1 was observed in ion trap scans performed during the experiment, but not in corresponding Orbitrap scans. This is likely due to the protein's high molecular weight. Larger proteins are often not detected in high resolution analyses because they undergo more collisions with latent gaseous molecules present in the Orbitrap analyzer.

To fully test the utility of our method in identifying unknown proteins in a mixture, a search using ProSight™ was run against the *E. coli* database. The search

yielded 49 assignments. Of these, 46 were ribosomal proteins. The three additional proteins were not proteins from the 70S ribosome but are closely associated with ribosomal function. These identifications were further verified to be correct by manual analysis. In other words, using our methodology and search technique, we obtained no false positive results. For comparison purposes we performed a second experiment that did not utilize IIPT or parallel ion parking. A search of this data yielded only 38 identifications.

Accession #	Protein Name	MW (kDa)	ProSight Identified	Manual ppm Error	Number of Fragment Ions Observed	Modifications
P0A7Q6	L36	4,535	X	3.61	39	
P0A7P5	L34	5,380	X	3.79	13	
P0A7N9	L33	6,372	X	4.23	78	-Met, +Methyl (αN or Lys2)
P0A7N4	L32	6,446	X	4.27	61	-Met
P0AG51	L30	6,542	X	4.75	34	-Met
P0A7M6	L29	7,273	X	4.20	67	
P0A7Q1	L35	7,289	X	3.41	21	-Met
P0A7M9	L31	7,871	X	4.06	44	
P68679	S21	8,500	X	5.65	45	-Met
P0A7T7	S18	8,986	X	3.67	45	-Met, +Acetyl (αN)

Accession # [a]	Protein Name	MW (kDa) [b]	ProSight Identified [c]	Manual ppm Error [d]	Number of Fragment Ions Observed [e]	Modifications [f]
P0A7M2	L28	9,006	X	4.39	6	-Met
P0A7L8	L27	9,124	X	3.63	49	-Met
P0A7T3	S16	9,191	X	4.57	31	
P0A7U7	S20	9,684	X	4.86	61	-Met
P0AG63	S17	9,704	X	4.95	80	-Met
P0ADZ4	S15	10,269	X	5.36	51	-Met
P0A7U3	S19	10,430	X	4.27	94	-Met
P68919	L25	10,693	X	5.47	70	
P0ADZ0	L23	11,199	X	4.50	79	
P60624	L24	11,316	X	4.90	83	-Met
P0AG48	L21	11,564	X	5.76	42	
P0AG59	S14	11,580	X	4.61	47	-Met
P0A7R5	S10	11,736	X	6.90	11	
P61175	L22	12,226	X	5.15	75	
P0A7K2	L12	12,288	X	4.68	66	-Met
	L12	12,288		4.18	23	-Met, +Methyl

Accession # [a]	Protein Name	MW (kDa) [b]	ProSight Identified [c]	Manual ppm Error [d]	Number of Fragment Ions Observed [e]	Modifications [f]
	L7	12,330	X	4.83	58	-Met, +Acet (αN)
	L7	12,330		3.68	5	-Met, +Acet (αN), +Methyl
	L12	12,288		4.33	No MS3	-Met, +Cu
	L12	12,288		3.18	18	-Met, +Me, +Cu
	L7	12,330		2.45	No MS3	-Met, +Acet (αN), +Cu
	L7	12,330		2.68	No MS3	-Met, +Acet (αN), +Methyl, +Cu
P0C018	L18	12,770	X	4.62	50	
P0A7S9	S13	13,099	X	5.15	48	-Met
P0A7K6	L19	13,133	X	6.26	33	-Met
P0A7L3	L20	13,497		4.03	5	-Met
P0ADY3	L14	13,541	X	3.62	24	
P0A7S3	S12	13,737	X	4.43	28	-Met, +Methylthiol (Asp88)
P0A7R9	S11	13,845		3.34	32	-Met, +Methyl (αN or Lys2)
P0A7W7	S8	14,127	X	6.74	36	-Met
P0AG44	L17	14,365	X	5.57	25	
P0A7X3	S9	14,856	X	5.86	31	-Met

Accession # [a]	Protein Name	MW (kDa) [b]	ProSight Identified [c]	Manual ppm Error [d]	Number of Fragment Ions Observed [e]	Modifications [f]
P0A7J7	L11	14,875	X	4.42	51	-Met, +3Methyl (αN or Lys2), +3Methyl (Lys3), +3Methyl
P02413	L15	14,980	X	3.20	51	D81T
P0ADY7	L16	15,281		3.90	9	+Methyl , +Oxid
P02358	S6	15,703	X			-14 Da, +1x, 2x, 3x Glu (C-term)
P0A7R1	L9	15,769	X	5.85	15	
P0AA10	L13	16,019	X	5.35	22	
P0A7W1	S5	17,603	X	4.89		-Met, -16 Da
P0A7J3	L10	17,712	X	5.27	61	-Met
P0AG55	L6	18,904	X	4.06	53	-Met
P02359	S7	20,019	X	4.32	43	-Met, -AA 157-179
P62399	L5	20,302	X	5.21	42	-Met
P60723	L4	22,087		4.15	12	
P60438	L3	22,244	X	3.80	30	+Methyl
P0A7V8	S4	23,469	X	5.68	17	-Met
P0A7L0	L1	24,730	X	4.31	24	-Met

