High Resolution Peptide Tandem Mass Spectrometry: Alternative Charge Carriers in Electron-based Activation Methods and Improved Antibody Sequence Variant Analysis

by

Tao Jiang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2016

Doctoral Committee:

Professor Kristina I. Håkansson, Chair
Professor Philip C. Andrews
Professor Zhan Chen
Associate Professor Brandon T. Ruotolo

Acknowledgement

First I would like to thank my advisor, Dr. Kristina Håkansson, for her mentorship during the past five years. Her enthusiasm and excellence encouraged me to devote myself to mass spectrometry research. With her support, my Ph.D. pursuit has been the most exciting and rewarding one I could have imagined.

I would like to thank my other committee members as well: Dr. Philip Andrews, Dr. Zhan Chen, and Dr. Brandon Ruotolo. Their valuable suggestions and comments contributed to all aspects of my research, as well as the generation of this thesis.

The Håkansson lab has always been filled with people I love to work with. Previous and current lab members offered me great help in scientific research and professional development. I'm grateful for all of you: Hangtian Song, Katie Hersberger, Di Gao, Wendi Hale, Phil McClory, Kevin Ileka, Qingyi Wang, Hye Kyong Kweon, and Nick Borotto.

My research was funded by the Department of Chemistry at the University of Michigan, NSF grants, and by Bristol-Myers Squibb. Parts of the experimental data for Chapter 2 were gathered at the National High Magnetic Field Laboratory. I would like to thank these organizations for providing me with the necessary research facilities and financial support.

Finally, I would like to thank my dearest family and friends, for shaping me into who I am now. My parents, Wenxin Jiang and Jiamei Mao, provided me with unmatchable love and insightful education. They enabled me to explore the greater world and to achieve myself. My wife, Xiaolin Wang, has always been supportive of my efforts. She gives me the warmth of a family in a land thousands of miles away from my home. My friends in Ann Arbor, as well as those in other parts of the United States or in China, offered me help and joy during the time we spent together. I would like to express my deepest gratitude to all of you.

Tao Jiang

May 11, 2016

Ann Arbor, Michigan

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Mass Spectrometry in Peptide and Protein Analysis

Proteomics is a subject aiming at comprehensive investigation of multiple aspects of proteins in biological systems.[1] According to results from the Human Genome Project, there are 20k-25k genes in the human genome that encode proteins.[2] In order to understand biological systems, identities, quantities, modifications, structures and interactions of proteins are all desirable information. Under natural circumstances, some of these proteins exist at low levels,[3] in complicated matrices,[4] in multiple modified forms,[5] or in multiple isoforms.[6] These difficulties and the associated multiple dimensions of information bring challenges towards analytical instruments and protocols for proteomics.

For pharmaceutical proteins, detailed characterization is important in both the R&D and production stages. For example, in the effort to express pharmaceutical proteins in bioreactors using, e.g., Chinese hamster ovary (CHO) cells[7] or *Escherichia coli* (E. coli) cells,[8] produced

protein needs to be characterized to ensure plasmid sequence integrity. In the production phase, maximum impurity levels for pharmaceutical protein products are set by the Food and Drug Administration (FDA) to ensure safety and efficacy of products. Characterization and quantification of these impurities, usually other proteins with erroneous sequence[9] or proteins from host cells,[10] requires analytical protocols to be not only sensitive and robust, but also at high throughput.

Mass spectrometry (MS) is a powerful tool to address protein characterization in both proteomics and pharmaceutical sciences.[11] "Soft" ionization methods, such as electrospray ionization (ESI, see Chapter 1.4) [12] and matrix-assisted laser desorption/ionization (MALDI),[13] can ionize peptides and proteins gently, preventing them from disintegrating. MS analysis relies on measuring mass-to-charge ratios of ions in the gas phase. Coupled with other separation or analytical methods, typically liquid chromatography (LC), modern mass spectrometry instruments and protocols can identify and quantify trace level proteins from various sample sources. Tandem mass spectrometry (MS/MS) can fragment peptide bonds in a predictable manner, thus by measuring fragments generated, sequence and modifications of peptides or proteins can be derived.

Two complementary approaches[14-15] are commonly used for protein analysis by mass spectrometry: bottom-up[16] and top-down.[17] In the bottom-up approach, generally there are three major steps: sample preparation, LCMS/MS experiments, and database searching. In the sample preparation step, protein samples are purified by LC and/or other separation methods, and then digested into peptides. In LC/MS/MS experiments, ESI is commonly used as ion source due to its high compatibility with LC. Database searching can identify peptides, based on their

MS/MS spectra. Peptides unique to a specific protein then reconstitute the information of the corresponding protein. In the top-down approach, proteins are introduced into the mass spectrometer as whole molecules, rather than digested into peptides. MS/MS experiments are performed to fragment the resulting protein molecular ions, and generated fragments are measured to recover information about the protein. Both these two methods provide protein sequence and modification information. Currently, the bottom-up approach is more commonly adopted, due to well-established supporting instrumentation and software, and less amount of sample is required for accurate sequence identification. However, the top-down approach has demonstrated superiority in modification characterization.[17]

1.2 Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS)

1.2.1 Basic Principles of FT-ICR MS

Invented in 1974,[18-19] Fourier transform ion cyclotron resonance (FT-ICR) MS features superior mass accuracy and resolving power among all types of mass spectrometers.[20-21] FT-ICR MS is based on cyclotron motion of charged ions in a confined space with a homogeneous magnetic field. When an ion with mass m, charge z*e (e is the charge of an electron/proton, 1.60217662e-19 Coulomb and z is an integer), and velocity **v** is placed in a magnetic field **B**, it will experience the Lorentz force, designated as **F**:

$$\mathbf{F}=ze\mathbf{v}\times\mathbf{B} \qquad (1)$$

This force will serve as a centripetal force, supporting a circular motion of the ion, in the plane perpendicular to the magnetic field. Designating the angular speed of this circular motion as ω, we have:

$$\mathbf{F}=mv^2/\mathbf{r}=m\omega\mathbf{v} \qquad (2)$$

Combining (1) and (2), we arrive at the cyclotron equation:

$$\omega=\frac{ze\mathbf{B}}{m} \qquad (3)$$

By substituting $f=\omega/2\pi$ (in which f = frequency in Hertz), the cyclotron equation can also be written as:

$$f=\frac{zeB}{2\pi m} \qquad (4)$$

In a homogeneous magnetic field, B is a fixed value. Therefore, cyclotron motion frequency of an ion is dictated by its mass-to-charge ratio (m/z) value. To measure this frequency (and thus the m/z ratio), ions are injected and trapped axially by a DC potential in an ICR cell (Figure 1.1), located at the bore of a superconducting magnet of known magnetic field strength.. Initial cyclotron motions have small radii ($\leqslant$1 mm) due to a well collimated ion beam.[22] However, the phase of this cyclotron motion is random for different ions with identical m/z ratios. By adding an external waveform RF electrical field with the same frequency as the ion cyclotron frequency, resonance absorption causes ions with the associated m/z ratio to acquire unified phase and also amplifies their cyclotron motion radius for detection.

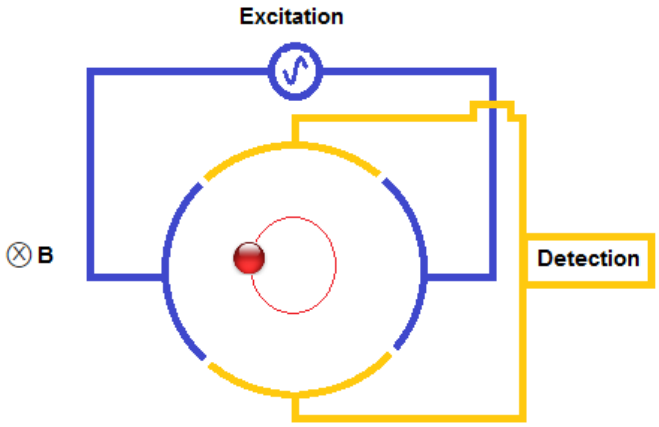


Figure 1.1. Schematic diagram of a cylindrical FT-ICR cell.

An FT-ICR cell is designed to perform both excitation (via the applied external waveform described above) and detection of ions undergoing cyclotron motion. Trapping plates that carry DC electric potentials are employed to confine ion packets (the red ball) axially inside the cell. Plates used to apply the excitation waveform are situated perpendicular to the magnetic field. The frequency of this waveform can be changed rapidly (~50 μs), in order to rapidly excite ions within a broad m/z range. This process is referred to as an r.f. chirp.[23] Excited ions, with larger cyclotron motion radius, generate image current on the detection plates. Such current is amplified and digitally recorded, as the time-domain signal. Fourier transform is applied to transform the time-domain signal into a frequency-domain signal, which can be translated into a mass spectrum, as shown in Figure 1.2.

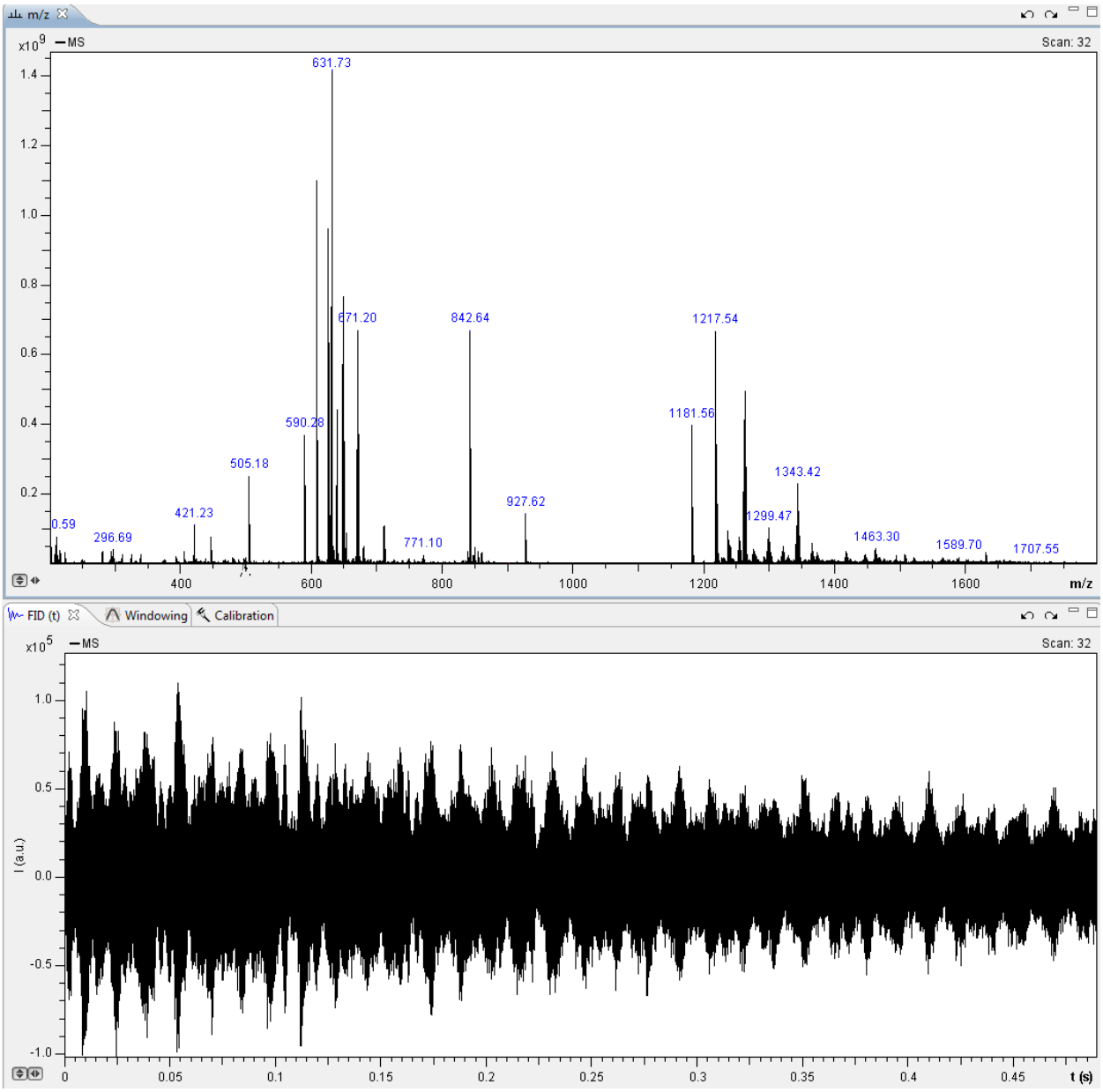


Figure 1.2. Fourier transform–ion cyclotron resonance m/z spectrum obtained by calibrated frequency-to-m/z conversion and time-domain image-current signal from opposed detection electrodes, displayed in Bruker FTMS processing software.

1.2.2 FT-ICR MS Instrumentation

Figure 1.3 shows the schematics of two 7 Tesla FT-ICR MS instruments at the University of Michigan, Apex and Solarix (Bruker Daltonics, Billerica, MA). 7 Tesla homogeneous magnetic

fields are provided by superconducting magnets. Both instruments are equipped with ESI as ion source, quadrupole as ion selection device, hexapole collision cell for external collision activated dissociation (CAD), FT-ICR cell as detector and a hollow cathode behind the FT-ICR cell for electron-based MS/MS experiments. On Solarix, a 10.6 μm CO_2 IR laser (Synrad, Mukilteo, WA) is present for IR heating of ions inside the FT-ICR cell, or for performing infrared multiphoton dissociation (IRMPD). An electron transfer dissociation (ETD) module is also present, enabling ETD and proton transfer reaction (PTR).

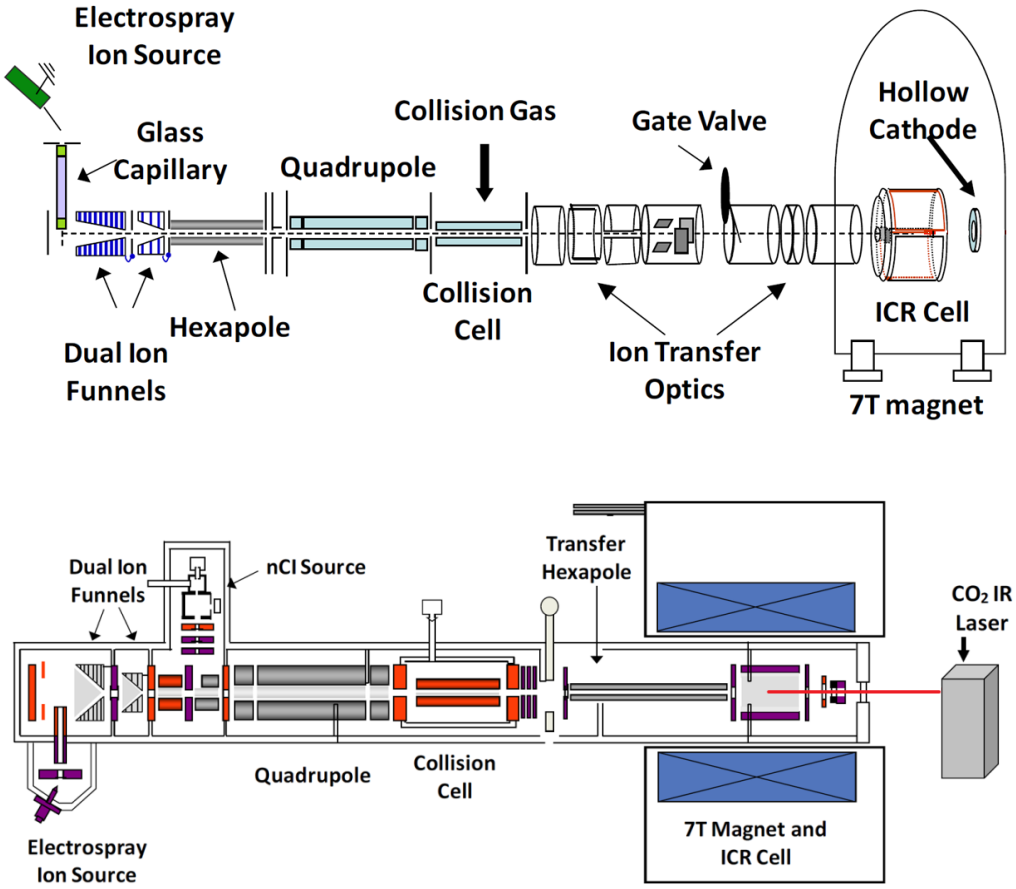


Figure 1.3. Schematic diagrams (Upper: Bruker Apex; Lower: Bruker Solarix) of 7 Tesla FT-ICR mass spectrometers at the University of Michigan.

1.2.3 Advantages and Application of FT-ICR MS

The primary advantages of FT-ICR MS are its ultrahigh resolution and high mass accuracy, allowing analysis of complex samples, such as protein digests[24] and petroleum.[25] With high resolution, the mass difference between different isotopes (e, g., a shift of ^{2}H vs. ^{13}C.) can be discriminated, allowing isotopic fine structure detection for specific elements in biological samples, such as sulfur in disulfide bonds.[20] High resolution can also be useful in the analysis of high-charge-state proteins or protein complexes,[26] for which charge state determination may require discrimination of isotopic signals.

Another advantage is that FT-ICR MS can be a versatile platform for MS/MS experiments. In addition to CAD, ETD and IRMPD, which can be performed on other types of mass spectrometers, FT-ICR can easily implement cathodes for electron-based MS/MS methods, such as electron capture dissociation (ECD) and electron detachment dissociation (EDD). These methods provide complementary information for peptide and protein sequencing compared with more conventional CAD, and are useful for modification detection (see Chapter 1.5). The confining effect of the magnetic field upon free electrons contributes to the success of electron-based MS/MS experiments in FT-ICR MS. In fact, when such experiments are implemented on other mass spectrometers, an external magnetic field is often required.

1.3 Orbitrap Hybrid Mass Spectrometry

1.3.1 Basic Orbitrap Principles

Orbitrap is a mass analyzer based on the Kingdon trap.[27-28] Similar in basic ideas of harmonic excitation and image current detection to the FT-ICR, ions in an orbitrap are placed in

a quadro-logarithmic electric field, instead of a homogeneous magnetic field. In order to generate such an electric field, spindle-shaped and barrel-shaped electrodes are placed coaxially to form an orbitrap (see Figure 1.4). The resulting electrostatic force serves as a centripetal force for circular movement around the orbitrap axis, while motion of ions along the axis is similar to a simple oscillator (analogous to axial motion in FT-ICR MS). Due to the quadro-logarithmic field the angular frequency of the ion axial motion is proportional to the square root of their z/m value, thus this axial frequency can be used for determination of m/z, similar to the circular movement frequency in FT-ICR MS.

1.3.2 Orbitrap Instrumentation

The specific orbitrap instrument used in this dissertation is a Thermo Orbitrap Elite (Thermo-Fisher Scientific, San Jose, CA) at Bristol-Myers Squibb, Bloomsbury, NJ. This instrument features a dual-analyzer system: an orbitrap for high resolution and high mass accuracy detection, and a linear quadrupole ion trap for fast CAD experiments and automated gain control. In addition to CAD in the ion trap, the instrument is capable of performing higher-energy collisional dissociation (HCD) and ETD in the HCD collision cell. This instrument is also equipped with an ESI source, as well as a quadrupole following the ion trap. In order to improve focusing of ion packets during orbitrap injection, a C-trap is placed in front of the orbitrap.[29]

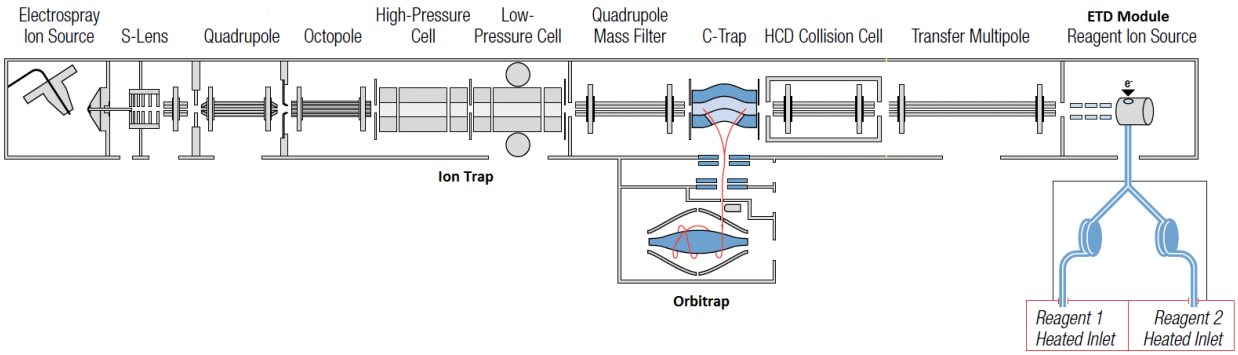


Figure 1.4. Schematic diagrams of a Thermo Orbitrap Elite mass spectrometer.

1.3.3 Advantages and Application of Hybrid Orbitrap MS

The major advantage of ion trap/orbitrap hybrid instruments is their rapid, automated LC/MS/MS functions while simultaneously allowing high mass accuracy precursor ion measurements.[30] In bottom-up proteomics, data-dependent LC/MS/MS is the most commonly applied strategy. For each LC/MS precursor ion spectrum, the top n (usually n = 10 for the current instrument) most abundant ions undergo MS/MS for peptide identification. In this process, high mass accuracy and resolution of the precursor LC/MS spectra reduce false positives and contribute to confident peptide identification. This need is fulfilled by the orbitrap as analyzer for LC/MS spectra. On the other hand, MS/MS experiments need to be fast, so that few peptides are missed in LC/MS/MS. The rapid activation and detection of the linear ion trap provide high-throughput MS/MS spectra with acceptable mass accuracy for database searching while accurate precursor ion masses are simultaneously obtained in the slower orbitrap analyzer.

1.4 Electrospray Ionization in Mass Spectrometry

Most mass spectrometry analyzers, including FT-ICR, orbitrap, and linear ion trap, require molecules to be measured as gas phase ions, under vacuum. In order to introduce biomolecules

that are not volatile into mass spectrometers, a "soft" ion source is required. Such sources need to protect large biomolecules from disintegrating and may even preserve their native structure.[31-33] ESI is one of these "soft" ionization methods.[34-35] For ESI, volatile solvents are typically applied to dissolve biological samples, such as peptides and proteins. Methanol, acetonitrile, and water are major solvents employed in this thesis. LC eluent, usually a mixture of aqueous and organic solvents, is also compatible with ESI. Acid or base is often added in solvents to promote protonation for cation generation or deprotonation for anion generation, respectively.

A scheme of the ESI process is shown in Figure 1.4. Dissolved samples flow through a spray needle with an electric potential (~1-6 kV) applied. Charged analyte molecules are gathered by such strong electric field, and pushed into the end of spray needle. Due to the applied electric field, ions of the same polarity gather at the tip of the spray needle. When Coulomb repulsion between these ions overcomes the surface tension of the solvent, a Taylor cone is formed, and flowing solvent bursts into small charged droplets. Heat is typically applied to promote solvent evaporation from these droplets. As the size of droplets decreases, their charge density increases, and Coulomb repulsion will again eventually overcome the surface tension of the droplets, resulting in Coulombic fission: larger droplets disintegrate into multiple smaller droplets, splitting the original charge. This process repeats until desolvated ions start to generate.

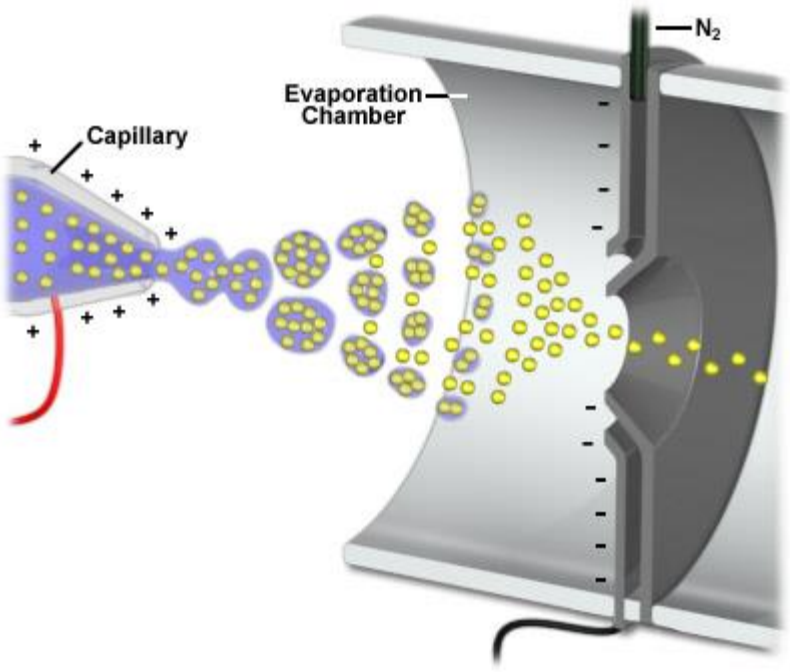


Figure 1.5. Schematic diagram of electrospray ionization, producing cations. Source: National High Magnetic Field Laboratory website (nationalmaglab.org/user-facilities/icr/techniques/esi).

The mechanism for desolvated ion generation in ESI is debated. Two major models, the ion evaporation model (IEM)[36] and the charge residue model (CRM),[37-38] fit experimental observation in different aspects. In IEM, Coulomb repulsion is strong enough to eject ions from droplets at some point before complete desolvation. In CRM, Coulombic fission repeats until one single ion is left in a droplet. Complete solvent evaporation of such droplets generates desolvated ions. Experimental results indicate that large intact biomolecular ions or biomolecular complex ions are produced by CRM,[39-40] while smaller ions are produced by IEM.[41] More recently, for nonpolar polymer chains, a chain ejection model (CEM) has been proposed.[42] Such nonpolar polymer chains may come from unfolding of charged proteins. Due to their hydrophobicity, they are unfavorable to reside within the droplet. Once they migrate to the surface of droplets, they are ejected into the gas phase due to Coulomb repulsion.

In addition to gentle ionization and LC compatibility, ESI generates multiply charged ions for larger analytes, constituting another advantage for FT-ICR and Orbitrap MS. The benefit of multiple charging includes the following aspects: a) Generated image current in FT-ICR and Orbitrap MS is proportional to the charge of ions, thus higher ion charge increases signal-to-noise ratio.[43] b) The limited m/z range of these mass analyzers can be overcome by analyzing ions that carry more charges. Also, FT-ICR analyzers show improved performance in the low m/z area, with higher mass accuracy, resolving power and limit of detection.[44] c) Some MS/MS methods, such as electron capture dissociation (ECD) and electron detachment dissociation (EDD), require multiply charged ions.

1.5 Tandem Mass Spectrometry (MS/MS, or MSn)

1.5.1 Basic MS^n Principles

Biopolymers are composed of repeating basic units, e.g., proteins are formed from linking amino acid residues and nucleic acids are formed by linking nucleotides. These basic building blocks can be arranged in different orders, resulting in a large number of isomers, which cannot be separated by single stage mass spectrometry analysis. In order to address this problem, tandem mass spectrometry can be utilized. The basic idea of tandem mass spectrometry is to fragment selected analyte ions, i.e., precursor ions, in a predictable manner, so that sequence information for these ions can be recovered, and isomeric forms of ions can be differentiated.

In most cases, an MS/MS experiment starts with precursor ion selection, because typically only some specific ions are of interest, and fragment ions from multiple precursor ions may complicate result interpretation. Quadrupoles are the most commonly adopted ion selecting

device.[45-46] When a quadrupole is operating in mass filtering mode, electric potentials are applied on all rods. These electric potentials have a DC bias and an RF frequency, and neighboring rods have different polarities. The DC bias and RF frequency determine a specific m/z window and only ions whose m/z are in this window can stably oscillate between the quadrupole rods, and thus reach the other end of the quadrupole. All other ions are ejected out of the quadrupole and pumped away.

In FT-ICR mass spectrometry, ion selection can also be done by stored waveform inverse Fourier transform (SWIFT) inside the ICR cell.[47-49] In SWIFT, frequencies of all unwanted ions in the ICR cell are derived from their m/z. These frequencies are compiled into a frequency domain waveform, which is then inversely Fourier transformed into a time domain waveform. The time domain waveform is used for generating a high amplitude excitation pulse on the excitation plates. Unwanted ions are thus excessively excited, hit the electrodes in the cell walls, become neutralized, and pumped away.

After precursor ion selection, remaining ions are activated to induce dissociation. Depending on whether ion selection and activation occur at the same location in the instrument or not, tandem mass spectrometry experiments can be tandem-in-space or tandem-in-time. In the tandem-in-space case, ion selection and activation occur at different locations within the instrument. For example, when we use the Solarix instrument to perform CAD, ion selection is performed with the quadrupole, while collisional activation occurs in the hexapole collision cell. In the tandem-in-time case, ion selection and activation occur in the same region of the instrument, but sequentially in time. Such selection-activation can be repeated n times, denoted

as MS^n. ICR cells can practically perform up to MS^4,[50-51] while ion traps can perform up to MS^{10} [52] by commercially available instruments.

For peptides and proteins, MS/MS-based cleavages of the peptide backbone, as well as the fragments produced by such cleavages, are most informative. In order to describe the fragments produced in peptide MS/MS by their cleavage sites, Roepstorff and Fohlman proposed a nomenclature for MS/MS fragments in 1984.[53] Figure 1.5 demonstrates peptide backbone fragments under this nomenclature.

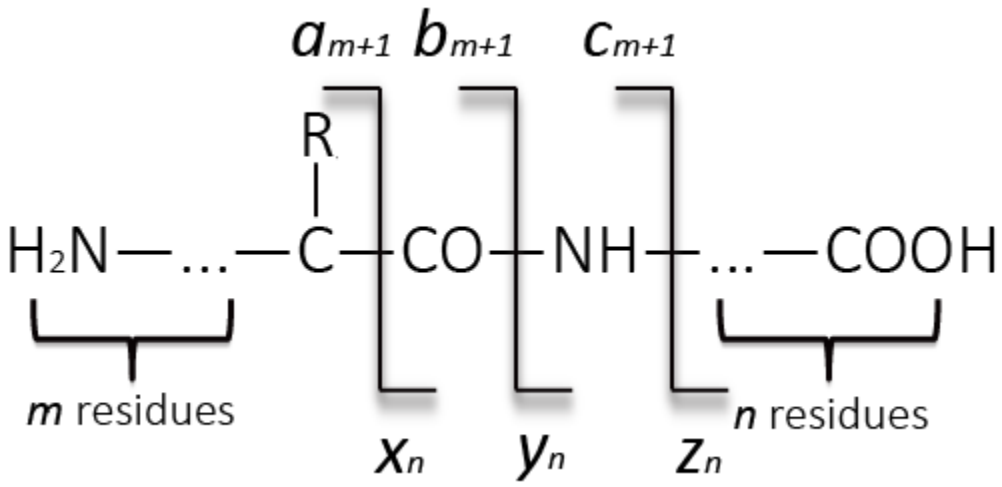


Figure 1.6. Nomenclature for peptide/protein fragment ions in tandem mass spectrometry.

According to the activation method applied, such cleavage may occur at different bonds of the peptide backbone.[54-57] Activation methods involved in this thesis are discussed in the following sections.

1.5.2 Collision Activated Dissociation (CAD)

CAD, also referred to as collision-induced dissociation (CID), is the most widely used MS/MS fragmentation method.[58] In CAD, precursor ions are activated by inelastic collisions

with neutral molecules, such as argon, followed by dissociation. A widely accepted model for peptide dissociation after collision in positive ion mode CAD is the mobile proton model.[59-62] In their ground state, peptide ions should have their protons located at the most basic sites, such as lysine and arginine residues and the N-terminus. When such ions are collisionally activated, the increased energy allows protons to be relocated to other sites within the peptide, such as the amide nitrogen of peptide bonds. Peptide bonds can therefore be weakened, and become breakable by nucleophilic attack by either the neighboring amide oxygen[63] or the nitrogen.[64] In negative ion mode CAD, the dissociation mechanism is different from that in positive mode CAD, and the results are usually less predictable.[65]

According to the collision energy range, there are two types of CAD: low-energy (1-100 eV) and high-energy (>1 keV). In this thesis, only low-energy CAD was involved, thus the acronym CAD stands for low-energy CAD only. There are multiple ways to implement CAD. The most common implementations are beam-type and ion trap-type CAD. In beam-type CAD, precursor ions are accelerated through a multipole filled with inert gas. During such CAD experiments, the multipole operates in RF-only mode to confine all precursor ions and fragment ions between the rods. On the FT-ICR MS instrument involved in this thesis, this process occurs at the collision cell and argon is used as the collision gas. Triple quadrupole and quadrupole-time-of-flight instruments are equipped with beam-type CAD as well. In ion trap-type CAD, ions are accelerated by resonant excitation, followed by collisions with inert background gas. Compared with beam-type CAD, ion trap-type CAD has lower energy deposit rate, requires longer activation time (10^{-1}~10^{-3} s for ion trap-type, while 10^{-4} s for beam-type), and produces more selective cleavages of the most labile bonds.[66] Similar to ion-trap type CAD, sustained off-resonance irradiation (SORI) CAD[67] is available in FT-ICR cells, for which an off-resonance

waveform is applied for excitation of specific ions. Due to the UHV condition in FT-ICR cells (~10^{-10} mbar), neutral inert gas have to be pulsed into the cell as collision gas. This procedure limits the application of SORI-CAD, because after a SORI-CAD experiment, UHV needs to be restored prior to detection, otherwise collisional damping of excited ions reduces resolution. The throughput of SORI-CAD is reduced by this pumping procedure.

In CAD, vibrational activation of precursor ions occurs.[68] The energy gain from each individual collision may be insufficient for breaking a bond, but such energy gain can be cumulative until a bond is broken. Ion internal energy is distributed across all bonds through intramolecular vibrational-energy redistribution.[69] This "slow-heating" process results in cleavage of the most labile bonds through the most energetically preferred pathways.[70] In the case of peptide and protein backbone cleavage, the preferred site is the peptide bond (CO-NH bond). Such cleavage produces *b* and *y* fragment ions. Neutral loss of small molecules, such as H_2O and NH_3, are also often observed.

The major advantage of CAD includes high reproducibility and reliability, versatile instrument compatibility, and predictable product ions. Based on these features, various MS/MS database searching tools have been developed, and put to use in proteomics research.[71] However, one major limitation of CAD is for protein modification studies. Many post-translational modifications (PTMs), such as phosphorylation, sulfation and glycosylation, are connected to the peptide backbone through bonds that are weaker than peptide bonds. Therefore, in the slow-heating process of CAD, these PTMs are preferentially cleaved. The corresponding mass spectra are dominated by PTM neutral loss instead of backbone cleavage fragment ions,

and the information about PTM sites is not retained in CAD MS/MS spectra.[72] In some cases, phosphorylation can migrate in CAD, further complicating data interpretation.[73-75]

1.5.3 Infrared Multiphoton Dissociation (IRMPD)

IRMPD is an alternative vibrational activation MS/MS method. Its basic principle is similar to that of CAD, while precursor ions are activated not through collision with neutral gas molecules, but through absorption of multiple photons from an IR laser beam. Each photon carries about 0.1 eV energy,[76] thus multiple photons have to be absorbed by precursor ions before fragmentation begins, also making IRMPD a slow-heating MS/MS method. IRMPD yields *b* and *y* ions as well as neutral losses and PTM losses, similar to CAD. The most commonly applied laser for IRMPD is a 10.6 μm CO_2 laser, as on Solarix.[77-80] Tunable lasers such as free electron lasers can also be used.[81] IRMPD has been implemented in FT-ICR cells,[82] ion traps[76] and an external hexapole ion reservoir.[83] FT-ICR cells are best suited for IRMPD, due to the UHV minimizing collisional cooling of activated precursor ions. In addition, no collision gas is needed in the FT-ICR cell for IRMPD. In this thesis, IRMPD was often combined with ECD, EDD and niECD as a supplementary activation method to disintegrate fragment product ions that are non-covalently bound together.[84-85]

1.5.3 Electron Capture Dissociation (ECD) and Electron Transfer Dissociation (ETD)

ECD[86] and ETD[87] are electron-based MS/MS methods, operating in positive ion mode. Instead of activating precursor ions to a higher vibrational energy level, multiply charged precursor ions acquire an electron, become charge reduced radicals, and dissociate into fragment

ions. In ECD, the electron is provided by irradiation with low-energy (~1eV) electrons generated by a cathode. In ETD, the electron is provided by anions, or ETD reagent. Aromatic molecules with multiple rings are commonly used as reagent, due to their low electron affinity. On Solarix, the ETD reagent is fluoranthene.[88] The ECD and ETD reactions can be written as follows:

ECD: $[M + nH]^{n+}$ (n>1) + $\mathbf{e}^-$ (<1 eV)→$[M + nH]^{(n-1)+\bullet}$→fragments

ETD: $[M + nH]^{n+}$ (n>1) + $\mathbf{A}^{-\bullet}$ →$[M + nH]^{(n-1)+\bullet}$ + A →fragments

In order to create charged fragment ions, multiply charged precursor ions are required for ECD/ETD. ECD/ETD show improved performance for high charge state precursor ions,[89-90] due to higher energy release and Coulomb repulsion between product ion pairs.

The ECD mechanism is under debate. There are two dominant models to explain ECD: the hot hydrogen mechanism[91] and the amide superbase mechanism.[92] In the hot hydrogen mechanism, the electron is captured by one of a peptide's protonated sites, such as the N-terminus or a lysine, arginine or histidine side chain. Neutralized protons gain internal energy and become mobile. When this "hot" hydrogen relocates to a backbone carbonyl, an amino ketyl radical (Cα-C$^{\bullet}$OH-NH-Cα) is formed. This radical immediately dissociates through N-Cα cleavage, producing *c* ions and $z^{\bullet}$ ions (with Cα as a radical). In the amide superbase mechanism, the electron is captured by a peptide bond, making the backbone amide a superbase. This process is facilitated by nearby protons, and the generated amide can draft a proton from its vicinity, producing an amino ketyl radical. The radical then dissociates in a similar manner to the hot hydrogen model. According to current experimental observations,[92-100] the two mechanisms may occur at the same time and preference of mechanism depends on the nature of a peptide and

its charge sites. In Chapter 2, the case when metal ions serve as alternative charge carriers is discussed. ETD is usually believed to have a similar if not the same mechanism as ECD.[101-102]

ECD/ETD provide complementary information compared with CAD. Radical-driven fragmentation produces $c/z^{\bullet}$ ions rather than the *b/y* ions from CAD. The difference in cleavage site makes ECD/ETD spectra and CAD spectra of the same precursor ions contain fragment ion pairs that differ from one another by ~15 Daltons. Such fragment ion pairs give information about whether a fragment ion contains the peptide N-terminus or C-terminus and thus can be applied for de novo sequencing of peptides and proteins.[103] Additional benefit of ECD comes from its presumed non-ergodic nature with. higher sequence coverage,[104] and more random cleavage sites compared to CAD, because cleavage does not necessarily happen at the most energetically preferable sites. For the same reason, ECD/ETD can produce fragment ions with labile PTMs retained. This feature makes ECD/ETD a powerful tool for PTM studies, especially labile PTMs such as phosphorylation,[105] sulfation[106] and glycosylation.[107] Information about their sites can be acquired.

1.5.5 Electron Detachment Dissociation (EDD)

EDD is the first electron-based MS/MS method operating in negative ion mode.[108] In EDD, precursor anions are irradiated by relatively high energy electrons (10-30 eV). These high energy electrons can eject an electron from the precursor ions, forming a charge reduced radical. For peptides and proteins, such radicals rapidly dissociate via Cα-C bond cleavage, forming *a*- and *x*-ions.[109] In some cases, neutral losses such as CO_2 or H_2O loss can also happen.[110] The EDD reaction can be written as:

EDD: $[M - nH]^{n-}$ (n>1) + $\mathbf{e}^-$ (10-30 eV)→$[M - nH]^{(n-1)+\bullet *}$ + $\mathbf{2e}^-$→fragments

As a radical-driven process, EDD can cleave peptide bonds, retaining labile PTMs on fragment ions.[111-112] This feature makes EDD suitable for the analysis of labile PTMs on acidic peptides, which are difficult to ionize in positive mode. However, EDD requires fine tuning of the electron energy level, limiting its compatibility with different analytes.[113] EDD suffers from inefficient backbone cleavages and non-informative neutral losses. In Chapter 3 of this thesis, alternative charge carriers are applied as an attempt to alleviate the latter issue.

1.5.6 Negative-ion Electron Capture Dissociation (niECD)

Negative ion electron capture dissociation is a relatively novel electron-based MS/MS technique, discovered in our lab.[114] Its reaction process is highly similar to ECD, but with anion precursors. Energetically it is unfavorable for an anion to capture an electron and increase its charge in the gas phase due to Coulombic repulsion. Therefore, only a narrow range of electron energy (3.5–6.5 eV) can trigger niECD, and the efficiency of this reaction is highly dependent on whether the precursor ions has a positive charging site to form a zwitterion. Once the electron is captured, an electron-rich radical similar to the radicals formed in ECD is generated, and undergoes similar dissociation pathways as the radicals in ECD, generating *c* and *z* fragment ions. The niECD reaction can be written as:

niECD: $[M - nH]^{n-} + \mathbf{e}^{-} (3.5\text{-}6.5\ eV) \rightarrow [M - nH]^{(n+1)-\bullet} \rightarrow$ fragments

niECD combines the analytical benefits of ECD/ETD and EDD. It produces highly predictable *c* and *z* ions, is PTM-compatible, and operates in negative mode, thus allowing acidic molecules to be more effectively ionized. niECD does not require multiply charged precursor ions, and its charge increase nature facilitates FT-ICR ion detection by increasing image current

amplitude. However, drawbacks still exist for this method. Not all peptides can be fragmented by niECD, and niECD efficiency varies significantly for different peptides. In Chapter 3 of this thesis, alternative charge carriers are applied as an attempt to alleviate this issue.

1.6 Mass Spectrometry Database Searching

In proteomics research, bottom-up LC/MS/MS is the most commonly applied method.[54, 115] It provides solid qualitative information about protein identification and peptide sequence. Combined with other techniques such as selected-reaction monitoring (SRM) [116] and isobaric tagging,[117] quantitation information can also be acquired. During each bottom-up LC/MS/MS run, one or multiple proteins are digested, producing dozens to hundreds of peptides. A mixture of these peptides is injected into an LC for separation, and separated peptides are analyzed by MS/MS. Currently top ten data dependent (DDA) LC/MS/MS is the most commonly applied strategy for LC/MS/MS.[118] In top ten data dependent LC/MS/MS, MS/MS scans occur after each precursor ion MS scan. The top ten most abundant ions in the MS scan are chosen as precursor ions for subsequent MS/MS scans. Once ten MS/MS scans are completed, or MS/MS scans on all precursor ions are completed, the instrument moves on and acquires the next MS scan. Ion trap-Orbitrap hybrid instruments are widely employed for LC/MS/MS experiments.[119] During LC/MS/MS on this type of instrument, MS detection is often performed in the Orbitrap to ensure high mass accuracy of precursor ions, as well as reliable quantification of detected ions. MS/MS experiments and their detection can be completed in the ion trap, in order to increase throughput. However, it is inevitable that some low-intensity peptides that co-elute with other peptides are missed by MS/MS. This missing peptide issue is aggravated when having multiple proteins in a sample, common in proteomics studies of biological samples.[120] Data-independent

acquisition (DIA) was invented to overcome drawbacks of top ten DDA.[118] In DIA, MS/MS is carried out for all peptide ions[121] or peptide ions within a certain m/z range that is larger than a typical MS/MS selection window.[118] Fragment ions can be extracted and assigned to precursor ions according to their elution time, and thus reconstitute a "pseudo-DDA" data set for analysis.[122]

In addition to DDA and DIA, targeted LC/MS/MS is a commonly used acquisition strategy. In targeted LC/MS/MS, an inclusion list with m/z values for analyte ions of interest and their elution time ranges directs MS/MS experiments. Only ions on the inclusion list will be subjected to MS/MS experiments. In Chapter 4, details about this strategy as well as how it can be applied in protein sequence variant analysis are discussed.

Currently, the most commonly applied data analysis strategy for LC/MS/MS data is database searching. By matching MS/MS spectra to those of proteolytic peptides, the probability of having a specific digested protein in the injected sample can be derived.[123] However, false positives exist in both MS/MS spectral matching and protein probability calculations. Target-decoy search strategy can massively reduce false positives in both steps for proteomics studies.[124] In this strategy, a decoy database is constructed according to the type of false positive to be removed. For example, by reversing the sequence of proteins, a decoy database focusing on removing *b/y* ion confusion can be built.[124] Other more advanced algorithms can build decoy databases for multiple false positive possibilities.[125-126] Search hits in both target and decoy databases with higher probability scores in the decoy database are removed.

1.7 Dissertation Overview

Research presented in this dissertation focus on two aspects. The first aspect is exploring how alternative charge carriers can affect ion-electron reactions in ECD and EDD. The second aspect is implementing a targeted LC/MS/MS strategy in antibody characterization. The first chapter provides a general overview of mass spectrometry-based proteomics, FT-ICR and Orbitrap MS involved in this thesis, the soft-ionization technique used in biomolecule analysis (ESI), various MS/MS techniques, and bottom-up LC/MS/MS database searching strategy. Chapter 2 discusses how metal-peptide interaction affects ECD. Metal-peptide complex ECD was extensively investigated to clarify how the interaction between peptide and different metal ions affect ECD, and the underlying mechanism of metal-peptide complex ECD. Chapter 3 discusses how salt anions affect ion-electron reactions in negative ion mode, i.e., EDD. Salt anion-adducted peptide ions were generated in negative mode and subjected to EDD. Influence of such adduction was assessed quantitatively. Chapter 4 introduces a novel workflow for low-level protein sequence variant detection in therapeutic protein samples. This chapter is based on my research at Bristol-Myers Squibb Co. in 2015. A targeted LC/MS/MS workflow was developed for low-level protein sequence variant analysis. Based on statistical analysis of protein sequence variant levels in multiple clones, probability of a protein sequence variant to be a DNA error can be derived. A summary of all results in this dissertation and future directions are presented in Chapter 5. Chapters 2 and 4 are to be submitted for journal publication at the point of this dissertation being finished.

1.8 References

1. Blackstock, W. P.; Weir, M. P., Proteomics: quantitative and physical mapping of cellular proteins. *Trends in Biotechnology* **1999,** *17*, 121-127.
2. International Human Genome Sequencing, C., Finishing the euchromatic sequence of the human genome. *Nature* **2004,** *431*, 931-45.