Accession # [a]	Protein Name	MW (kDa) [b]	ProSight Identified [c]	Manual ppm Error [d]	Number of Fragment Ions Observed [e]	Modifications [f]
P0A7V3	S3	25,983		3.98	8 (Mixed)	-Met
P0A7V0	S2	26,744		4.17	21	-Met
P60422	L2	29,860		5.44	23	-Met
P0AG67	S1	61,158		IT Only	No MS3	

Table 3.1 Escherichia Coli Ribosomal Protein Mixture Results.
[a] Accession number of E coli proteins from the uniprot database.
[b] Common names of 70S E coli ribosomal protein subunits.
[c] Calculated MW (average mass) based on protein sequence obtained from uniprot.
[d] Proteins identified by database search with ProSightPC are denoted by an X.
[e] Mass error in ppm of observed versus theoretical (including PTMs) most abundant charge state and isotope after IIPT and parallel ion parking.
[f] Number of fragment ions (b and y) observed in HCD MS^3 averaged spectrum.
[g] Observed post-translational modifications. Identified modification sites are in parentheses.

3.6 Conclusions

We have demonstrated the ability of IIPT coupled with parallel ion parking to perform gas-phase purification of proteins ionized by ESI. The introduction of a high mass reagent allows us to perform parking at masses throughout the m/z range offered by hybrid ion trap-Orbitrap instruments. In addition, we determined that it is necessary to partially excite the reagent anion as well as the desired product ion to ensure complete parking at a single charge state. We introduced the use of wide isolation windows and broadband parallel ion parking ranges to ensure the selection and parking of a majority of intact proteins, regardless of molecular weight. This allowed us to select and park multiple proteins of different masses concurrently without *a priori* knowledge of their

identity.

An automated HPLC-MS method which utilized parallel ion parking resulted in up to a 30-fold increase in S/N and provided a simplified spectrum from which precursors were selected for fragmentation. A ProSight™ search of the data against the *E. coli* database provided 49 positively identified proteins with no false positive results. A similar experiment in which parking was not employed resulted in only 38 proteins identified. Manual interpretation of the data yielded the positive identification of 53 out of 54 ribosomal proteins in a mixture. The identified proteins ranged in molecular weight from 4.5 to 30 kDa. While it did not yield comprehensive sequence coverage, HCD fragmentation resulted in our ability to site-localize multiple post-translational modifications. Our ability to differentiate eight different proteoforms of ribosomal protein L7/L12 highlights the advantages of analyzing proteins as intact molecules. This was accomplished over a one hour HPLC gradient and used only 25 fmol of each protein.

Parallel ion parking offers three distinct advantages over intact protein MS techniques that do not utilize the technique: 1) Increased signal-to-noise ratio of each protein increases the likelihood that low-level intact species will be observed, 2) simplified spectra from which data-dependent selections occur increases the likelihood that each protein will be selected for fragmentation, and 3) higher quality isotopic distributions arising from increased ion populations in a single charge state allow for more confident assignment of a protein's intact mass. This methodology offers an

unrivaled tool for observing and identifying low abundance proteins, furthering the understanding of complex biological systems.

3.7 References

1. Catherman AD, Skinner OS, Kelleher NL. Top down proteomics: facts and perspectives. *Biochem Biophys Res Commun.* **2014,** *445,* 683-693.

2. Fornelli L, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, Tsybin YO (2012). Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation orbitrap FTMS. *Mol Cell Prot,* 11(12): 1758-1767.

3. Fernandez de la Mora J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal Chim Acta*. **2000,** *406*, 93-104.

4. Coon JJ, Ueberheide B, Syka JEP, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci*. **2005,** *102*, 9463-9468.

5. McLuckey SA, Stephenson JL, Jr. Ion/ion chemistry of high-mass multiply charged ions. *Mass Spectrom Rev.* **1998,** *17*, 369-407.

6. Stephenson JL, Jr., McLuckey SA. Ion/ion proton transfer reactions for protein mixture analysis. *Anal Chem*. **1996,** *68*, 4026-4032.

7. McLuckey SA, Stephenson JL, Jr., Asano KG. Ion/ion proton-transfer kinetics: implications for analysis of ions derived from electrospray of protein mixtures. *Anal Chem.* **1998,** *70,* 1198-1202.

8. Chrisman PA, Pitteri SJ, McLuckey SA. Parallel ion parking of protein mixtures. *Anal Chem.* **2006***, 78*, 310-316.

9. Earley L, Anderson LC, Bai DL, Mullen C, Syka JE, English AM, Dunyach JJ, Stafford GC, Jr, Shabanowitz J, Hunt DF, Compton PD. Front-end electron transfer dissociation: A new ionization source. *Anal Chem,* **2013,** *85,* 8385-8390.

10. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA.* **2004***, 101,* 9528-9533.

11. Zhao X, Ryjkov VL, Schuessler HA. Parametric excitations of trapped ions in a linear rf ion trap. *Physical Review A.* **2002,** *66*, 063414.

12. Gunawardena HP, He M, Chrisman PA, Pitteri SJ, Hogan JM, Hodges BD, McLuckey SA. Electron transfer versus proton transfer in gas-phase ion/ion reactions of polyprotonated peptides. *J Am Chem Soc.* **2005,** *127*, 12627-12639.

13. Coon JJ, Syka JEP, Schwartz JC, Shabanowitz J, Hunt DF. Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions. *Int J Mass Spectrom.* **2004,** *236,* 33-42.

14. F2 Chemicals, Ltd. Web. Accessed December 20th, 2016.http://www.f2chemicals. com/perfluoro13dimethylcyclohexane.html

15. F2 Chemicals, Ltd. Web. Accessed December 20th, 2016. http://www.f2chemicals. com/perfluorodecalin.html

16. F2 Chemicals, Ltd. Web. Accessed December 20th, 2016. http://www.f2chemicals.

com/perfluoromethyldecalin.html

17. Doudna JA, Rath VL. Structure and function of the eukaryotic ribosome: the next frontier. *Cell.* **2002,** *109,* 153-156.

18. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell: Reference Edition.* **2008.** Garland Science.

19. Nesterchuck MV, Sergiev PV, Dontsova OA. Posttranslational modifications of ribosomal proteins in escherichia coli. *Acta Naturae.* **2011,** *3,* 22-33.

20. Isono K, Isono S. Ribosomal protein modification is Escherichia coli: studies of a mutant lacking the N-terminal acetylation of protein S18. *Mol Gen Genet.* **1980,** *177,* 645-651

21. Chi A, Bai DL, Geer LY, Shabanowitz J, Hunt DF. Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. *Int J of Mass Spectrom.* **2007,** *259,* 197-203.

22. Ledvina AR, Beauchene NA, McAlister GC, Syka JEP, Schwartz JC, Griep-Raming J, Westphall MS, Coon JJ. Activated-ion ETD (AI-ETD) improves the ability of ETD to identify peptides in a complex mixture. *Anal Chem.* **2010,** *82*, 10068-10074.

23. Stark GR, Stein WH, Moore S. Reactions of the cyanate present in aqueous urea with amino acids and proteins. *J Biol Chem.* **1960,** *235,* 3177-3181.

24. Terhorst C, Moller W. The primary structure of an acidic protein from 50-S ribosomes of Escherichia coli which is involved in GTP hydrolysis dependent on elongation factors G and T. *Eur J Biochem.* **1973,** *34*, 138-152.

25. Polevoda B, Sherman F. The diversity of acetylated proteins. *Genome Biol.* **2002,** *3(5),* 1-6.

26. Dognin MJ, Wittman-Liebold B. Purification and primary structure determination of the N-terminal blocked protein, L11, from Escherichia coli ribosomes. *Eur J Biochem.* **1980,** *112,* 131-151.

27. Zhang J, Sprung R, Pei J, Tan X, Kim S, Zhu H, Liu CF, Grishin NV, Zhao Y. Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. *Mol Cell Proteomics.* **2009,** *8,* 215-225.

28. Strader MB, Costantino N, Elkins CA, Chen CY, Patel I, Makusky AJ, Choy JS, Court DL, Markey SP, Kowalak JA. A Proteomic and Transcriptomic Approach Reveals New Insight Into beta-Methylthiolation of Escherichia coli Ribosomal Protein S12. *Mol Cell Proteomics.* **2011,** *10(3),* M110.005199. DOI: 10.1074/mcp.M110.005199.

29. Muranova TA, Muranov AV, Markova LF, Ovchinnikov YA. The Primary Structure of Ribosomal Protein L3 from Escherichia coli 70S Ribosomes. *FEBS Lett.* **1978,** *96,* 301-305.

30. Brosius J, Chen R. The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes. *FEBS Lett.* **1976,** *68,* 105-109.

31. Ge W, Wolf A, Feng T, Ho C, Sekirnik R, Zayer A, Granatino N, Cockman ME, Loenarz C, Loik ND, Hardy AP, Claridge TDW, Hamed RB, Chowdhury R, Gong L, Robinson CV, Trudgian DC, Jiang M, Mackeen MM, McCullagh JS, Gordiyenko Y, Thalhammer A, Yamamoto A, Yang M, Liu-Yui P, Zhang Z, Schmidt-Zachmann M,

Kessler BM, Ratcliffe PJ, Preston GM, Coleman ML, Schofield CJ. Oxegenase-catalyzed ribosome hydroxylation in prokaryotes and humans. *Nat Chem Biol.* **2012,** *8,* 960-962.

32. Kang WK, Icho T, Isono S, Kitakawa M, Isono K. Characterization of the gene RimK responsible for the addition of glutamic acid residues to the C-terminus of ribosomal protein S6 is Escherichia coli. *Mol Gen Genet.* **1989,** *217*, 281-288.