3. Wienkoop, S.; Weckwerth, W., Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in Arabidopsis thaliana. *J. Exp. Bot.* **2006,** *57*, 1529-1535.
4. Anderson, N. L.; Anderson, N. G., The human plasma proteome - History, character, and diagnostic prospects. *Mol. Cell. Proteomics* **2002,** *1*, 845-867.
5. Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006,** *127*, 635-648.
6. Warner, S. C.; Waite, J. H., Expression of multiple forms of an adhesive plaque protein in an individual mussel, Mytilus edulis. *Mar. Biol.* **1999,** *134*, 729-734.
7. Wurm, F. M., Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **2004,** *22*, 1393-1398.
8. Zhu, Z. P.; Zapata, G.; Shalaby, R.; Snedecor, B.; Chen, H.; Carter, P., High level secretion of a humanized bispecific diabody from Escherichia coli. *Bio-Technology* **1996,** *14*, 192-196.
9. Zeck, A.; Regula, J. T.; Larraillet, V.; Mautz, B.; Popp, O.; Gopfert, U.; Wiegeshoff, F.; Vollertsen, U. E. E.; Gorr, I. H.; Koll, H.; Papadimitriou, A., Low Level Sequence Variant Analysis of Recombinant Proteins: An Optimized Approach. *PLoS One* **2012,** *7*.
10. Doneanu, C. E.; Xenopoulos, A.; Fadgen, K.; Murphy, J.; Skilton, S. J.; Prentice, H.; Stapels, M.; Chen, W. B., Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. *mAbs* **2012,** *4*, 24-44.
11. Pandey, A.; Mann, M., Proteomics to Study Genes and Genomes. *Nature* **2000,** *405*, 837-846.
12. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., ELECTROSPRAY IONIZATION FOR MASS-SPECTROMETRY OF LARGE BIOMOLECULES. *Science* **1989,** *246*, 64-71.
13. Castro, J. A.; Koster, C.; Wilkins, C., Matrix-assisted Laser Desorption Ionization of High-mass Molecules by Fourier Transform Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1992,** *6*, 239-241.
14. Chait, B. T., Mass spectrometry: Bottom-up or top-down? *Science* **2006,** *314*, 65-66.
15. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. In *Annual Review of Biomedical Engineering*, Annual Reviews: Palo Alto, 2009; Vol. 11, pp 49-79.
16. Zhang, Y. Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013,** *113*, 2343-2394.
17. Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W., Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* **2002,** *124*, 672-678.
18. Comisarow, M. B.; Marshall, A. G., Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **1974,** *25*, 282-283.
19. Comisarow, M. B.; Marshall, A. G., Frequency-Sweep Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **1974,** *26*, 489-490.
20. Gorman, J. J.; Wallis, T. P.; Pitt, J. J., Protein disulfide bond determination by mass spectrometry. *Mass Spectrom. Rev.* **2002,** *21*, 183-216.
21. Shi, S. D.-H.; Hendrickson, C. L.; Marshall, A. G., Counting Individual Sulfur Atoms in a Protein by Ultrahigh-Resolution Fourier Transform Ion Cyclotron Resonance Mass

Spectrometry: Experimental Resolution of Isotopic Fine Structure in Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95*, 11532-11537.
22. Mitchell, D. W.; Smith, R. D., Cyclotron Motion Of 2 Coulombically Interacting Ion Clouds with Implications to Fourier-Transform Ion-Cyclotron Resonance Mass-spectrometry. *Phys. Rev. E* **1995,** *52*, 4366-4386.
23. Amster, I. J., Fourier transform mass spectrometry. *J. Mass Spectrom.* **1996,** *31*, 1325-1337.
24. Haas, W.; Faherty, B. K.; Gerber, S. A.; Elias, J. E.; Beausoleil, S. A.; Bakalarski, C. E.; Li, X.; Villen, J.; Gygi, S. P., Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell. Proteomics* **2006,** *5*, 1326-1337.
25. Marshall, A. G.; Rodgers, R. P., Petroleomics: The next grand challenge for chemical analysis. *Accounts Chem. Res.* **2004,** *37*, 53-59.
26. Xie, Y. M.; Zhang, J.; Yin, S.; Loo, J. A., Top-down ESI-ECD-FT-ICR mass spectrometry localizes noncovalent protein-ligand binding sites. *J. Am. Chem. Soc.* **2006,** *128*, 14432-14433.
27. Makarov, A., Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000,** *72*, 1156-1162.
28. Hu, Q. Z.; Noll, R. J.; Li, H. Y.; Makarov, A.; Hardman, M.; Cooks, R. G., The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **2005,** *40*, 430-443.
29. Perry, R. H.; Cooks, R. G.; Noll, R. J., Orbitrap Mass Spectrometry: Instrumenttaion, Ion Motion and Applications. *Mass Spectrom. Rev.* **2008,** *27*, 661-699.
30. Makarov, A.; Denisov, E.; Kholomeev, A.; Baischun, W.; Lange, O.; Strupat, K.; Horning, S., Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **2006,** *78*, 2113-2120.
31. Loo, J. A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **1997,** *16*, 1-23.
32. van den Heuvel, R. H.; Heck, A. J., Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **2004,** *8*, 519-526.
33. Kaddis, C. S.; Loo, J. A., Native protein MS and ion mobility large flying proteins with ESI. *Anal. Chem.* **2007,** *79*, 1778-1784.
34. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989,** *246*, 64-71.
35. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F., Electrospray Ionization - Principles and Practice. *Mass Spectrom. Rev.* **1990,** *9*, 37-70.
36. Iribarne, J. V.; Thomson, B. A., On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics* **1976,** *64*, 2287-2294.
37. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B., Molecular Beams of Macroions. *The Journal of Chemical Physics* **1968,** *49*, 2240-2249.
38. Mack, L. L.; Kralik, P.; Rheude, A.; Dole, M., Molecular Beams of Macroions. II. *The Journal of Chemical Physics* **1970,** *52*, 4977-4986.
39. Fernandez de la Mora, J., Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta* **2000,** *406*, 93-104.
40. Smith, R. D.; Loo, J. A.; Ogorzalek Loo, R. R.; Busman, M.; Udseth, H. R., Principles and Practice of Electrospray Ionization - Mass Spectrometry for Large Polypeptide and Proteins. *Mass Spectrom. Rev.* **1991,** *10*, 359-451.

41. Gamero-Castaño, M.; Fernández de la Mora, J., Direct measurement of ion evaporation kinetics from electrified liquid surfaces. *The Journal of Chemical Physics* **2000,** *113*, 815-832.
42. Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S., Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013,** *85*, 2-9.
43. Fung, Y. M. E.; Adams, C. M.; Zubarev, R. A., Electron Ionization Dissociation of Singly and Multiply Charged Peptides. *J. Am. Chem. Soc.* **2009,** *131*, 9977-9985.
44. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrom. Rev.* **1998,** *17*, 1-35.
45. Belov, M. E.; Nikolaev, E. N.; Anderson, G. A.; Udseth, H. R.; Conrads, T. P.; Veenstra, T. D.; Masselon, C. D.; Gorshkov, M. V.; Smith, R. D., Design and Performance of an ESI Interface for Selective External Ion Accumulation Coupled to a Fourier Transform Ion Cyclotron Mass Spectrometer. *Anal. Chem.* **2001,** *73*, 253-261.
46. Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F., Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and use in the Comparative Analysis of Histone H3 Post-Translational Modifications *J. Proteome Res.* **2004,** *3*, 621-626.
47. Guan, S.; Marshall, A. G., Stored Waveform Inverse Fourier Transform (SWIFT) Ion Excitation in Trapped-ion Mass Spectrometry: Theory and Applications. *Int. J. Mass Spectrom. Ion Processes* **1996,** *157/158*, 5-37.
48. Wang, T. C. L.; Ricca, T. L.; Marshall, A. G., Extension of Dynamic-range in Fourier-Transform Ion-Cyclotron Resonance Mass-spectrometry via Stored Wave-form Inverse Fourier-Transform Excitation. *Anal. Chem.* **1986,** *58*, 2935-2938.
49. Chen, L.; Marshall, A. G., Effect of time-domain dynamic range on stored waveform excitation for Fourier transform ion cyclotron resonance mass spectrometry. *Rapid communications in mass spectrometry : RCM* **1987,** *1*, 39-42.
50. Zhang, J.; Schubothe, K.; Li, B.; Russell, S.; Lebrilla, C. B., Infrared Multiphoton Dissociation of O-Linked Mucin-Type Oligosaccharides. *Anal. Chem.* **2005,** *77*, 208-214.
51. Huang, Y.; Pasa-Tolic, L.; Guan, S.; Marshall, A. G., Collision-Induced Dissociation for Mass Spectrometric Analysis of Biopolymers: High-Resolution Fourier Transform Ion Cyclotron Resonance MS4. *Anal. Chem.* **1994,** *66*, 4385-4389.
52. Zucker, S. M.; Lee, S.; Webber, N.; Valentine, S. J.; Reilly, J. P.; Clemmer, D. E., An Ion Mobility/Ion Trap/Photodissociation Instrument for Characterization of Ion Structure. *J. Am. Soc. Mass Spectrom.* **2011,** *22*, 1477-1485.
53. Roepstorff, P.; Fohlman, J., Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984,** *11*, 601-601.
54. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422*, 198-207.
55. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001,** *19*, 242-247.
56. de Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Frohlich, F.; Walther, T. C.; Mann, M., Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **2008,** *455*, 1251-U60.
57. Sleno, L.; Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **2004,** *39*, 1091-1112.

58. McLuckey, S. A., Principles of Collisional Activation in Analytical Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1992,** *3*, 599-614.
59. Biemann, K.; Martin, S. A., MS/MS of Peptides. *Mass Spectrom. Rev.* **1987,** *6*, 1-76.
60. Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H., Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.* **1996,** *118*, 8365–8374.
61. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A., Special Feature: Commentary - Mobile and Localized Protons: a Framework for Understanding Peptide Dissociation. *J. Mass Spectrom.* **2000,** *35*, 1399-1406.
62. Paizs, B.; Suhai, S., Fragmentation Pathways of Protonated Peptides. *Mass Spectrom. Rev.* **2005,** *24*, 508-548.
63. Vaisar, T.; Urban, J., Gas-phase fragmentation of protonated mono-N-methylated peptides. Analogy with solution-phase acid-catalyzed hydrolysis. *J. Mass Spectrom.* **1998,** *33*, 505-524.
64. Eckart, K.; Holthausen, M. C.; Koch, W.; Spiess, J., Mass spectrometric and quantum mechanical analysis of gas-phase formation, structure, and decomposition of various b(2) ions and their specifically deuterated analogs. *J. Am. Soc. Mass Spectrom.* **1998,** *9*, 1002-1011.
65. Bowie, J. H.; Brinkworth, C. S.; Dua, S., Collision-Induced Fragmentations of the (M-H)(-) Parent Anions of Underivatized Peptides: An Aid to Structure Determination and some Unusual Negative Ion Cleavages. *Mass Spectrom. Rev.* **2002,** *21*, 87-107.
66. Wells, J. M.; McLuckey, S. A., Collision-Induced Dissociation (CID) of Peptides and Proteins. *Biol. Mass Spectrom.* **2005,** *402*, 148-185
67. Gauthier, J. W.; Trautman, T. R.; Jacobson, D. B., Sustained Off-resonance Irradiation for CAD Involving FTMS - CAD Technique that Emulates Infrared Multiphoton Dissociation. *Anal. Chim. Acta* **1991,** *246*, 211-225.
68. Schwartz, R. N.; Slawsky, Z. I.; Herzfeld, K. F., Calculation of Vibrational Relaxation Times in Gases. *J. Chem. Phys.* **1952,** *20*, 1591-1599.
69. Stannard, P. R.; Gelbart, W. M., INTRAMOLECULAR VIBRATIONAL-ENERGY REDISTRIBUTION. *J. Phys. Chem.* **1981,** *85*, 3592-3599.
70. McLuckey, S. A.; Goeringer, D. E., Slow heating methods in tandem mass spectrometry. *J. Mass Spectrom.* **1997,** *32*, 461-474.
71. Sadygov, R. G.; Cociorva, D.; Yates, J. R., Large-scate database searching using tandem mass spectra: Looking up the answer in the back of the book. *Nat. Methods* **2004,** *1*, 195-202.
72. Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G., Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* **2007,** *4*, 798-806.
73. Palumbo, A. M.; Reid, G. E., Evaluation of Gas-Phase Rearrangement and Competing Fragmentation Reactions on Protein Phosphorylation Site Assignment Using Collision Induced Dissociation-MS/MS and MS3. *Anal. Chem.* **2008,** *80*, 9735-9747.
74. Mischerikow, N.; Altelaar, A. F. M.; Navarro, J. D.; Mohammed, S.; Heck, A. J. R., Comparative Assessment of Site Assignments in CID and ETD Spectra of Phosphopeptides Discloses Limited Relocation of Phosphate Groups. *Mol. Cell. Proteomics* **2010,** *9*, 2140-2148
75. Aguiar, M.; Haas, W.; Beausoleil, S. A.; Rush, J.; Gygi, S. P., Gas-Phase Rearrangements Do Not Affect Site Localization Reliability in Phosphoproteomics Data Sets. *J. Proteome Res.* **2010,** *9*, 3103–3107.
76. Brodbelt, J. S.; Wilson, J. J., Infrared Multiphoton Dissociation in Quadrupole Ion Traps. *Mass Spectrom. Rev.* **2009,** *28*, 390-424.

77. Adamson, J. T.; Hakansson, K., Infrared Multiphoton Dissociation and Electron Capture Dissociation of High-Mannose Type Glycopeptides. *J. Proteome Res.* **2006,** *5*, 493-501.
78. Yang, J.; Hakansson, K., Characterization of Oligodeoxynucleotide Fragmentation Pathways in Infrared Multiphoton Dissociation and Electron Detachment Dissociation by Fourier Transform Ion Cyclotron Double Resonance. *Eur. J. Mass Spectrom.* **2009,** *15*, 293-304.
79. Kalli, A.; Hakansson, K., Preferential cleavage of S-S and C-S bonds in electron detachment dissociation and infrared multiphoton dissociation of disulfide-linked peptide anions. *Int. J. Mass Spectrom.* **2007,** *263*, 71-81.
80. Zhou, W.; Hakansson, K., Gas-Phase Ion-Electron and Ion-Photon Reactions of Pronase-Derived Glycopeptides. *Glycobiology* **2010,** *20*, 1473-1474.
81. Valle, J. J.; Eyler, J. R.; Oomens, J.; Moore, D. T.; van der Meer, A. F. G.; von Helden, G.; Meijer, G.; Hendrickson, C. L.; Marshall, A. G.; Blakney, G. T., Free electron laser-Fourier transform ion cyclotron resonance mass spectrometry facility for obtaining infrared multiphoton dissociation spectra of gaseous ions. *Review of Scientific Instruments* **2005,** *76*, 023103.
82. Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W., Infrared Multiphoton Dissociation of Large Multiply-charged Ions for Biomolecule Sequencing. *Anal. Chem.* **1994,** *66*, 2809-2815.
83. Hofstadler, S. A.; Sannes-Lowery, K. A.; Griffey, R. H., Infrared Multiphoton Dissociation in an External Ion Reservoir. *Anal. Chem.* **1999,** *71*, 2067-2070.
84. Horn, D. M.; Ge, Y.; McLafferty, F. W., Activated Ion Electron Capture Dissociation for Mass Spectral Sequencing of Larger (42 kDa) Proteins. *Anal. Chem.* **2000,** *72*, 4778-4784.
85. Mikhailov, V. A.; Cooper, H. J., Activated Ion Electron Capture Dissociation (AI ECD) of Proteins: Synchronization of Infrared and Electron Irradiation with Ion Magnetron Motion *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 763-771.
86. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998,** *120*, 3265-3266.
87. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A* **2004,** *101*, 9528-9533.
88. Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F., Analysis of Phosphorylation Sites on Proteins from Saccharomyces cerevisiae by Electron Transfer Dissociation (ETD) Mass Spectrometry *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, 2193-2198.
89. Iavarone, A. T.; Paech, K.; Williams, E. R., Effects of Charge State and Cationizing Agent on the Electron Capture Dissociation of a Peptide. *Anal. Chem.* **2004,** *76*, 2231-2238.
90. Liu, J.; McLuckey, S. A., Electron transfer dissociation: Effects of cation charge state on product partitioning in ion/ion electron transfer to multiply protonated polypeptides. *Int. J. Mass Spectrom.* **2012,** *330*, 174-181.
91. Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W., Electron Capture Dissociation of Gaseous Multiply-charged Proteins is Favored at Disulfide Bonds and Other Sites of High Hydrogen Atom Affinity. *J. Am. Chem. Soc.* **1999,** *121*, 2857-2862.
92. Syrstad, E. A.; Turecek, F., Toward a General Mechanism of Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2005,** *16*, 208-224.

93. Turecek, F., N-C-alpha Bond Dissociation Energies and Kinetics in Amide and Peptide Radicals. Is the Dissociation a Non-Ergodic Process? *J. Am. Chem. Soc.* **2003,** *125*, 5954-5963.
94. Sawicka, A.; Skurski, P.; Hudgins, R. R.; Simons, J., Model Calculations Relevant to Disulfide Bond Cleavage via Electron Capture Influenced by Positively Charged Groups. *J. Phys. Chem. B* **2003,** *107*, 13505-13511.
95. Breuker, K.; Oh, H. B.; Lin, C.; Carpenter, B. K.; McLafferty, F. W., Nonergodic and Conformational Control of the Electron Capture Dissociation of Protein Cations. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101*, 14011-14016.
96. Sobczyk, M.; Anusiewicz, W.; Berdys-Kochanska, J.; Sawicka, A.; Skurski, P.; Simons, J., Coulomb-Assisted Dissociative Electron Attachment: Application to a Model Peptide. *J. Phys. Chem. A* **2005,** *109*, 250-258.
97. Patriksson, A.; Adams, C.; Kjeldsen, F.; Raber, J.; van der Spoel, D.; Zubarev, R. A., Prediction of N-C-alpha Bond Cleavage Frequencies in Electron Capture Dissociation of Trp-cage Dications by Force-field Molecular Dynamics Simulations *Int. J. Mass Spectrom.* **2006,** *248*, 124-135.
98. Chen, X. H.; Turecek, F., The Arginine Anomaly: Arginine Radicals are Poor Hydrogen Atom Donors in Electron Transfer Induced Dissociations *J. Am. Chem. Soc.* **2006,** *128*, 12520-12530
99. Jones, J. W.; Sasaki, T.; Goodlett, D. R.; Turecek, F., Electron Capture in Spin-trap Capped Peptides. An Experimental Example of Ergodic Dissociation in Peptide Cation-Radicals *J. Am. Soc. Mass Spectrom.* **2007,** *18*, 432-444
100. Leib, R. D.; Donald, W. A.; Bush, M. F.; O'Brien, J. T.; Williams, E. R., Internal Energy Deposition in Electron Capture Dissociation Measured using Hydrated Divalent Metal Ions as Nanocalorimeters *J. Am. Chem. Soc.* **2007,** *129*, 4894-4895.
101. Gunawardena, H. P.; He, M.; Chrisman, P. A.; Pitteri, S. J.; Hogan, J. M.; Hodges, B. D. M.; McLuckey, S. A., Electron transfer versus proton transfer in gas-phase ion/ion reactions of polyprotonated peptides. *J. Am. Chem. Soc.* **2005,** *127*, 12627-12639.
102. Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N., Peptide sequencing and characterization of post-translational modifications by enhanced ion-charging and liquid chromatography electron-transfer dissociation tandem mass spectrometry. *Anal. Chem.* **2007,** *79*, 9243-9252.
103. Savitski, M. M.; Nielsen, M., L.; Kjeldsen, F.; Zubarev, R. A., Proteomics-grade de novo Sequencing Approach. *J. Proteome Res.* **2005,** *4*, 2348-2354.
104. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W., Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem.* **2000,** *72*, 563-573.
105. Stensballe, A.; Norregaard-Jensen, O.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A., Electron Capture Dissociation of Singly and Multiply Phosphorylated Peptides. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 1793-1800.
106. Kelleher, N. L.; Zubarev, R. A.; Bush, K.; Furie, B.; Furie, B. C.; McLafferty, F. W.; Walsh, C. T., Localization of Labile Posttranslational Modifications by Electron Capture Dissociation: The Case of gamma-carboxyglutamic Acid. *Anal. Chem.* **1999,** *71*, 4250-4253.
107. Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A., Localization of O-glycosylation Sites in Peptides by Electron Capture Dissociation in a Fourier Transform Mass Spectrometer. *Anal. Chem.* **1999,** *71*, 4431-4436.

108. Budnik, B. A.; Haselmann, K. F.; Zubarev, R. A., Electron Detachment Dissociation of Peptide Di-anions: an Electron-hole Recombination Phenomenon. *Chem. Phys. Lett.* **2001,** *342*, 299-302.
109. Kjeldsen, F.; Silivra, O. A.; Ivonin, I. A.; Haselmann, K. F.; Gorshkov, M.; Zubarev, R. A., C(alpha)-C Backbone Fragmentation Dominates in Electron Detachment Dissociation of Gas-Phase Polypeptide Polyanions. *Chem. Eur. J.* **2005,** *11*, 1803-1812.
110. Anusiewicz, I.; Jasionowski, M.; Skurski, P.; Simons, J., Backbone and Side-Chain Cleavages in Electron Detachment Dissociation (EDD). *J. Phys. Chem. A* **2005,** *109*, 11332-11337.
111. Kleinnijenhuis, A. J.; Kjeldsen, F.; Kallipolitis, B.; Haselmann, K. F.; Jensen, O. N., Analysis of Histidine Phosphorylation Using Tandem MS and Ion-Electron Reactions. *Anal. Chem.* **2007**, 7450-7456.
112. Kweon, H. K.; Hakansson, K., Metal Oxide-Based Enrichment Combined with Gas-Phase Ion-Electron Reactions for Improved Mass Spectrometric Characterization of Protein Phosphorylation. *J. Proteome Res.* **2008,** *7*, 745-755.
113. Yang, J.; Hakansson, K., Characterization and Optimization of Electron Detachment Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Int. J. Mass Spectrom.* **2008,** *276*, 144-148
114. Yoo, H. J.; Wang, N.; Zhuang, S.; Song, H.; Håkansson, K., Negative-Ion Electron Capture Dissociation: Radical-Driven Fragmentation of Charge-Increased Gaseous Peptide Anions. *J. Am. Chem. Soc.* **2011,** *133*, 16790-16793.
115. Domon, B.; Aebersold, R., Review - Mass spectrometry and protein analysis. *Science* **2006,** *312*, 212-217.
116. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100*, 6940-6945.
117. Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004,** *3*, 1154-1169.
118. Geromanos, S. J.; Vissers, J. P. C.; Silva, J. C.; Dorschel, C. A.; Li, G. Z.; Gorenstein, M. V.; Bateman, R. H.; Langridge, J. I., The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **2009,** *9*, 1683-1695.
119. Scherl, A.; Shaffer, S. A.; Taylor, G. K.; Hernandez, P.; Appel, R. D.; Binz, P. A.; Goodlett, D. R., On the benefits of acquiring peptide fragment ions at high measured mass accuracy. *J. Am. Soc. Mass Spectrom.* **2008,** *19*, 891-901.
120. Michalski, A.; Cox, J.; Mann, M., More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC−MS/MS. *J. Proteome Res.* **2011,** *10*, 1785-1793.
121. Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K., UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* **2006,** *20*, 1989-1994.

122. Tsou, C. C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A. C.; Nesvizhskii, A. I., DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **2015,** *12*, 258-+.
123. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** *20*, 3551-3567.
124. Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007,** *4*, 207-214.
125. Wang, G.; Wu, W. W.; Zhang, Z.; Masilamani, S.; Shen, R. F., Decoy Methods for Assessing False Positives and False Discovery Rates in Shotgun Proteomics. *Anal. Chem.* **2009,** *81*, 146-159.
126. Ahrne, E.; Ohta, Y.; Nikitin, F.; Scherl, A.; Lisacek, F.; Muller, M., An improved method for the construction of decoy peptide MS/MS spectra suitable for the accurate estimation of false discovery rates. *Proteomics* **2011,** *11*, 4085-4095.

Chapter 2

Side Chain, Peptide Acidity, and Stoichiometry Effects in Electron Capture Dissociation of Di- and Tri-Valent Metal-Adducted Aβ16 Peptides

2.1 Introduction

As mentioned in Chapter 1, ECD and ETD are of great interest, because these methods offer extensive information about peptide or protein sequence while preserving PTM information. In ECD/ETD experiments, performance is closely related to the charge state[1-2] and mass[2-3] of precursor ions. ECD and ETD fragmentation efficiency is known to be more favorable when precursor ions carry more positive charge. Williams et al. discovered that trivalent metals are capable of supercharging proteins to a higher charge state,[4] which typically is favorable for ECD/ETD. However, the presence of a high valence metal, potentially capable of electron capture, may unfavorably influence ECD/ETD performance. ECD/ETD of such metal-peptide complexes is of interest due to the potential analytical application as well as the gain of further

mechanistic insight into electron based MS/MS. Metal adduction has also been shown to allow retention of highly labile sulfation in ECD of sulfopeptides.[5-6]

Previous research on metal-assisted ECD has shown that different metal ions have different impacts on peptide ECD outcome: the presence of metals such as Li,[1] Na,[1] Mg,[7]Ca,[7] and La[8] yields *c/z* type fragments, or such fragments with metal ions attached, while the presence of metals such as Hg,[9] Cu,[10] and Eu[8] cause peptides to undergo unconventional ECD pathways, to produce other types of fragments, or no fragments at all. Chan and co-workers proposed that metal-electron recombination energy and the energy barrier for proton relocation in the presence of a metal determine the type of ECD behavior associated with a particular metal ion.[9] The coordination sphere provided by metal-binding amino acid side chains has been shown to have the potential to alter ETD outcome for copper-adducted peptides.[11] Further understanding of metal-assisted ECD/ETD is desired to allow application towards metalloprotein analysis, including specific metal binding site determination.[1]

Several peptide and protein properties can influence how these biomolecules bind metal ions. Here, we investigate the effects of amino acid side chain substitutions, peptide acidity, and metal-to-peptide ratio. Histidine side chains are involved in many metal binding pockets and domains, such as zinc finger domains,[12] metal binding pockets of the superoxide dismutase family,[13] and in hemoglobin. The imidazole ring in the histidine side chain can donate lone-pair electrons from its nitrogen atom, and thus contribute to the binding of diverse transition metal ions. Carboxylate groups in acidic side chains and at peptide C-termini can also contribute to the coordination of metal ions. In addition, the mobile protons provided by neutral carboxylic acid groups can serve as alternative electron capture sites.[7] Furthermore, Prell et al. showed, based

on IRMPD spectroscopy,[14] that metal-π interactions with aromatic side chains contribute to the stability of La- and Eu-peptide complexes.

The aforementioned coordination sphere factors all potentially contribute to metal-peptide complex stability. Protein gas-phase stability has previously been reported to influence the variety of *c*/*z*-type product ions observed in intact protein ECD.[15] Non-covalent interactions, such as hydrogen bonds and salt bridges, can prevent complementary fragment ion pairs from separating and thus being observed as separate product ions. Metal coordination can potentially involve strong non-covalent interaction and, therefore, the gas-phase stability of metal-peptide complexes may also influence ECD outcome.

To further investigate ECD of metal-peptide complexes, we chose the metals Mg^{2+}, Ni^{2+}, La^{3+} and Eu^{3+}. Mg^{2+} binds non-specifically to peptides, while Ni^{2+} has selectivity towards histidine side chains. Eu^{3+} has been reported to hinder conventional ECD pathways, while the high valence metal La^{3+} did not prevent ECD from occurring.[8] It was proposed that this ECD outcome difference is due to the higher electron recombination energy of Eu^{3+}. In contrast to this work, which involved peptides that non-selectively bind metals, we chose to examine a versatile metal binding peptide, beta-amyloid 1-16 (Aβ16, DAEFRHDSGYEVHHQK) in both wildtype and mutant forms.

2.2 Experimental

2.2.1 Materials

In order to assess the influence of histidine side chains in metal-peptide complex ECD, the H6, H13, and H14 residues of Aβ16 were mutated into phenylalanine, which has a similar size as

histidine but much lower electron donating potential. This mutated peptide is designated as Aβ16F (sequence: DAEFRFDSGYEVFFQK). Amidation of all acidic sites in Aβ16, including acidic side chains and the C-terminus was also performed to undermine coordination to O-affinitive metal ions, as well as to eliminate mobile protons. This mutated peptide is designated as Aβ16B (sequence: NAQFRHNSGYQVHHQK-NH_2). The F4, H6, Y10, H13, and H14 residues of Aβ16 were also mutated into non-aromatic residues (alanine, lysine, and serine) to eliminate metal-π interactions. This mutated peptide is designated as Aβ16A (sequence: DAEARKDSGSEVKKQK). All synthetic peptides were purchased from GenicBio Ltd.. Metal salts were purchased from Sigma-Aldrich (St. Louis, MO).

2.2.2 Sample Preparation

Metal-peptide complexes were prepared using 0.1-10 μM peptide mixed with 20-200 fold excess of metal acetate or metal bromide salt. The electrospray solution was 1:1 (v/v) methanol-water. Samples were introduced into the mass spectrometer by direct infusion.

2.2.3 FT-ICR Mass Spectrometry

Metal ion-peptide complexes were generated by ESI at a flow rate of 70-150 μL/h. Nebulizing gas (N_2, 1-2 L/h) and drying gas (N_2, 2-5L/h) were utilized. The ESI source temperature was tuned, typically between 120-180 ˚C, to ensure metal-peptide complexes formation. Resulting triply charged metal-peptide complexes were mass-selected, accumulated in a hexapole for 1-8 s, pushed into the ICR cell, via multipole transfer, and trapped in the ICR cell for ECD, activated-ion ECD, or IRMPD, followed by fragment ion detection. ECD in a Bruker SolariX 7 Tesla Q-FT-ICR mass spectrometer was performed with 0.8-1 eV electrons for 100 ms, while ECD in a 9.4 Tesla Q-FT-ICR mass spectrometer at the National High Magnetic Field

Laboratory (NHMFL) was performed with 2-3 eV electrons for 100 ms. The Solarix instrument is equipped with a hollow cathode,[16] operating at 1.6 A, while the NHMFL instrument is equipped with a solid cathode, operating at 3.4 A. Activated-ion ECD was realized by IR laser activation before and/or after ECD, using a 10.6 μm CO_2 laser, firing at 10-15 W for 20-200 ms, depending on metal-peptide complex nature. IRMPD was performed with laser power ranging from 7-17.5 W. For CAD, metal ion-peptide complexes were mass selected by a quadrupole, and collisionally activated in a hexapole collision cell at 2-11 V collision voltage. Product ions generated by CAD in the collision cell were transferred into the ICR cell for detection.

2.2.4 Data Processing and Analysis

Mass spectra generated on the SolariX instrument were acquired using Bruker SolarixControl software. These spectra were collected in broadband mode, m/z range 200-2000 with 512K data points. Data processing was performed with Bruker DataAnalysis software. Mass spectra generated on the NHMFL instrument were acquired by Predator Acquisition software,[17] using the same conditions as for the SolariX instrument. Data processing was performed with Predator Analysis software.[17] Fragment ion peaks within 0.001 m/z, or 10 ppm, error were considered acceptable. The S/N ratio threshold was set to 3. All ECD fragmentation efficiencies were calculated as total fragment ion signal intensity divided by the precursor ion abundance prior to ECD at the same scan number and time. Product ion assignment and efficiency calculations were performed by an in-house written R program. In order to quantitatively compare ECD fragmentation efficiency in the presence of various metals, we calculated relative fragmentation efficiency as follows:

$$Relative\ frag.eff. = \frac{frag.eff.Metallated - frag.eff.Protonated}{frag.eff.Protonated}$$

2.3 Results and Discussion

2.3.1 Metal Binding Efficiency of Mutated Peptides

Different metal ions are known to have different solution-phase peptide affinities.[18] [19] For Aβ16, the solution-phase binding affinity is known to be Cu(II)>Ni(II)>Mg(II). In our experiments, we found that the formation efficiency of metal-peptide complexes via ESI also varies between metals and peptides. Figure 2.1 demonstrates the percentage distribution of various ionic forms (solely protonated, singly-, and doubly metallated) d of wild type and mutant Aβ16, generated with a standard ESI source. We can see that the formation of gas-phase metal-peptide complexes is correlated with the solution-phase binding affinity of these metals with Aβ16: with stronger solution-phase binding affinity, a gas-phase metal-peptide complex is more likely to be generated. Notably, $[peptide + 2M_{II} - H]^{3+}$ complexes are generated, particularly for Cu(II). Such doubly metallated complexes show significantly reduced abundance for Aβ16F,, which may be explained by the lacking histidine side chains eliminating a potential specific binding sites on Aβ16.[19] Under some ESI conditions, we could observe complexes containing more than two metal ions (Figure 2.2). Each type of $[peptide +M_{II} + H]^{3+}$ and $[peptide + 2M_{III} - 3H]^{3+}$ were subjected to ECD to investigate how different metal ions affect Aβ16 ECD, as well as how the same metal ions affect wildtype vs. mutated Aβ16.

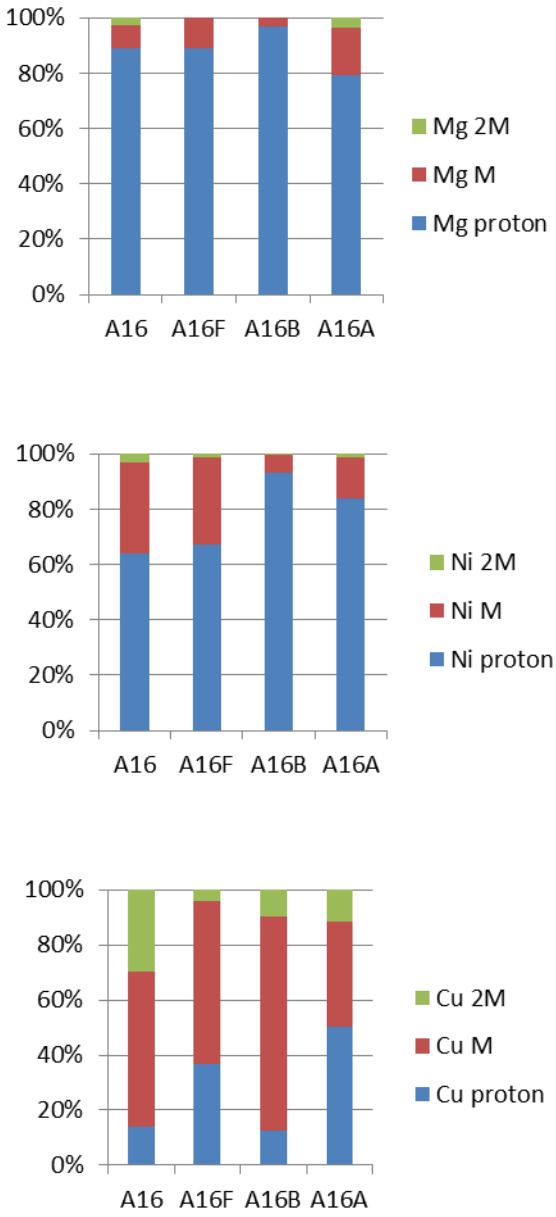


Figure 2.1. Observed distributions of protonated and metal-adducted wild type and mutated Aβ16 following ESI in the presence of Mg, Ni and Cu(II). Proton, M, and 2M represent protonated peptides, singly metallated peptides, and doubly metallated peptides, respectively.

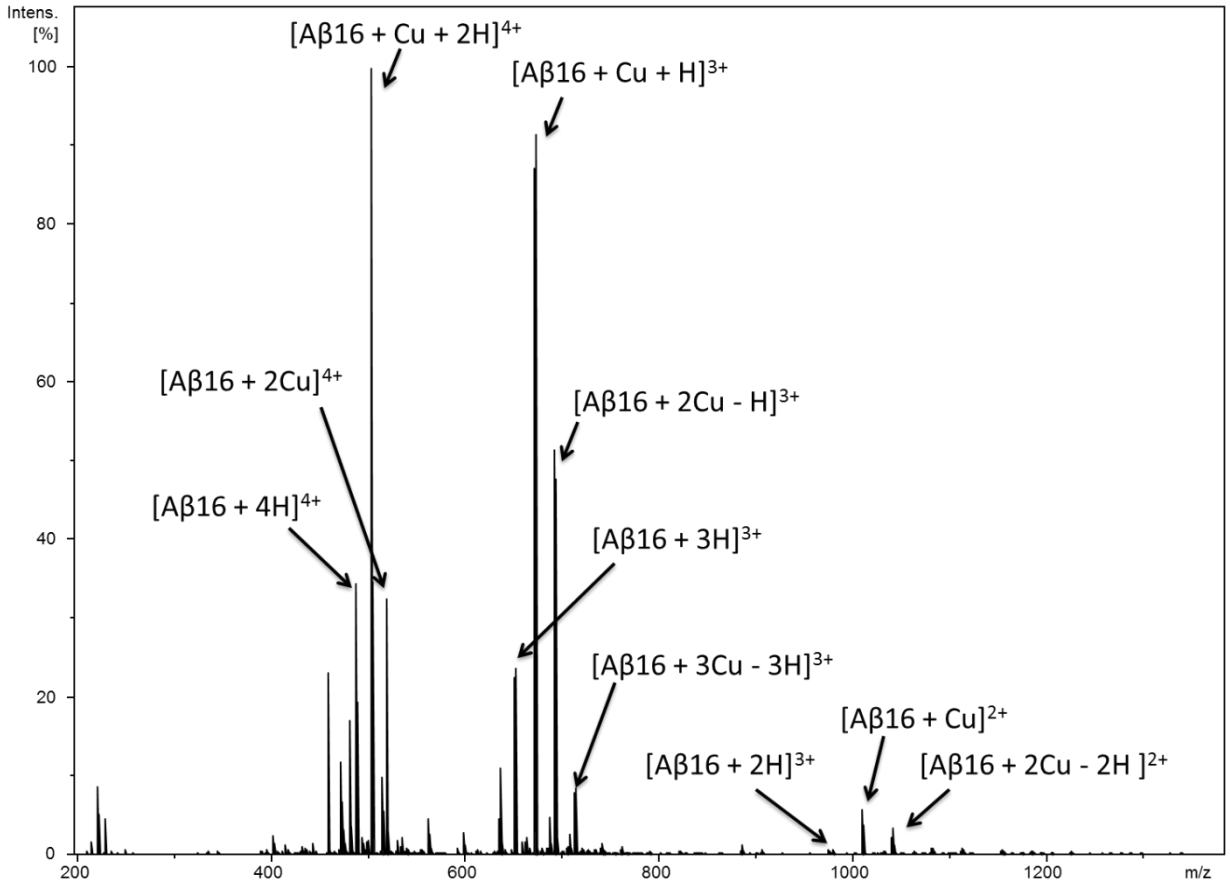


Figure 2.2. ESI-FT-ICR mass spectrum of Cu-adducted Aβ16. Multiple metal-peptide complexes of the same charge state can be observed.

2.3.2 Metal-Peptide Complex ECD

Due to the presence of metal ions as alternative charge carriers, ECD of metal-peptide complexes can show different outcomes compared with ECD of the corresponding protonated peptide: metal-adducted fragments may be observed and metal ions may reduce ECD fragmentation efficiency, even when protons are still present in the complex due to preferred electron capture at the metal.

Figure 2.3 shows ECD of Mg- and Cu(II)-adducted Aβ16.

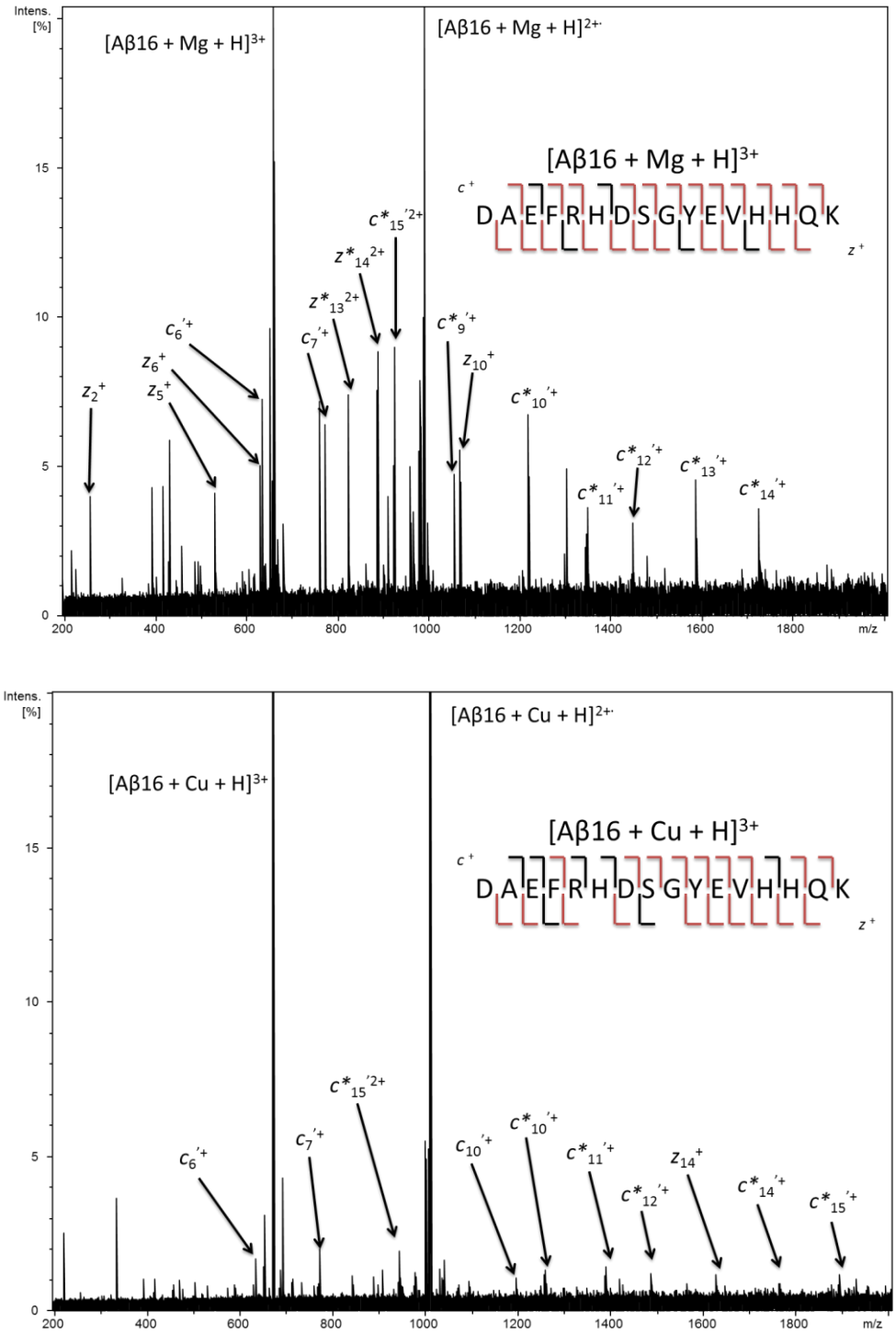


Figure 2.3. ECD FT-ICR tandem mass spectra of triply charged Mg- (top) and Cu- (bottom) adducted Aβ16. The y axes are zoomed to 20% for improved view of the fragment ions. Fragments labeled with * contain a metal ion. Such fragments are labeled in red in the cleavage schemes.

From these spectra we can see that, although both precursor ions contain a divalent metal and a proton, the fragmentation efficiency of Mg-adducted Aβ16 is higher than for Cu-adducted Aβ16. This phenomenon can be explained by both ECD mechanisms described in Chapter 1: in the hot hydrogen mechanism, metal ions serve as competitive electron capture sites. Energy release upon electron capture by metal ions varies depending on the metal. The Mg second ionization energy is 1450.7 kJ/mol, while Cu's second ionization energy is 1957.9 kJ/mol. However, considering that the Cu(II) ion is more than twice as massive as the Mg ion, Mg(II) ions would be more mobile, or "hotter" than Cu(II) ions following electron capture, thus Mg(II) electron capture may allow for more efficient ECD. In the amide superbase mechanism, metal ions may abstract the captured electron from the peptide bond, thus preventing efficient ECD. Because Cu(II) has a higher affinity for the peptide backbone amides compared with Mg(II), electron transfer from a peptide bond to Cu(II) may be more likely. However, the observation of an additional hydrogen in *c* ions indicate that protons participate in the ECD process of these metal(II)- Aβ16 complexes, consistent with a previously proposed mechanism for metal-assisted ECD.[7]

Each metal ion has a unique mass, electron recombination energy, and coordination preference. It is not possible to evaluate these factors of metal ions separately in our experiments. However, previous research has shown that some metal ions are less favorable for fragment generation in ECD than others, including Cu(II),[20] Hg,[9] and Eu.[8] Some other metals, including Li,[1] Na,[1] and Mg,[7] are considered "ECD compatible", meaning their presence do not harm ECD fragmentation efficiency and sequence coverage significantly. There are also metals with an intermediate behavior, such as La and Ni. Previous research on metal-assisted ECD showed that ECD compatibility of divalent metals follows the order Mg>Ni>Cu,[7, 9, 11]

consistent with my results for Aβ16. For trivalent metals, ECD compatibility has been shown to be La>Eu.[8] We found a similar dependence for Aβ16, as demonstrated by the ECD spectra in Figure 2.4.

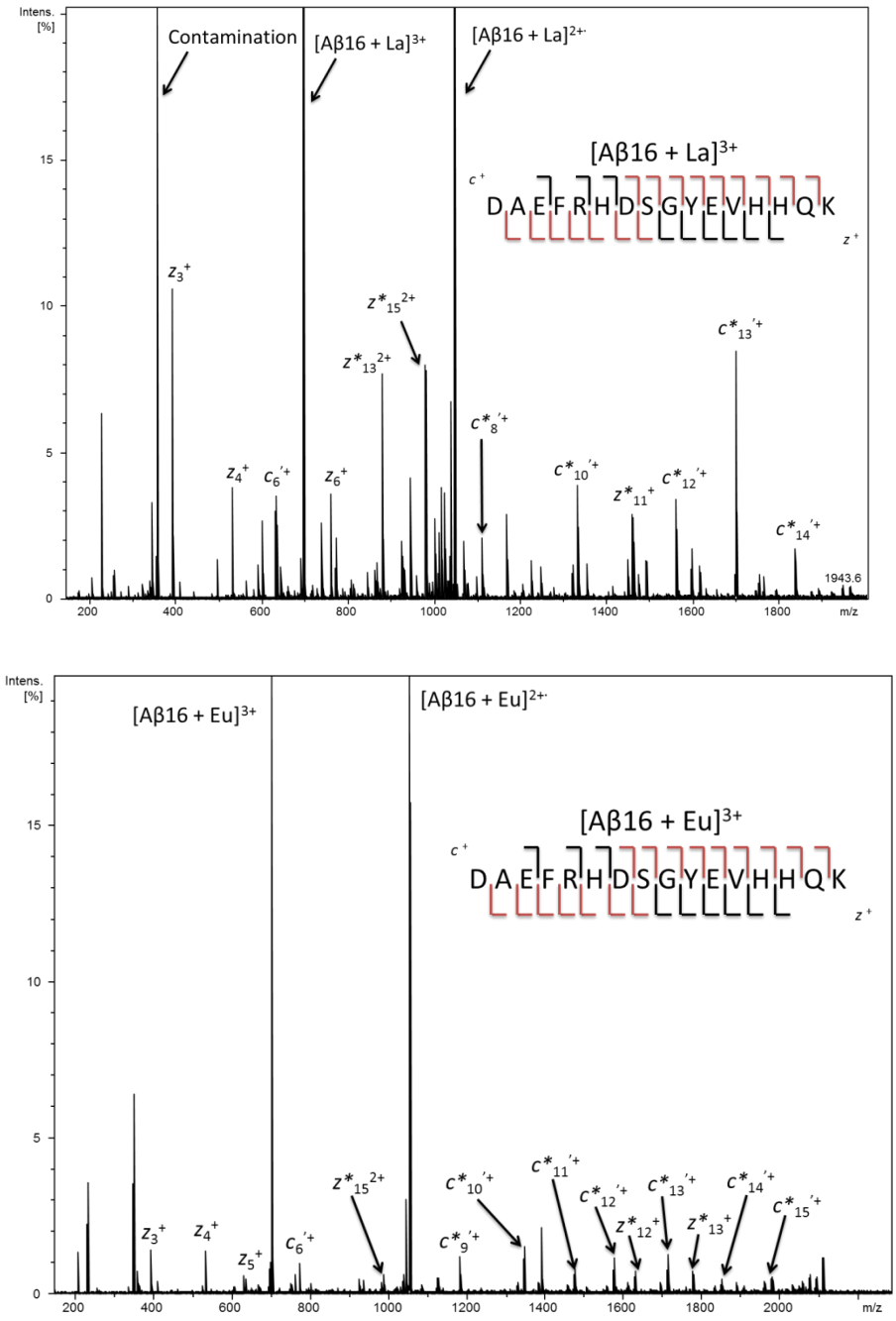


Figure 2.4. ECD FT-ICR spectra of triply-charged La- (top) and Eu- (bottom) adducted Aβ16.

2.3.3 Effects of Histidine Side Chains on Metal-Peptide Complex ECD

Histidine side chains constitute a critical component of various metal binding sites in metalloproteins. The histidine imidazole ring provides lone pair electrons, making it an ideal N-donor coordinator in peptides and proteins. In addition to the well-established binding of transition metals such as nickel(II), copper(II), and manganese(II), histidine is capable of binding high valence metals, such as platinum(IV)[21] and vanadium(V).[22] For Aβ16, histidine is a known metal binding site in vivo [23] and in the gas phase.[24] Therefore, removal of histidine from Aβ16 may undermine its specific metal binding capability, increasing the chance of direct electron capture at the metal ions due to improved electron access, and thus influencing the ECD outcome. We replaced histidine with phenylalanine (to yield Aβ16F), a non-specific binding side chain of similar size, to investigate specific binding effects in metal-peptide complex ECD. Figure 2.5 shows relative ECD fragmentation efficiency of Mg(II)-, Ni(II)-, Cu(II)-, La(III)-, and Eu(III)-adducted wild type Aβ16 and Aβ16F. For trivalent metals, the replacement of histidine with phenylalanine does not appear to change how metals influence peptide ECD fragmentation as no significant change in overall fragmentation efficiency was observed between Aβ16 and Aβ16F. Results for Mg and Ni are inconclusive due to larger error bars, however it appears that all these metals are non-specifically coordinated to the examined Aβ16 peptides. By contrast, for Cu(II), which is known to be a histidine-binding metal and has been shown to yield *b*/*y* ions instead of *c/z* ions upon ECD,[20] the removal of histidine significantly reduced *c/z* type fragmentation efficiency, presumably due to weakened metal coordination causing an increased chance of electron capture by the metal.

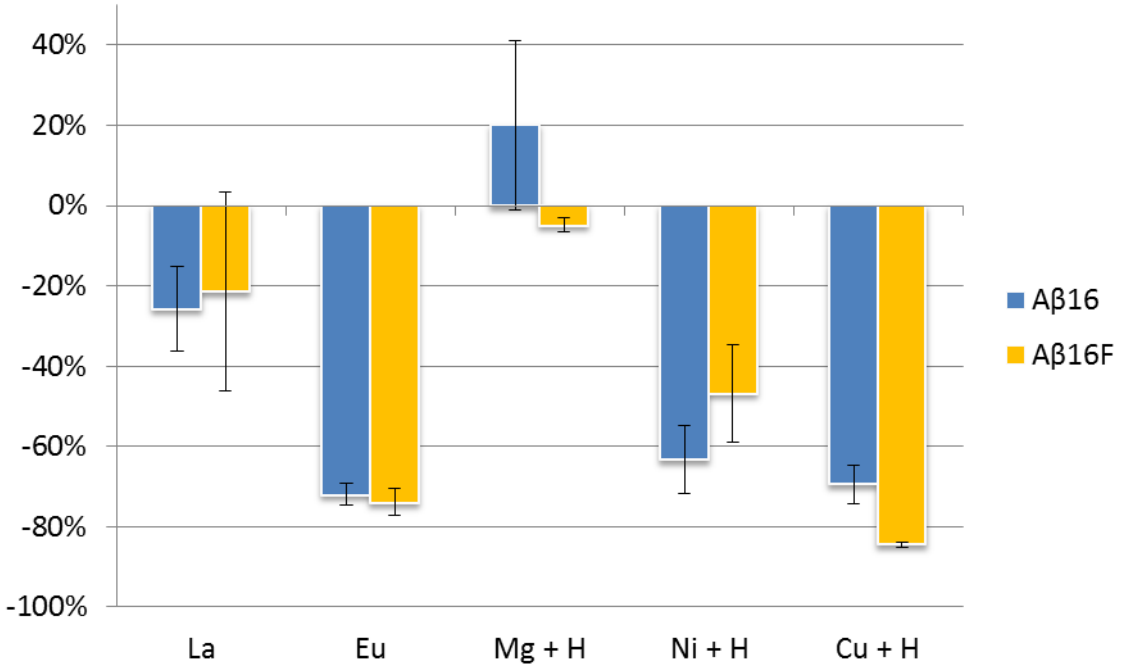


Figure 2.5. Relative ECD fragmentation efficiencies of Aβ16 and Aβ16F, adducted by different metals.

2.3.4 Effects of Peptide Acidity and Mobile Protons on Metal-Peptide Complex ECD

Wild type Aβ16 contains multiple glutamic acid and aspartic acid side chains. In [peptide + $metal^{n+}$]$^{n+}$ type complexes, these side chains may either remain neutral, or deprotonate and coordinate metal ions.[7] In the latter case, measured precursor ion m/z ratios may correspond to zwitterionic species with deprotonated acidic side chains and other protonated sites. The latter mobile protons may capture electrons and induce conventional ECD pathways for metal-peptide complexes. The multiple glutamic acid and aspartic acid side chains in Aβ16 increase the probability of this scenario. By amidating all glutamic acid side chains, aspartic acid side chains, and the Aβ16 C-terminus (to yield Aβ16B), zwitterion formation in the presence of metals should be much less likely than for wild type Aβ16. In order to quantify the influence of this

mutation on metal-peptide complex ECD, we examined the ECD fragmentation efficiency of Mg(II)-, Ni(II)-, Cu(II)-, La(III)-, and Eu(III)-adducted Aβ16 and Aβ16B, as well as of the corresponding two triply protonated peptides.

Figure 2.6 shows the relative ECD fragmentation efficiency of these peptides/metal-peptide complexes. From this Figure, we can see that, with potential mobile protons eliminated, ECD fragmentation efficiencies of metallated peptides unexpectedly increased for La(III)-, Mg(II)-, Ni(II)-, and Cu(II)-adducted peptides. Amidation of all carboxylic acid side chains introduces many N-donors to the peptide, which may contribute to improving the coordination sphere of metal ions, thus switching the ECD behavior in the presence of the aforementioned metals towards that reported for Cu-adducted Aβ peptides in ETD.[11]Improved coordination, shields the metals from direct electron capture, thus potentially explaining the observed unexpected improved ECD outcome.

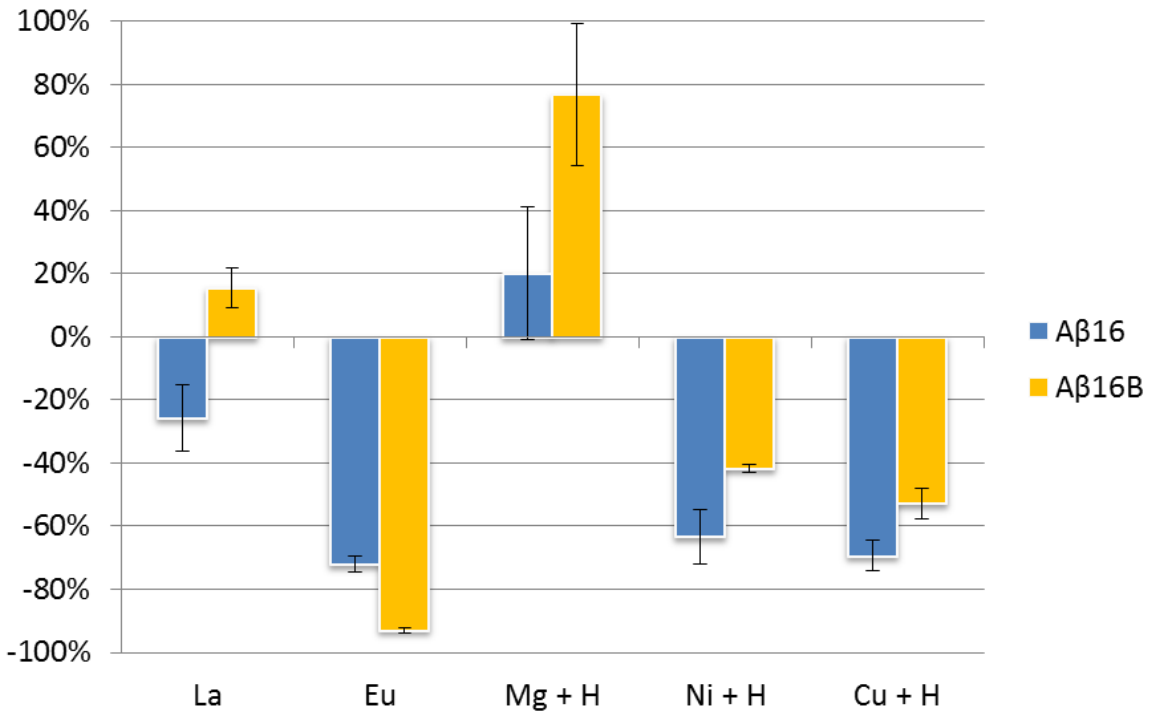


Figure 2.6. Relative ECD fragmentation efficiencies of Aβ16 and Aβ16B bound to different types of metals.

By contrast, for Eu-adducted Aβ16B, ECD fragmentation efficiency was significantly reduced (60% less) compared with ECD of Eu-adducted Aβ16 with 50% lower sequence coverage (the corresponding spectrum is shown in Figure 2.7). This result is similar to previous reports on Eu(III)-adducted peptide ECD.[8] It may be explained by unique properties of Eu(III), compared with the other examined metals. Eu(III) has four f-electrons with asymmetrical wave functions that make coordination less energetically favorable. La(III) also has a higher ion radius than the other metals, including Eu(III), as well as a higher known maximum coordination number (12 vs. 9 for Eu), thus in the presence of many metal-binding side chains, La(III) may be more effectively shielded from direct electron capture, similar to the divalent metals.

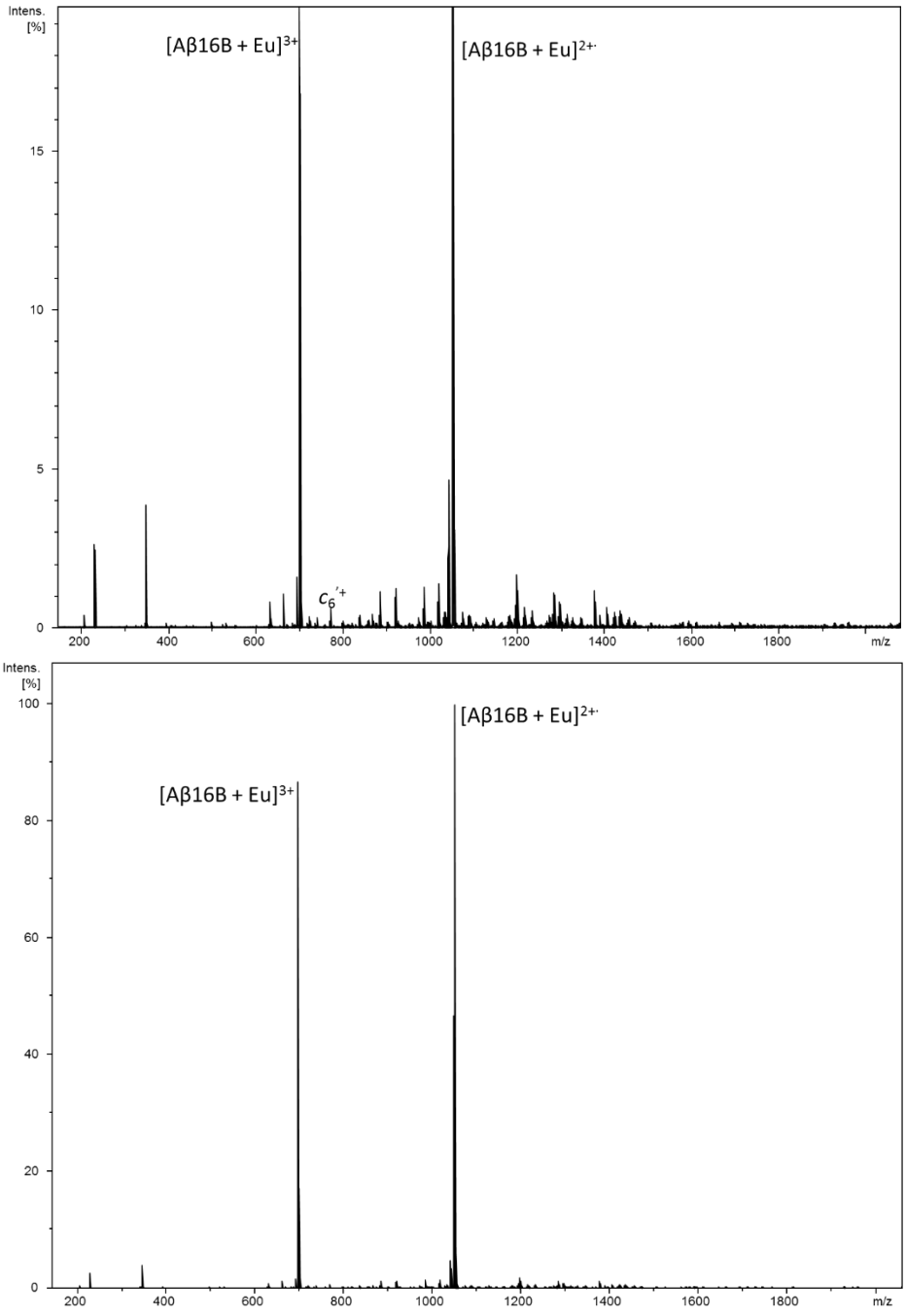


Figure 2.7. ECD FT-ICR tandem mass spectrum of $[A\beta16B + Eu]^{3+}$ (bottom) with a zoomed-in view (top). Few fragments are observed, while the charge-reduced precursor ion is dominant.

2.3.5 Effect of Aromatic Side Chains on Metal-Peptide Complex ECD

Metal-π interactions are well understood in solution-phase organometallics. For gas-phase metal-peptide complexes, Prell et al.[14] demonstrated via IRMPD spectroscopy that backbone carbonyl oxygens, acidic side chain oxygens, and metal-π interactions all contribute to form the lowest energy structure for La(III) binding to small peptides. Aβ16 contains five aromatic side chains: three histidines, a phenylalanine, and a tyrosine, which possibly contribute to the coordination sphere for trivalent metal ions. We mutated these residues into related non-aromatic residues: histidine to lysine, phenylalanine to alanine, and tyrosine to serine to yield Aβ16A. We examined the ECD fragmentation efficiency of protonated and Mg(II)-, Ni(II), La(III)-, and Eu(III)-adducted wild type Aβ16 and Aβ16A. The results are shown in Figure 2.8. With aromatic side chains removed, ECD efficiency dropped for trivalent metal-adducted peptides with a more significant drop for Eu(III)-adducted peptides compared with La(III). However for divalent metal-adducted Aβ16A, ECD efficiency increased. This result indicates that reduced π-donors affect coordination of trivalent metals, and that less coordinated metal ions impede the generation of *c/z* ions in ECD, particularly for Eu(III). For divalent metals, lysine appears to provide improved coordination, thus increasing ECD fragmentation efficiency of metal(II)-Aβ16A complexes. Such effect is especially significant on Mg(II)-adducted Aβ16A.

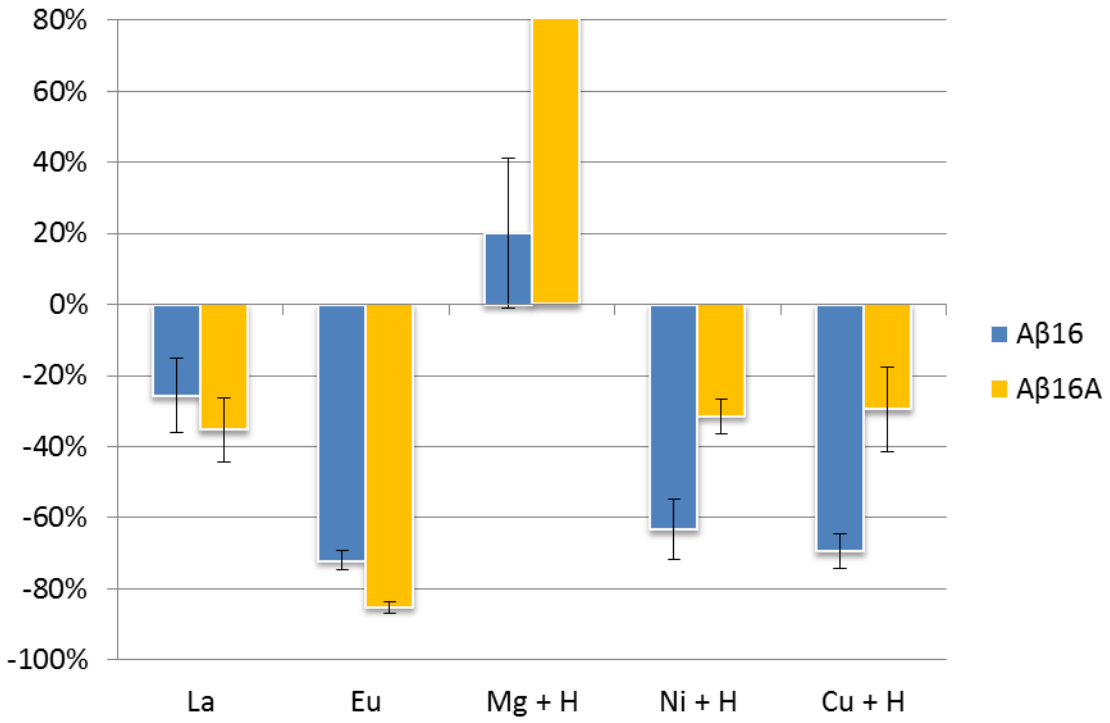


Figure 2.8. Relative ECD fragmentation efficiency of Aβ16 and Aβ16A bound to different types of metals. Notably, the relative fragmentation efficiency of Mg-adducted Aβ16A is 147%.

2.3.6 Precursor Stability Effects in Metal-Peptide Complex ECD

In addition to the copper(II) results shown in section 2.3.1, we observed abundant peptide ions with multiple trivalent metal adducts upon ESI. ECD of doubly La-adducted Aβ16 yields fewer *c/z* types fragments than its singly lanthanum-adducted counterpart (Figure 2.9). For Eu adduction, ECD of doubly Eu-adducted Aβ16 yields no *c/z* types fragments at all (Figure 2.10).

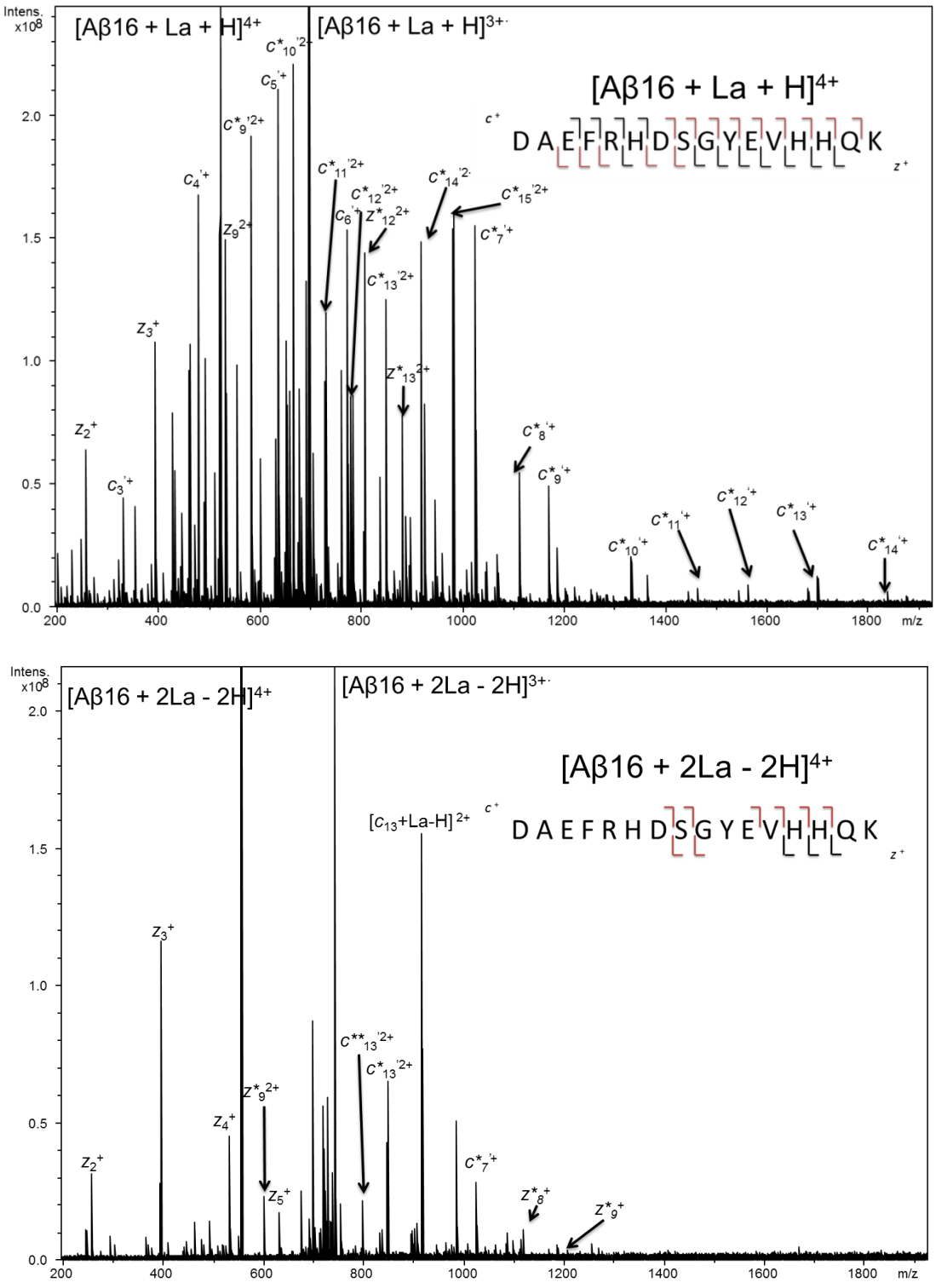


Figure 2.9. ECD FT-ICR tandem mass spectra of $[Aβ16 + La + H]^{4+}$ (top) and $[Aβ16 + 2La - 2H]^{4+}$ (bottom). ECD fragmentation efficiency and sequence coverage dropped significantly due to the second metal ion adduction.

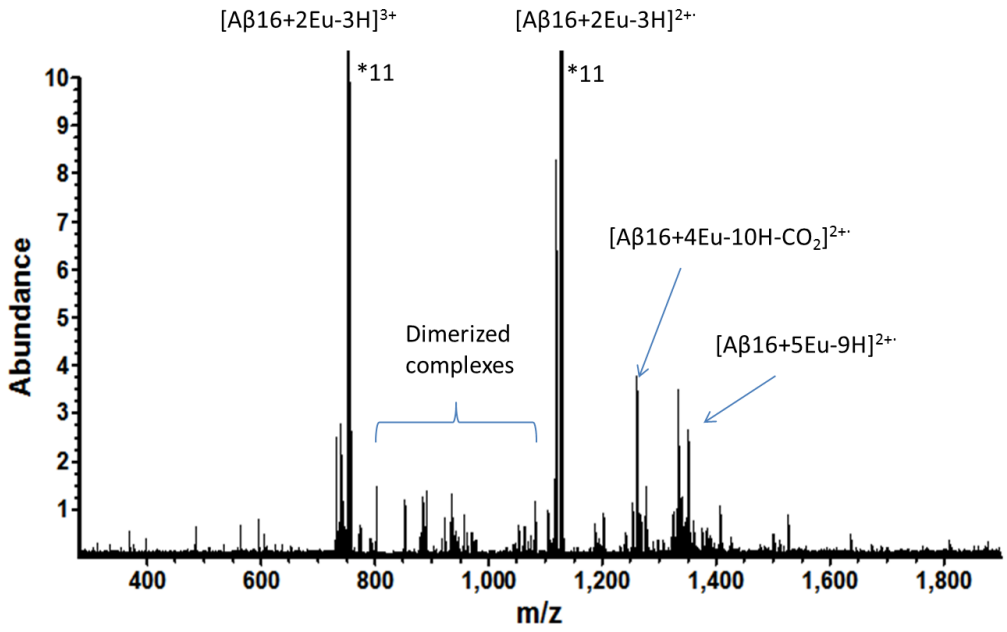


Figure 2.10. ECD FT-ICR tandem mass spectrum of $[Aβ16 +2Eu-3H]^{3+}$. Only neutral losses and dimerized complexes ae observed.

Our first hypothesis for explaining the reduced ECD efficiency upon multiple metal adductions is that, due to the limited size of Aβ16, the second metal ion may be much less coordinated. An uncoordinated metal ion can capture electrons more readily, thus becoming reduced without causing any backbone cleavages. In this scenario, we would expect the doubly metallated complexes to be less stable to heat activation, such as collisions or IR laser irradiation.

In order to further investigate this effect, CAD experiments at various collision energies were conducted on the corresponding complexes. Figure 2.11 shows the precursor ion abundance decrease trend upon CAD as function of collision voltage for Aβ16 adducted with one or two trivalent metals ions. We can see that doubly metal-adducted complexes require higher collisional voltages for dissociation compared with singly metal-adducted complexes. The 2Eu-peptide complex is more stable than the 2La-peptide complex. These results contradict our original hypothesis, suggesting that electron sinking may not be the dominant effect that shuts

down ECD in 2Eu-peptide complexes. Introduction of a second metal ion stabilizes the complex, thus potentially preventing complementary product ion pairs from separating and thus not be detected. Similar effects have been observed in ECD of intact proteins.[15]

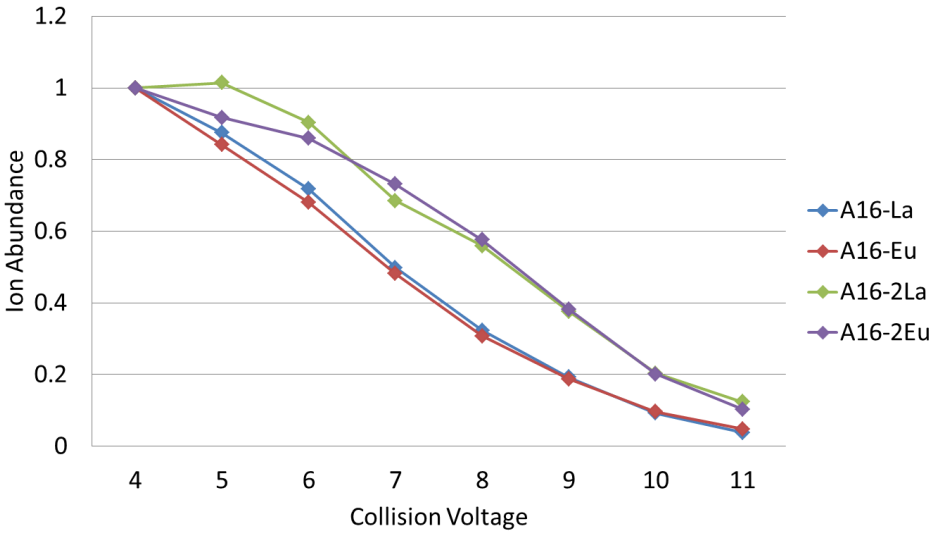


Figure 2.11. Precursor ion abundance loss upon CAD of Aβ16 with one or two lanthanum or europium adducts at different collision energies.

For intact proteins, IR activation to disrupt non-covalent interactions can significantly improve fragmentation efficiency and sequence coverage in ECD. Based on the same idea, we performed IR laser activation prior to and after electron irradiation. Figure 2.12 shows the ECD spectra with and without IR activation. In the latter case, the 2Eu-complex showed comparable ECD performance to the 2La-complex, both in terms of fragmentation efficiency and sequence coverage. This result indicates that metal coordination can be retained in complementary ECD fragment ion pairs and thus preventing the separated products from being observed. IR activation of the 2La-complex did not alter fragmentation efficiency, nor sequence coverage, indicating that other factors, such as electron recombination energy differences, may also contribute to the observed difference in ECD performance for singly vs. doubly-metallated Eu and La complexes.

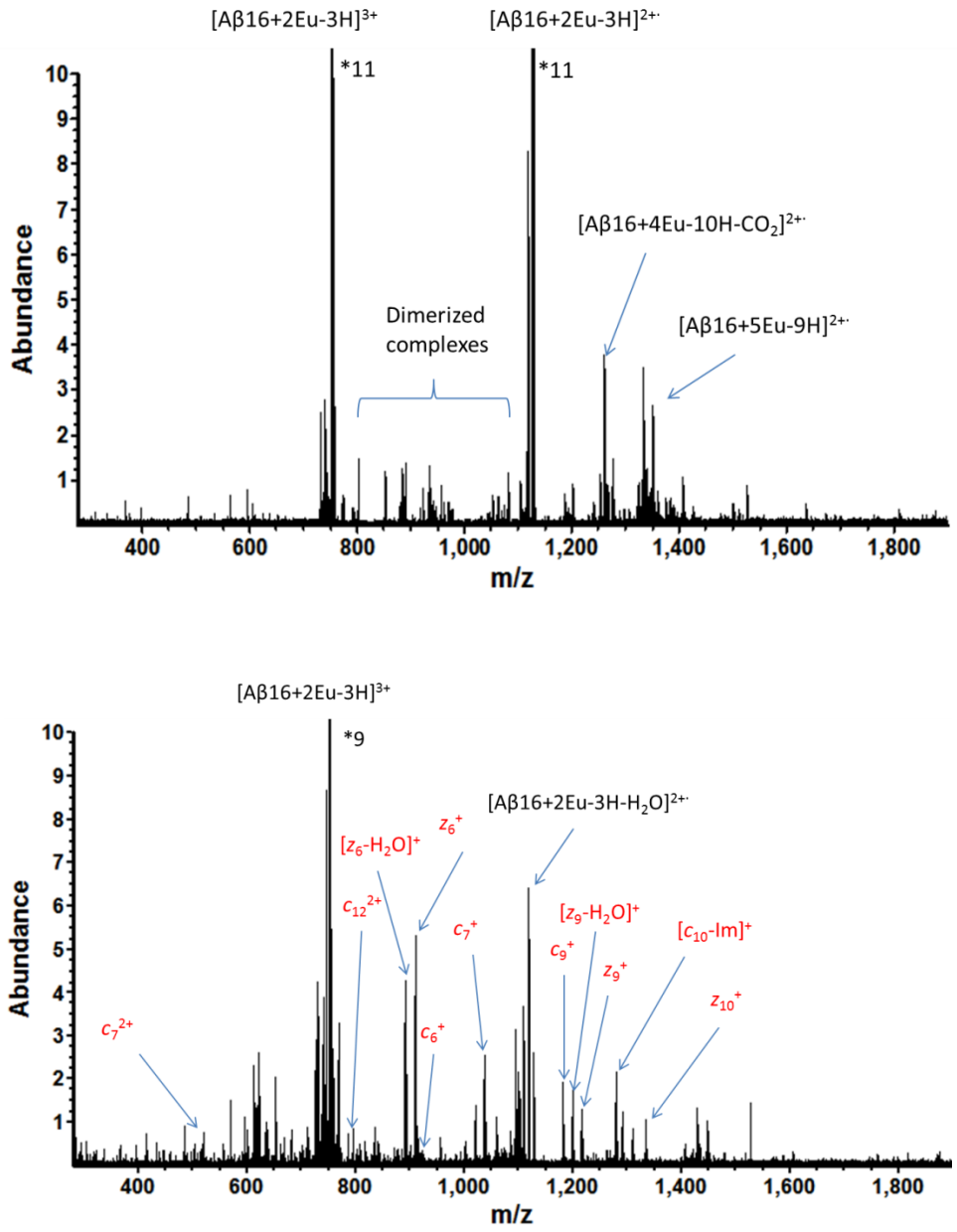


Figure 2.12. ECD FT-ICR tandem mass spectra of triply charged, doubly Eu-adducted Aβ16 without (top) and with (bottom) IR laser activation prior to and after ECD. All observed fragments contain europium. IR activation resulted in observation of *c/z* type fragments whereas no such fragments were observed in the absence of activation.

2.4 Conclusions

From our ECD fragmentation efficiency assessment of wild type and mutant metal-Aβ16 complexes, alteration of the metal binding sphere in metal- Aβ16 complexes affects ECD

outcome. The degree of this effect depends on the metal type and the specific coordination properties of the peptide. Weaker shielding of metals by the peptides resulted in less efficient ECD. This phenomenon may be explained by preferred electron capture at less coordinated metal ions, a pathway that may not result in backbone fragmentation.

For Aβ16 peptides adducted with two metal ions, we observed a dramatic drop in ECD fragmentation efficiency for lanthanum and elimination of ECD backbone fragments for europium, similar to other previously investigated singly Eu-adducted peptides.[8] Through CAD-based gas-phase stability probing, the possibility of uncoordinated metal ions serving as electron capture sites does not appear likely. Instead, increased stability of the 2Eu-Aβ16 complex vs 1Eu-Aβ16 through increased non-covalent interactions appears to explain the shut-down of backbone ECD fragmentation.

The latter coordination-based stability factor in trivalent metal-peptide ECD may also provide insight on our understanding of into ECD of other metal-peptide/protein complexes. Most metal ions have higher electron recombination energy than protons (13.6 eV). For example, the electron recombination energy of Mg(II) is 15.0 eV, Ni(II) is 18.2 eV, La(III) is 19.2 eV, and Eu(III) is 24.9 eV. With increased electron recombination energy, electron capture at metal ions would release additional energy upon electron capture, and thus potentially facilitate peptide dissociation. However, the absence of such effects has been attributed to metal ions serving as electron sinks and thus preclude backbone fragment-generating pathways. The experiments presented here on wild type and mutated Aβ16 peptides in combination with previous experimental results support the latter mechanism.

However, an alternative explanation may be that metal coordination improves the stability of the biomolecule, and thus prevents complementary ECD fragments from being separated and detected. Similar behavior has been observed in intact protein ECD experiments,[15] in which non-covalent interactions between complementary fragments prevent them from being separated and observed. Depending on the metal ion nature, some metals release sufficient energy upon electron capture to overcome the increased stability aided by their coordination. For some other metals, the energy release upon electron capture is insufficient to offset the increased stability due to coordination effects. In the latter case, we mainly observe charge-reduced species. Due to varied coordination ability for different peptides, and varied electron recombination energy for different metals understanding the peptide properties contributing to ECD outcome is important. The effects of metal coordination are in addition to the previous explanation based on direct electron capture at metal ions because electron capture and complementary fragment dissociation may occur separately. In summary, by utilizing metals with high electron recombination energy and limited coordination ability, ECD efficiency may be optimized.

2.5 References

1. Iavarone, A. T.; Paech, K.; Williams, E. R., Effects of Charge State and Cationizing Agent on the Electron Capture Dissociation of a Peptide. *Anal. Chem.* **2004,** *76*, 2231-2238.
2. Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J., Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry *Mol. Cell. Proteomics* **2007,** *6*, 1942-1951.
3. Zhang, H.; Cui, W. D.; Wen, J. Z.; Blankenship, R. E.; Gross, M. L., Native Electrospray and Electron-Capture Dissociation FTICR Mass Spectrometry for Top-Down Studies of Protein Assemblies. *Anal. Chem.* **2011,** *83*, 5598-5606.
4. Flick, T. G.; Williams, E. R., Supercharging with Trivalent Metal Ions in Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1885-1895.
5. Liu, H.; Hakansson, K., Electron Capture Dissociation of Tyrosine O-Sulfated Peptides Complexed with Divalent Metal Cations. *Anal. Chem.* **2006,** *78*, 7570-7576.
6. Liu, H. C.; Yoo, H. J.; Hakansson, K., Characterization of Phosphate-Containing Metabolites by Calcium Adduction and Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2008,** *19*, 799-808.

7. Fung, Y. M. E.; Liu, H.; Chan, T. W. D., Electron Capture Dissociation of Peptides Metalated with Alkaline-Earth Metal Ions. *J. Am. Soc. Mass Spectrom.* **2006,** *17*, 757-771.
8. Flick, T. G.; Donald, W. A.; Williams, E. R., Electron Capture Dissociation of Trivalent Metal Ion-Peptide Complexes. *J. Am. Soc. Mass Spectrom.* **2013,** *24*, 193-201.
9. Chen, X. F.; Chan, W. Y. K.; Wong, P. S.; Yeung, H. S.; Chan, T. W. D., Formation of Peptide Radical Cations (M+center dot) in Electron Capture Dissociation of Peptides Adducted with Group IIB Metal Ions. *J. Am. Soc. Mass Spectrom.* **2011,** *22*, 233-244.
10. Liu, H.; Hakansson, K., Divalent Metal-Peptide Interactions Probed by Electron Capture Dissociation of Trications. *J. Am. Soc. Mass Spectrom.* **2006,** *17*, 1731-1741.
11. Dong, J.; Vachet, R., Coordination Sphere Tuning of the Electron Transfer Dissociation Behavior of Cu(II)–Peptide Complexes. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 321-329.
12. Wolfe, S. A.; Nekludova, L.; Pabo, C. O., DNA recognition by Cys(2)His(2) zinc finger proteins. *Annu. Rev. Biophys. Biomolec. Struct.* **2000,** *29*, 183-212.
13. Richardson, J. S.; Thomas, K. A.; Rubin, B. H.; Richardson, D. C., Crystal-structure of bovine Cu, Zn superoxide-dismutase at 3A resolution - chain tracing and metal ligands *Proc. Natl. Acad. Sci. U. S. A.* **1975,** *72*, 1349-1353.
14. Prell, J. S.; Flick, T. G.; Oomens, J.; Berden, G.; Williams, E. R., Coordination of Trivalent Metal Cations to Peptides: Results from IRMPD Spectroscopy and Theory. *J. Phys. Chem. A* **2010,** *114*, 854-860.
15. Horn, D. M.; Ge, Y.; McLafferty, F. W., Activated Ion Electron Capture Dissociation for Mass Spectral Sequencing of Larger (42 kDa) Proteins. *Anal. Chem.* **2000,** *72*, 4778-4784.
16. Tsybin, Y. O.; Witt, M.; Baykut, G.; Kjeldsen, F.; Hakansson, P., Combined infrared multiphoton dissociation and electron capture dissociation with a hollow electron beam in Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003,** *17*, 1759-1768.
17. Blakney, G. T.; Hendrickson, C. L.; Marshall, A. G., Predator data station: A fast data acquisition system for advanced FT-ICR MS experiments. *Int. J. Mass Spectrom.* **2011,** *306*, 246-252.
18. Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G., Metal Chelate Affinity Chromatography, A New Approach to Protein Fractionation. *Nature* **1975,** *258*, 598-599.
19. Hochuli, E.; Dobeli, H.; Schacher, A., New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighboring Histidine-residues. *Journal of Chromatography* **1987,** *411*, 177-184.
20. Chen, X. F.; Wang, Z.; Li, W.; Wong, Y. L. E.; Chan, T. W. D., Effect of structural parameters on the electron capture dissociation and collision-induced dissociation pathways of copper(II)-peptide complexes. *Eur. J. Mass Spectrom.* **2015,** *21*, 649-657.
21. Jovanovic, S.; Petrovic, B.; Bugarcic, Z. D., UV-Vis, HPLC, and H-1-NMR studies of the substitution reactions of some Pt(IV) complexes with 5 '-GMP and L-histidine. *J. Coord. Chem.* **2010,** *63*, 2419-2430.
22. Messerschmidt, A.; Wever, R., X-ray structure of a vanadium-containing enzyme: Chloroperoxidase from the fungus Curvularia inaequalis. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93*, 392-396.
23. Multhaup, G.; Schlicksupp, A.; Hesse, L.; Beher, D.; Ruppert, T.; Masters, C. L.; Beyreuther, K., The amyloid precursor protein of Alzheimer's disease in the reduction of copper(II) to copper(I). *Science* **1996,** *271*, 1406-1409.

24. Ma, G. L.; Huang, F.; Pu, X. W.; Jia, L. Y.; Jiang, T.; Li, L. Z.; Liu, Y. Z., Identification of PtCl2(phen) Binding Modes in Amyloid-beta Peptide and the Mechanism of Aggregation Inhibition. *Chem.-Eur. J.* **2011,** *17,* 11657-11666.

Chapter 3

Effects of Salt Anions as Charge Carriers in Peptide Electron Detachment Dissociation

3.1 Introduction

As introduced in Chapter 1, EDD[1] and niECD[2] are valuable alternative MS/MS methods that operate in negative ion mode. Both of these two ion-electron reactions feature radical-driven processes and improved PTM compatibility [3-5] compared with CAD in negative ion mode. Product ions from these MS/MS methods are *a/x*-type ions for EDD[6] and *c/z*-type ions for niECD,[2] both types are complementary to the *c*/*y*-type ions produced in negative ion CAD (although negative ion CAD also often have unpredictable outcomes[7]). Negative ion mode operation allows improved detection of acidic biomolecules and sulfate modifications are also more stable in negative mode.[5] However, both of EDD and niECD still have drawbacks preventing them from similarly wide application as, e.g., ECD. For some peptides, charge-reduced, non-dissociated precursor ions, along with neutral losses such as CO_2 or H_2O loss,[8] are

dominant in EDD spectra. As shown in Figure 3.1, in EDD of angiotensin I, the major product ions are charge reduced precursor and CO_2 neutral losses. Fragmentation efficiency is very low due to competition with the neutral loss pathway. Because carboxylic acid groups in peptides are the most likely deprotonation sites, preferred electron detachment from deprotonated carboxylic acid groups likely explain the preference for this neutral loss.[1, 6, 9]

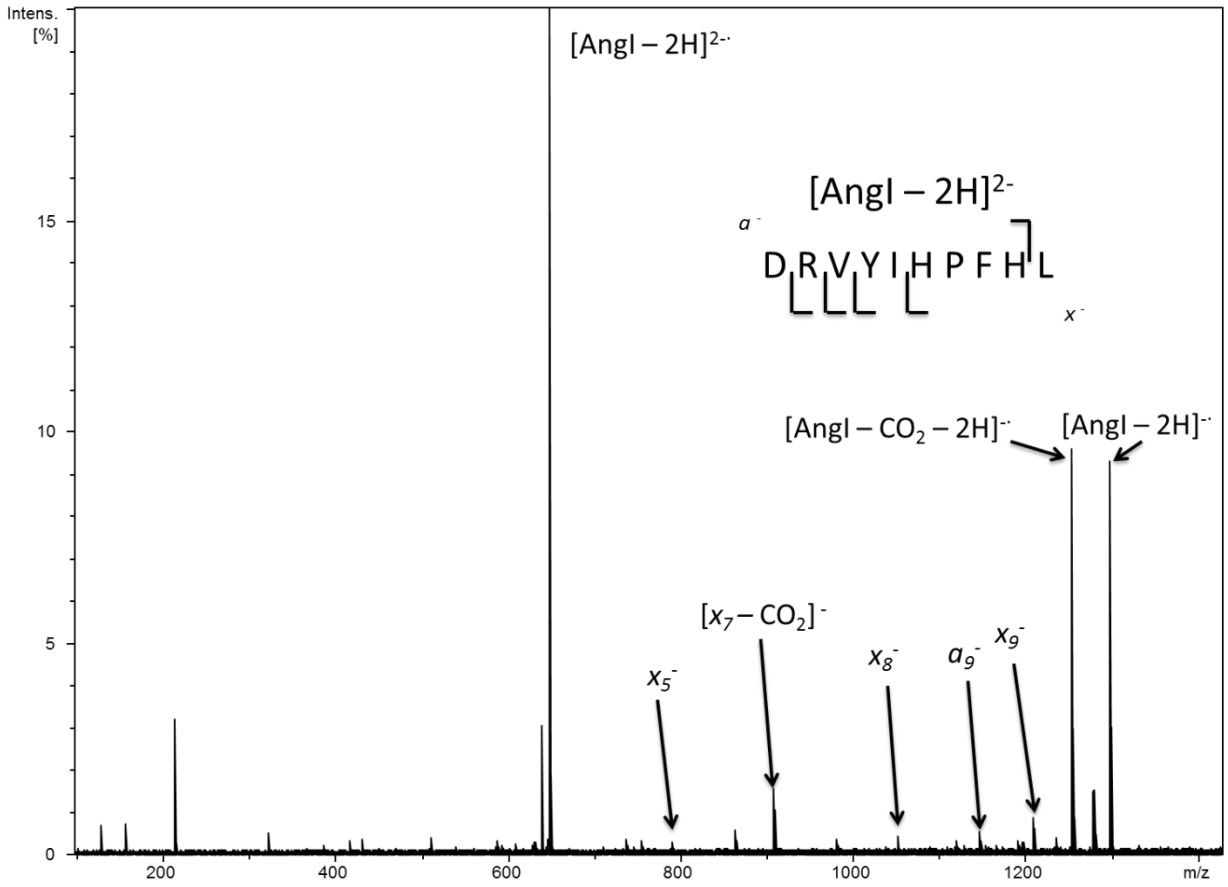


Figure 3.1. EDD spectrum of angiotensin I, zoomed in to 20% abundance to more clearly show the resulting fragment ions.

In niECD, the electron capture process is proposed to require a zwitterionic gaseous peptide structure.[2] Generation of zwitterions in the gas phase requires both acidic and basic moieties within peptides, preferably in multiple positions. Peptides also need to have a low overall charge state to more readily accept incoming electrons. Not all peptides satisfy these criteria.

In an effort to solve the neutral loss issue in EDD and the peptide incompatibility issue in niECD, use of alternative charge carriers, i.e. salt anions, is investigated. In this Chapter, we focus on improving EDD with salt anion adduction. We generated various peptide-anion complexes using negative mode ESI and assessed their performance in EDD. Assessment criteria include sequence coverage and fragmentation efficiency.

3.2 Experimental

3.2.1 Materials

Peptides used in this chapter include: angiotensin I (H-DRVYIHPFHL-OH), sulfated hirudin 55-65 (H-DFEEIPEEsYLQ-OH), free-acid LHRH (pEHWSYGLRPG-OH), sulfated cholecystokinin fragment 26-33 (CCKS, H-DsYMGWMDF-NH2), and phosphorylated DAM1 221-241 (H- SFVLNPTNIGMpSKSSQGHVTK-OH). CCKS and hirudin were purchased from Advanced ChemTech (Louisville, NY). All other peptides were from Sigma-Aldrich (St. Louis, MO).

3.2.2 Sample preparation

Peptide-anion complexes were prepared using 0.1-10 μM peptide mixed with a 20-fold excess of ammonium salt. Spray solutions consisted of 1:1 (v/v) methanol-water or acetonitrile-water. Samples were introduced into the mass spectrometer by direct infusion.

3.2.3 FT-ICR Mass Spectrometry

All experiments in this Chapter were performed on a Bruker Solarix 7 Tesla Q-FT-ICR mass spectrometer. Peptide-anion complexes were generated by negative mode ESI of sample

solutions at flow rates of 120-150 μL/h. Nebulizing gas (N_2, 1-2 L/h) and drying gas (N_2, 2-5 L/h) were applied in the ESI source. Source temperature was tuned for each compound to maximize complex formation, usually between 100-150 ˚C. The resulting doubly or singly charged peptide-anion complexes were mass-selected, accumulated in a multipole for 1-8 s, pushed into the ICR cell via high voltage ion optics, and trapped in the ICR cell for EDD or niECD. EDD was performed with ~20 eV electrons from a hollow cathode for 2 s. niECD was performed with ~4.5 eV electrons for 5 s. The cathode operated at either 1.7 or 1.6 A. For each set of fragemtation efficiency data, EDD/niECD experimental parameters were the same to ensure consistency. IR laser activation was used for AI-EDD and AI-niECD. The laser fired at 15 W with 0.1 s duration after the electron pulse event.

3.2.4 Data Processing and Analysis

Mass spectra were collected using Bruker SolarixControl software in broadband mode, m/z range 200-2000 with 512K data points. Data processing was performed with Bruker DataAnalysis software. Fragment ion peaks within 0.001 m/z or 10 ppm error were considered acceptable. The S/N ratio threshold was set to 3. All EDD fragmentation efficiencies were calculated as total fragment ion abundance divided by the precursor ion abundance prior to EDD for the same scan numbers and time. This processing of MS/MS efficiency was performed with an in-house R program. In order to quantitatively compare EDD fragmentation efficiency, we defined relative fragmentation efficiency as follows:

$$Relative\ frag.eff. = \frac{frag.eff.Metallated - frag.eff.Protonated}{frag.eff.Protonated}$$

Only predictable fragments (*a/x*- ions for EDD) were included in the calculations.

3.3 Results and Discussion

3.3.1 Anion Adduction Accessibility

Anion adduction has been observed in both positive[10] and negative[11] ESI. Such adduction is often considered undesired in mass spectrometry analysis because such analytical artifacts can lead to false positives. Thus, anion adduction must be considered in database searching strategies, as demonstrated in the Appendix of this thesis for protein sequence variant analysis. In a model proposed by Liu et al.,[11] anion binding sites were proposed to be guided by the relative gas phase basicity of peptide moieties and anions. Carboxyl groups, which typically carry negative charges, were proposed to be major binding sites for these anions.

Based on this model, in order to generate stable peptide-anion complexes for ion-electron reactions, three peptides of different acidity (hirudin, LHRH, and angiotensin I) were chosen for anion accessibility study. The ratio of acidic moieties to basic moieties in these three peptides scale as hirudin>LHRH>angiotensin I. Their EDD and niECD compatibility were found to also follow the same trend: hirudin>LHRH>angiontensin I. We chose common salt anions of different acidity as adduction candidates, including halogens (Cl^-, Br^-, I^-), perchlorate (ClO_4^-), sulfate (SO_4^{2-}), and bicarbonate (HCO_3^-).

From negative mode ESI-MS spectra, anion adduction accessibility test results are summarized in Table 3.1. Generally, chloride (Cl^-), bromide (Br-), and sulfate (SO_4^{2-}) ions are more prone to adduct to peptide anions in the gas phase. The gas phase peptide acidity does not seem to affect adduction accessibility as significantly as the nature of salt anions.

	Cl^-	Br^-	I^-	ClO_4^-	SO_4^{2-}	HCO_3^-
Angiotensin	√	√			√	
LHRH	√	√		√	√	
Hirudin		√			√	

Table 3.1 Anion adduction accessibility of common salt anions for three peptides. Check marks represent >10% relative peptide abundance observed in adducted form.

Noticeably, we observed multiple anion-adducted peptide complex ions. Figure 3.2 shows the mass spectrum of bromide-adducted LHRH complexes. According to the theory by Liu et al.,[11] preferred anion adduction sites on peptides are moieties with similar gas phase basicity as the anion (Figure 3.3). Because LHRH only has one acidic site, the experimental observation of multiple adducts suggests that bromide ions are compatible with multiple peptide moieties for anion adduction. This hypothesis also explains why bromide appears to have wide compatibility with different peptides.

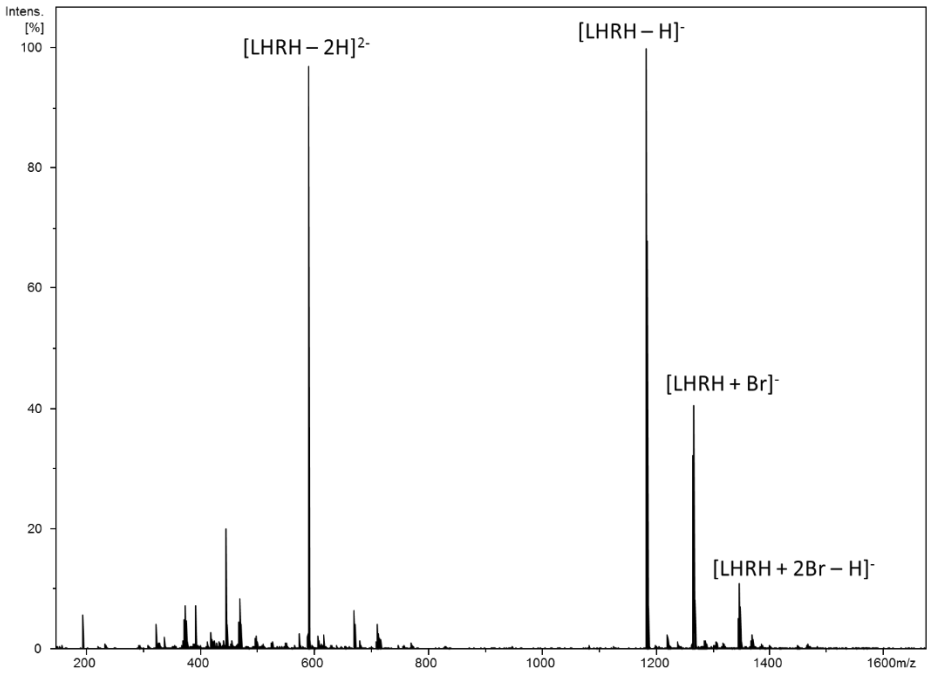


Figure 3.2. Mass spectrum of bromide-adducted LHRH complexes.

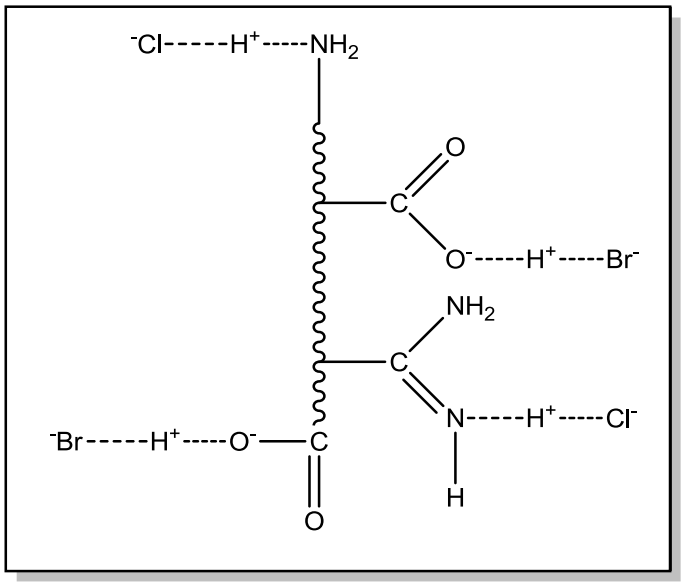


Figure 3.3. Proposed peptide-anion complex formed in a negatively charged electrospray droplet.

3.3.2 EDD of Anion-adducted Peptides

With available anion-peptide complexes, we tested how anion-adduction influences EDD performance. Two key factors were applied for performance assessment: sequence coverage and fragmentation efficiency. Because EDD is an electron-recombination radical process triggered by high-energy electrons, it can cleaves all possible peptide backbone bonds, thus providing relatively complete sequence coverage for short peptides. The ideal performance-enhancing anion adduct would either retain or improve sequence coverage, and increase fragmentation efficiency.

Figure 3.4 shows the EDD sequence coverage for deprotonated, Cl^-, Br^-, SO_4^{2-}, and ClO_4^- adducted LHRH. We chose LHRH for this assessment because it has the widest anion-adduction availability.

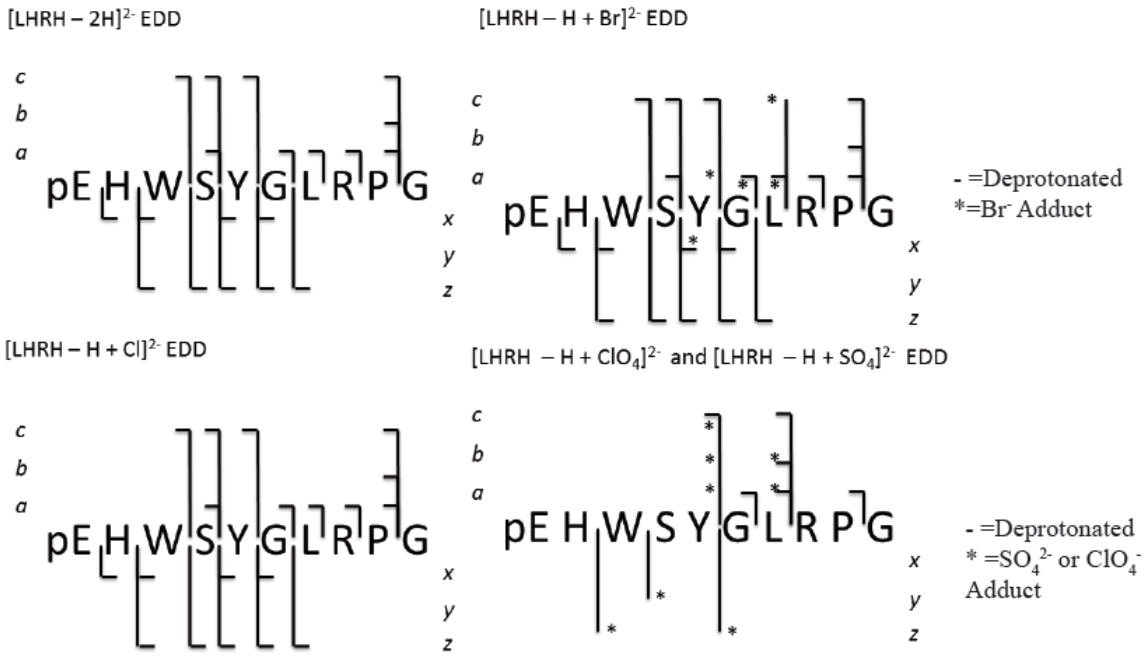


Figure 3.4 EDD Fragmentation patterns for different anion-adducted forms of LHRH.

From these sequence coverage results, perchlorate and sulfate adduction have a negative impact on EDD sequence coverage. As shown in Figure 3.5, both sequence coverage and fragmentation efficiency drop drastically when perchlorate ions are present. *x*-type ions, which constitute typical EDD product ions, are completely eliminated. The same phenomenon is also observed with sulfate-adducted peptide anions. These experimental results indicate that large oxygen-containing salt anions may be incompatible with EDD. Because the purpose of our research is to improve ion-electron reaction performance by using anion adduction, we did not design any additional experiments to further explore the underlying reason of the observed incompatibility. A reasonable speculation may be that multiple oxygen atoms can form a strong negative charge center that undergoes preferentially electron detachment, and/or interfere with productive radical rearrangement processes on the peptide backbone.

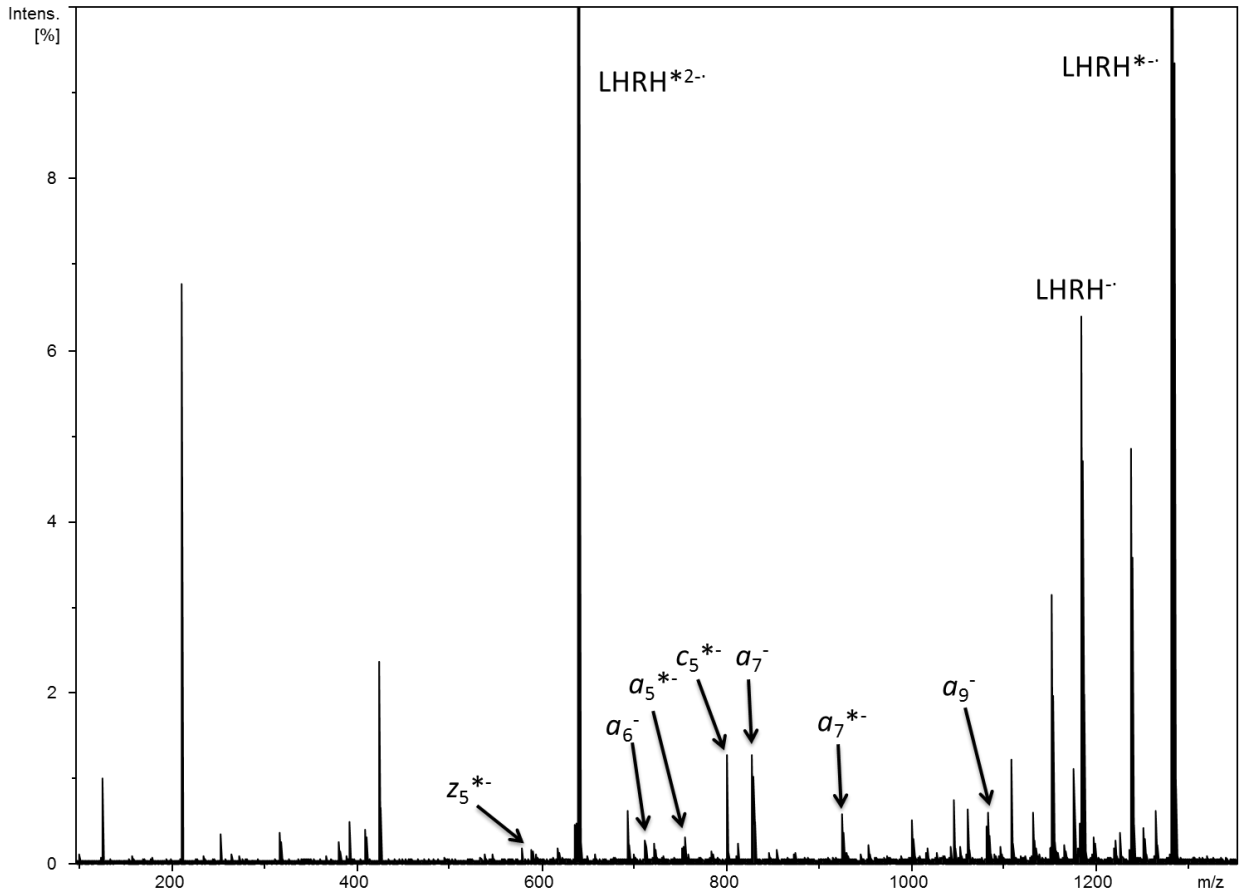


Figure 3.5. EDD spectrum of doubly charged perchlorate (ClO_4^-) adducted LHRH, zoomed-in to 10% abundance to more clearly show the fragment ions. Ions labeled with asterisks (*) contain perchlorate adducts.

For halogen-adducted peptides, both chloride and bromide adduction maintained EDD sequence coverage. In regards to spectral complexity, chloride-adducted peptides contain less chloride-adducted fragments than bromide-adducted peptides. However, the chloride ion has inferior ion adduction efficiency compared with bromide (see Figure 3.6). To achieve the same precursor ion abundance, chloride-adducted peptides require longer ion accumulation time or higher sample concentration than the same bromide-adducted peptide. Therefore, bromide was chosen as the most promising EDD-compatible ion for further investigation.

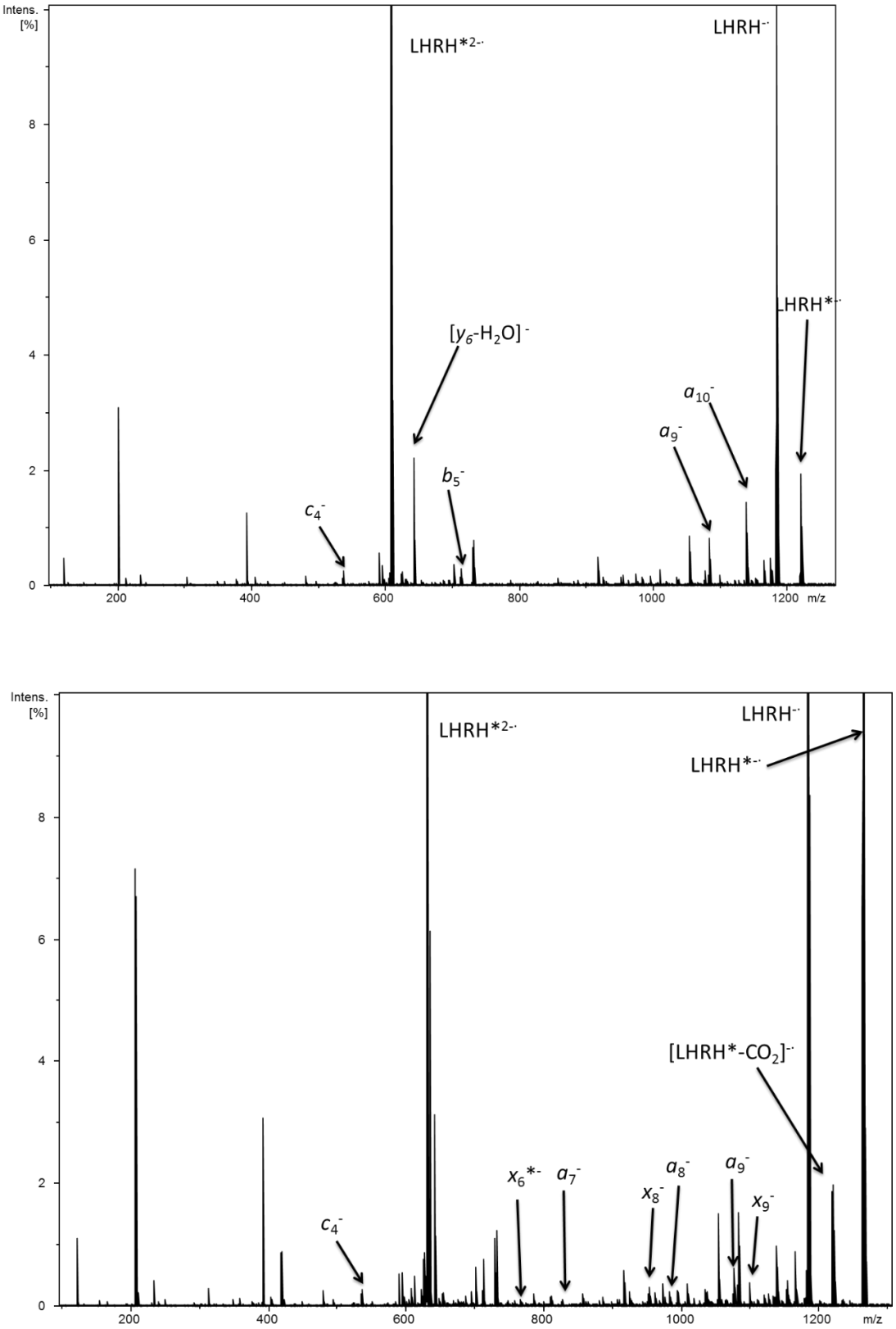


Figure 3.6. EDD spectra of Cl^- (top) and Br-adducted LHRH.

From Figure 3.6 we can also see that, for bromide-adducted LHRH, EDD product ion abundance of the charge reduced precursor and neutral losses both increase compared with chloride-adducted LHRH. Bromide adduction was also explored for other peptides and a similar effect was observed. The phosphopeptide DAM1 can stably attach multiple bromide ions and is used as a model here to demonstrate how bromide adduction influences the abundance of MS/MS product ions. Figure 3.7 shows the change in ion-electron reaction product ion abundance for DAM1 upon adduction with bromide. From this figure we can see that structural fragmentation yielding backbone sequence information is not affected by bromide adduction, while the charge reduced precursor ion abundance increased significantly with increased numbers of bromide adducts. This phenomenon suggests that bromide adduction facilitates electron detachment, possibly by adding an additional reaction pathway. However, such a reaction pathway does not appear to cause backbone cleavages, or the anion-peptide complex is more stable than the peptide alone so that complementary fragment pairs are not separated and thus observed. The latter phenomenon was previously observed in top down ECD,[12-13] in which noncovalent interactions can prevent fragment ions from being observed. IR laser activation was applied to increase the internal energy of the charge-reduced precursor ions, thus breaking noncovalent interactions and allowing fragment ions to separate and be observed.

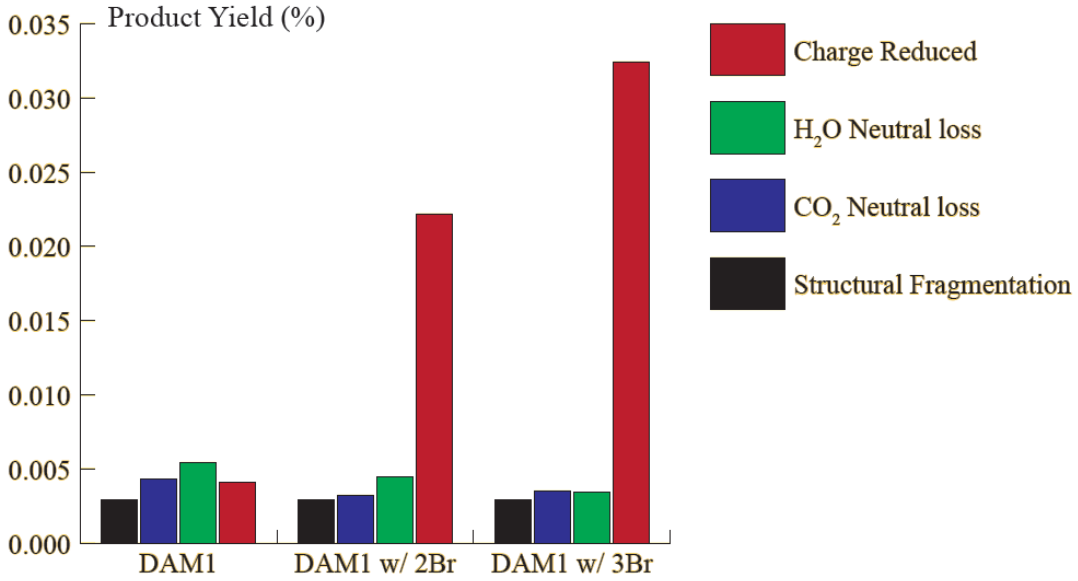


Figure 3.7. Quantified EDD reaction efficiency for the bromide-adducted phosphopeptide DAM1.

3.4 Conclusion

Results from this Chapter show that EDD of salt anion adducted peptides is possible for certain types of salt anions. For chloride and bromide ions, EDD fragmentation patterns of the corresponding peptide anion complexes are similar to those of deprotonated peptide anions. For other anions, such as perchlorate and sulfate ions, EDD is not operable upon anion adduction.

Chloride and bromide adduction did not increase peptide anion fragmentation efficiency. Instead of this expected abundance increase of backbone fragment ions, we observed significant abundance increase of charge reduced precursor ions. Therefore, no additional information or improved sensitivity was achieved through the adduction of salt anions. There are two possible causes of this observation: complementary fragment ion pairs may be bound by salt bridges, or salt anions enable an alternative non-dissociative electron detachment pathway. In the first case, fragments should be released upon IR laser activation, i.e., AI-EDD experiments. However, due

to instrument issue, such experiments remain to be performed, see Chapter 5 for detailed experimental plans.

3.5 References

1. Budnik, B. A.; Haselmann, K. F.; Zubarev, R. A., Electron Detachment Dissociation of Peptide Di-anions: an Electron-hole Recombination Phenomenon. *Chem. Phys. Lett.* **2001,** *342*, 299-302.
2. Yoo, H. J.; Wang, N.; Zhuang, S.; Song, H.; Hakansson, K., Negative Ion Electron Capture Dissociation: Radical-Driven Fragmentation of Charge-Increased Gaseous Peptide Anions. *J. Am. Chem. Soc.* **2011,** *133*, 16790-16793.
3. Kleinnijenhuis, A. J.; Kjeldsen, F.; Kallipolitis, B.; Haselmann, K. F.; Jensen, O. N., Analysis of Histidine Phosphorylation Using Tandem MS and Ion-Electron Reactions. *Anal. Chem.* **2007**, 7450-7456.
4. Kweon, H. K.; Hakansson, K., Metal Oxide-Based Enrichment Combined with Gas-Phase Ion-Electron Reactions for Improved Mass Spectrometric Characterization of Protein Phosphorylation. *J. Proteome Res.* **2008,** *7*, 745-755.
5. Hersberger, K. E.; Hakansson, K., Characterization of O-Sulfopeptides by Negative Ion Mode Tandem Mass Spectrometry: Superior Performance of Negative Ion Electron Capture Dissociation. *Anal. Chem.* **2012,** *84*, 6370-6377.
6. Kjeldsen, F.; Silivra, O. A.; Ivonin, I. A.; Zubarev, R. A. In *Electron Detachment Dissociation in Quadrupole Ion Trap Highlights Dominant C(alpha)-C Backbone Fragmentation in Polypeptide Radical Anions*, The 52nd ASMS Conference on Mass Spectometry and Allied Topics, Nashville, TN, May 23-27; Nashville, TN, 2004; pp CD-ROM.
7. Bowie, J. H.; Brinkworth, C. S.; Dua, S., Collision-Induced Fragmentations of the (M-H)(-) Parent Anions of Underivatized Peptides: An Aid to Structure Determination and some Unusual Negative Ion Cleavages. *Mass Spectrom. Rev.* **2002,** *21*, 87-107.
8. Anusiewicz, I.; Jasionowski, M.; Skurski, P.; Simons, J., Backbone and Side-Chain Cleavages in Electron Detachment Dissociation (EDD). *J. Phys. Chem. A* **2005,** *109*, 11332-11337.
9. Kjeldsen, F.; Silivra, O. A.; Ivonin, I. A.; Haselmann, K. F.; Gorshkov, M.; Zubarev, R. A., C(alpha)-C Backbone Fragmentation Dominates in Electron Detachment Dissociation of Gas-Phase Polypeptide Polyanions. *Chem. Eur. J.* **2005,** *11*, 1803-1812.
10. Flick, T. G.; Merenbloom, S. I.; Williams, E. R., Anion Effects on Sodium Ion and Acid Molecule Adduction to Protein Ions in Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2011,** *22*, 1968-1977.
11. Liu, X. H.; Cole, R. B., A New Model for Multiply Charged Adduct Formation Between Peptides and Anions in Electrospray Mass Spectrometry *J. Am. Soc. Mass Spectrom.* **2011,** *22*, 2125-2136.
12. Horn, D. M.; Ge, Y.; McLafferty, F. W., Activated Ion Electron Capture Dissociation for Mass Spectral Sequencing of Larger (42 kDa) Proteins. *Anal. Chem.* **2000,** *72*, 4778-4784.

13. Mikhailov, V. A.; Cooper, H. J., Activated Ion Electron Capture Dissociation (AI ECD) of Proteins: Synchronization of Infrared and Electron Irradiation with Ion Magnetron Motion *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 763-771.

Chapter 4

Application of Targeted LC/MS/MS in Pharmaceutical Protein Sequence Variant Analysis

4.1 Introduction

Therapeutic proteins are becoming an important contributor to efforts against various diseases, such as cancer,[1] hepatitis,[2] and acquired immunodeficiency syndrome (AIDS).[3] A common strategy for producing therapeutic proteins is through protein expression in host cells which have been transfected with the correspondent gene sequences. During manufacture, any deviation in the primary and higher-order structure from the designed protein can generate product-related variants. Such variants, which could be a low proportion after purification, can still potentially cause safety concerns through immunogenicity, or impact efficacy through many aspects such as binding affinity, stability, and pharmacokinetics. Therefore, controlling the levels of certain product-related variants is crucial and regulated by health authorities across the world. Protein sequence variance (PSV) is a type of product-related variant that deviates from

the designed sequence and may cause concerns in safety and efficacy. PSV could originate from multiple steps in protein expression, starting from deoxyribonucleic acid (DNA) replication,[4] messenger ribonucleic acid (mRNA) transcription,[5] to peptide synthesis in the ribosome.[6] In general, two categories of PSV have been reported: DNA mutation and translational misincorporation. During transfection and DNA amplification, deviation or mutation in the sequence may cause inheritable variation to the cell line. During translation, undesired tRNAs may mismatch with the mRNA, or a proportion of tRNA can be mischarged by incorrect amino acids. In both cases, an undesired amino acid residue will be brought to the peptide chain.

Due to the mechanistic difference between the two categories of PSV, the observed occurrence rate and distribution are different between PSVs from DNA mutation and translational misincorporation. According to previous reports, the level of translational misincorporation (mismatch and mischarge) is significantly higher than natural DNA replication error, up to 10^4 fold.[7-9] Distribution-wise, for a given protein expressed by a certain organism, individual DNA replication error rate is relatively lower than misincorporation. Thus it is even more unlikely that multiple clones, if they share no common ancestry after transfection, have the same mutation at the same site.[7] For large-scale protein manufacturing, once a DNA mutation is confirmed in the master cell bank (MCB), it could be difficult to remove the resulting protein variant by adjusting cell culture conditions and purification processes. On the other hand, levels of translational misincorporation can be affected by codon usage[10] and culture condition[11] of protein producing cells. For a panel of single clones expressing the same protein under identical conditions, the same type of translational misincorporation can occur in all clones. Translational misincorporation can be reduced or adjusted by optimizing cell culture conditions. For improved understanding of the monoclonal MCB and for developing strategies to reduce PSVs, it is

important not only to detect PSVs in the early stage of biologics development, but also to determine whether they are caused by DNA mutations or translational misincorporation.

Detection of low-level protein sequence variants requires sensitive analytical techniques. Currently, LC/MS/MS based peptide mapping is employed to provide information on both location and quantification of sequence variants. Changes in protein or peptide sequence such as amino acid substitution, except between leucine and isoleucine, will result in a change of mass. Such changes can be detected by LC/MS/MS. Top-ten DDA (see Chapter 1.6) is the most common data acquisition strategy, in which the top ten most abundant peaks in each MS scan are isolated and subjected to MS/MS experiments. Peptides can thus be identified by their corresponding MS/MS spectra. Increased resolution and mass accuracy facilitate identification of lower level sequence variants by providing more confident LC/MS identification of peptides.[12] However, the limitation of this method in sequence variant detection is that sequence variant peptides are usually low in abundance, and may co-elute with other peptides, thus there is no guarantee that their peaks are among the top ten most abundant peaks in the spectrum. Also, DDA inserts MS/MS scans between MS scans, which reduces the number of points in LC/MS extracted ion chromatogram (EIC) peaks. For many sequence variants in low abundance, EIC peak area is the only way to quantify their abundance and DDA-based analysis may thus suffer reliability and reproducibility issues. A previous report using multiple LC/MS/MS runs with dynamic exclusion of peptides found in the previous run improved PSV detection sensitivity,[13] however it brought large numbers of false positives, most of which need to be manually examined and removed.

Here we introduce a workflow to sensitively detect low-abundance PSVs and assess the probability that the observed PSVs originates from DNA mutation or translational misincorporation. The first step is DDA on protein material expressed from a group of single clones cultured under the same conditions, and to search for the major sequence variants. The second step is codon selected targeted LC/MS/MS, in which we search for additional sequence variants with the same codon change as major sequence variants. Then, from the distribution of sequence variants from the same codon change, we can determine how well a PSV fits into this distribution. Based on the assumption that translational misincorporation rates of all amino acids encoded by the same codon should follow a uniform pattern across different clones, PSVs outside the distribution is likely to be caused by DNA mutations. By this approach, we not only discovered more sequence variants, but also obtained more information on the properties of the codon changes that generate sequence variants, which are not commonly addressed in mass spectrometry-based analytical methods. Information about codon change helps to determine the underlying cause of sequence variants, and provides insight into cell line selection and cell culture condition optimization.

4.2 Experimental

4.2.1 Materials

The monoclonal antibodies (Antibody A, B, C, 1 and 2) were expressed in recombinant CHO cells and produced using standard mammalian cell cultivation, followed by chromatographic purification. Hydrochloric acid (HCl), dithiothreitol (DTT) and iodoacetamide (IAM) used in peptide mapping were purchased from Sigma-Aldrich (St. Louis, MO). HPLC grade reagents were purchased from Avantor (Center Valley, PA), sequencing grade modified

trypsin was purchased from Promega (Madison,WI). Synthetic peptides were ordered from Genscript, LLC.

4.2.2 Sample Preparation

Antibody samples were denatured with guanidine HCl followed by reduction with DTT and alkylation with IAM. The reduced and alkylated antibody was loaded onto a Zeba gel desalting 96-well spin-plate (Thermo Fisher Scientific, Rockford, IL) and eluted with a buffer containing 50 mM Tris/HCl, 2.0 M urea, 1.0 mM $CaCl_2$, pH 7.6. Sequencing grade trypsin (1%, w/w, enzyme to protein) was added to the sample prior to incubation at 37 °C for 2 h. The digestion was stopped by adding 1.0 M HCl to pH 2–3. The tryptic digests were separated on an Acquity UPLC BEH300 C18 column (1.7 μm, 2.1×150 mm, Waters, Milford, MA) at a flow rate of 0.2 mL/min using a Waters Acquity UPLC instrument (Waters, Milford, MA). The column was equilibrated with 1% acetonitrile containing 0.1% formic acid in water, and peptides were eluted with an increasing gradient of acetonitrile from 0 to 80% over 120 min.

4.2.3 Peptide Mapping

Top ten DDA LC/MS/MS analyses were conducted using a Thermo Orbitrap Elite (see Figure 1.3). MS spectra were acquired using the orbitrap FTMS analyzer, while CAD MS/MS fragments were generated and detected in the linear ion trap.[14] LC/MS/MS data searching was performed by Trans Proteomic Pipeline [15] or Protein Metrics Byonic.[16] Manual MS/MS viewing and matching for uncertain hits were done by Molecular Weight Calculator. Sequence variants identified in this manner were designated as major sequence variants.

In order to lower detection limit and achieve smoother peaks for quantification, separate LC/MS experiments were conducted on the same samples, using the same sample preparation procedure and LC gradient as DDA experiments. Thermo SIEVE software was utilized for feature detection in these experiments. An in-house R program was developed to search for sequence variants with the same codon as major sequence variants for detected features. For all detected MS1 matches, this program verified the isotopic distribution of the first two isotopic peaks with converted mzXML files from raw data files, in order to reduce false positives. With the search results, this program can also generate an inclusion list for targeted LC/MS/MS, providing verification for all sequence variants. Targeted LC/MS/MS experiments were conducted under the exact same conditions as DDA experiments, except that only precursor ions on the inclusion list were chosen for MS/MS. Quantification was done through Thermo Quan Browser, based on LC/MS EIC peak area percentage of sequence variant peptide over sum of the sequence variant peptide and its corresponding wild type peptide.[17] The workflow is shown in Figure 4.1.

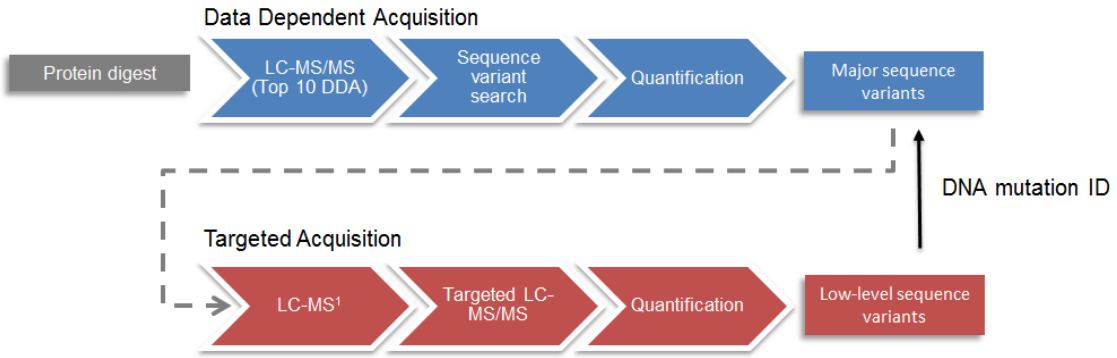


Figure 4.1. Workflow for low-level sequence variant detection. The blue part is the conventional bottom-up LC/MS/MS peptide mapping workflow for sequence variant searching. The red part is the codon-based targeted LC/MS/MS workflow for detecting low level sequence variants. Results from the low level sequence variant detection can be used to determine the origin of major sequence variants (DNA mutation or not).

4.2.4 Quantification Verification

Sequence variant and wild type peptides from Antibody B with codon change AGC(Ser) to AAC(Asn), GTC(Val) to CTC(Leu) and GGC(Gly) to GAC(Asp) were synthesized and mixed in ~1 μM each. TGC(Cys) to TAC(Tyr) variants were also found ,but synthetic peptide experiments were not conducted due to complicating factors of incomplete cysteine reduction/alkylation in the workflow that cannot be removed. To ensure that each UV peak on the HPLC UV spectrum contains only one peptide, 25 synthetic peptides were divided into 5 groups and run separately. 215 nm UV peak area and MS extracted ion chromatogram (EIC) peak area of each peptide at the charge state used for sequence variant quantification were acquired for each peptide. MS-UV abundance coefficient C for each sequence variant was calculated by:

$$C = \frac{UV_{SV}/MS_{SV}}{UV_{WT}/MS_{WT}}$$

in which UV_{SV}, MS_{SV}, UV_{WT}, MS_{WT} are the UV peak areas of the sequence variant peptide, MS EIC peak area of the sequence variant peptide, UV peak area of the wild type peptide, and the MS EIC peak area of the wild type peptide, respectively.

4.3 Results and Discussion

4.3.1 Sequence Variant Detection of Monoclonal Antibodies

The LC/MS/MS peptide mapping workflow described above was applied to antibody proteins A, B, and C. Twelve pre-RCB (Research cell bank) clones of antibody A, C and eigh pre-RCB clones of antibody B were analyzed. Table 4.1 shows the total numbers of sequence

variants detected for each type of codon, in each protein. LC/MS with prolonged ion accumulation time and no MS/MS provided improved sensitivity and quantification quality. Additional sequence variants, which are not detected in DDA experiments due to low peak abundance, can be detected with this workflow For example, for antibody B, DDA identified a heavy chain V305L variant, which has ~0.03% abundance; through LC/MS searching of sequence variants with this codon change, followed by targeted MS/MS, a heavy chain V273L variant with less than 0.01% abundance could be identified on the same instrument. Mapping of sequence variant abundances with the same codon change throughout all clones and amino acid residues contributes to the confidence of DNA mutant identification. For previously reported common translational misincorporation, such as Cys to Tyr and Ser to Asn, when one was found in DDA experiments, we saw more sequence variants at different amino acid residues through the targeted workflow.

Antibody A	AGC(S) to AAC(N)	TGC(C) to TAC(Y)	TTC(F) to GTC(V)
SV detected by Top 10 DDA	2	2	1
SV detected by Top 10 DDA + Targeted MS/MS	8	7	1
SV not detected	12	4	11

Antibody B	AGC(S) to AAC(N)	TGC(C) to TAC(Y)	GTC(V) to CTC(L)/ATC(I)	GGC(G) to GAC(D)	AAA(K) to AAC/AAT(N)
SV detected by Top 10 DDA	1	2	1	1	1
SV detected by Top 10 DDA + Targeted MS/MS	7	4	5	3	1
SV not detected	14	5	18	9	9

Antibody C	GTC/GTG/GTT(V) to CTC/CTG/CTT(L)	AAC/AAT(N) to AAG(K)	AAC/AAT(N) to AGC/AGT(S)	AGC(S) to AAC(N)	GCC(A) to TCC(S)	AAG(K) to AGG(R)
SV detected by Top 10 DDA	1	1	2	1	1	1
SV detected by Top 10 DDA + Targeted MS/MS	15	5	13	6	1	2
SV not detected	26	14	6	17	16	13

Table 4.1. Sequence variants identified in two stages of our workflow. Codons from potential DNA error are marked in red.

4.3.2 Sequence Variant Abundance Distribution

For each type of codon change, through observation of sequence variant abundance across clones at different amino acid residues on the whole protein sequence, we found two distinct distribution patterns, as demonstrated in Figure 4.2. If a sequence variant was observed in only one amino acid residue of the protein, it is probable that this sequence variant comes from a DNA mutation, since otherwise translational misincorporation should have invoked sequence variants sharing the same codon change in other amino acid residues of the protein, and in other clones. For individually developed clones, one type of DNA mutation usually occurs on only one of the clones, if these clones do not share an erroneous vector. In Figure 4.2, the codon containing such DNA mutations can be easily identified. We detected such mutation in antibody

B. Such DNA mutation may originate from vector or culture cell DNA mutation during replication. From our workflow, we were able to detect DNA mutation less than 0.5%.

The second pattern originates from translational misincorporation, in which sequence variants were detected on multiple amino acid residues and in all clones. In this pattern, for all possible amino acid residues corresponding to a particular codon change, the distribution of sequence variant abundance across clones is similar at each amino acid residue where translational misincorporation is detected. Column graphs showing such distributions for more translational misincorporation codon changes can be found in Figure 4.3.

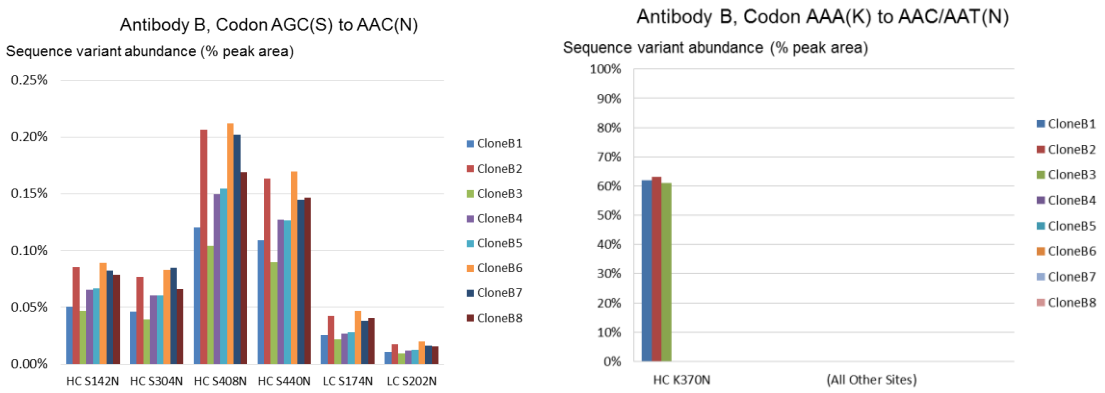


Figure 4.2. Typical sequence variant abundance distribution of codons containing misincorporation (left) and codons containing DNA mutation (right). Abundance values are in % MS EIC peak area. Note the similarity in distribution of sequence variant percentages across clones in the left graph.

Figure 4.3. Sequence variant abundance distribution pattern of antibody A, B and C, in % EIC peak area. Only translational misincorporation codons with three or more sequence variants found are shown.

Figure 4.3 Continued.

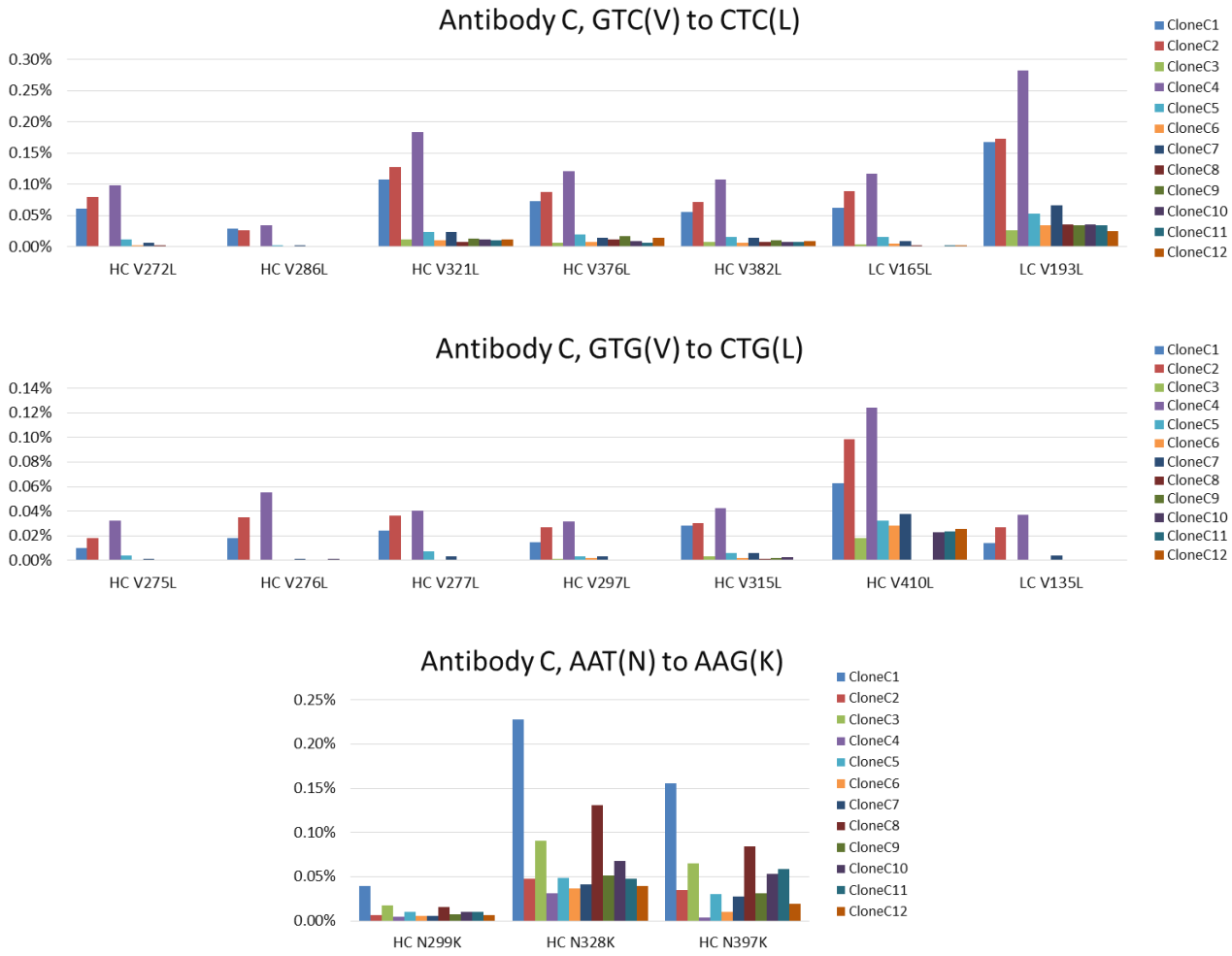


Figure 4.3 Continued.

The consensus between distribution patterns can be explained by the uniformity in codon usage and amino acid utilization in each individual cell line. Although expressing the same protein under identical culture conditions, there are still differences between cell lines in their capability to utilize cell culture material. Such differences cause the same translational misincorporation to have different abundance in each clone. However, for DNA errors, it is very unlikely to observe the same DNA error occurring at multiple clones and multiple amino acid residues. Recognizing such patterns, sequence variant abundance distribution of each codon can be used to identify DNA mutation.

The difference of translational misincorporation sequence variant abundance between different amino acid positions of the same codon and the same clone is also an interesting observation. To rule out the possibility that such difference is caused by the slight difference in ionization efficiency between sequence variant peptides and wild type peptides due to differences in sequence and elution time/spray solvent composition, quantitation was verified with UV detection as standard, as described in chapter 4.2.4. Quantification coefficient (designated as *C*) stands for the ratio of sequence variant abundance measured in UV peak area over the abundance measured in MS EIC peak area. Table 4.2 shows *C* of sequence variant peptides with codon change AGC(Ser) to AAC(Asn), GTC(Val) to CTC(Leu) and GGC(Gly) to GAC(Asp). From this Table, we can see that, for most sequence variants, their abundance values from MS EIC peak area are fairly close to values from UV peak area: 10 out of 14 sequence variants have *C* within 0.9-1.1. If we use these coefficients to convert MS EIC peak area abundance values into UV peak area abundance values, we can see that, for codon change AGC(Ser) to AAC(Asn) and GTC(Val) to CTC(Leu), the difference in sequence variant abundance between amino acid positions is still significant. For GGC(Gly) to GAC(Asp), the difference is insignificant, as shown in Figure 4.4. Such difference in tendency of showing translational misincorporation at amino acid positions corresponding to the same codon is also consistent for each clone. In the case when sequence variant abundance varies within codon change, it is unlikely for this codon change to correspond to a tRNA mischarging, because, in that case, all amino acid residues should be affected to the same degree. A possible explanation for this phenomenon is that tRNA binding is correlated with the local structure of the existing protein sequence as well as the micro environment of the ribosome.

Sequence Variant	Variant peptide sequence	Wild type peptide sequence	C
HC S142N	STSENTAALGCLVK	STSESTAALGCLVK	1.880
HC S304N	VVNVLTVLHQDWLNGK	VVSVLTVLHQDWLNGK	1.041
HC S408N	TTPPVLDSDGSFFLYNK	TTPPVLDSDGSFFLYSK	1.027
HC S440N	NLSLSPG	SLSLSPG	1.031
LC S174N	DSTYNLSSTLTLSK	DSTYSLSSTLTLSK	1.336
LC S202N	VYACEVTHQGLNSPVTK	VYACEVTHQGLSSPVTK	0.955
HC V37L	LSC**********WIR	LSC**********WVR	1.037
HC V129L	GPSIFPLAPCSR	GPSVFPLAPCSR	0.934
HC V273L	TPEVTCVVVDVSHEDPEIK	TPEVTCVVVDVSHEDPEVK	0.901
HC V305L	VVSILTVLHQDWLNGK	VVSVLTVLHQDWLNGK	0.998
HC V308L	VVSVLTILHQDWLNGK	VVSVLTVLHQDWLNGK	2.137
HC G147D	STSESTAALDCLVK	STSESTAALGCLVK	1.696
HC G281D	FNWYVDDVEVHNAK	FNWYVDGVEVHNAK	0.929
LC G200D	VYACEVTHQDLSSPVTK	VYACEVTHQGLSSPVTK	1.009

Table 4.2. MS-UV abundance coefficient for sequence variants of Antibody B. The sequence of the heavy chain V37L is not disclosed because that the peptide contains complementarity determining region (CDR) that is not yet filed.

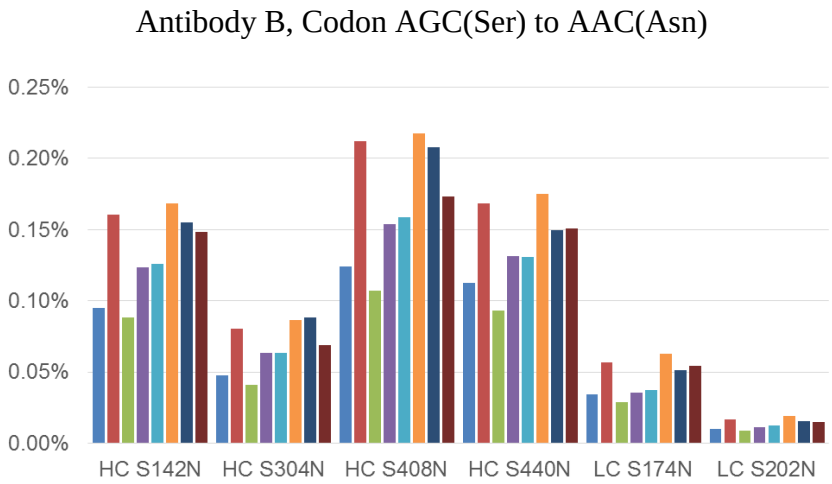
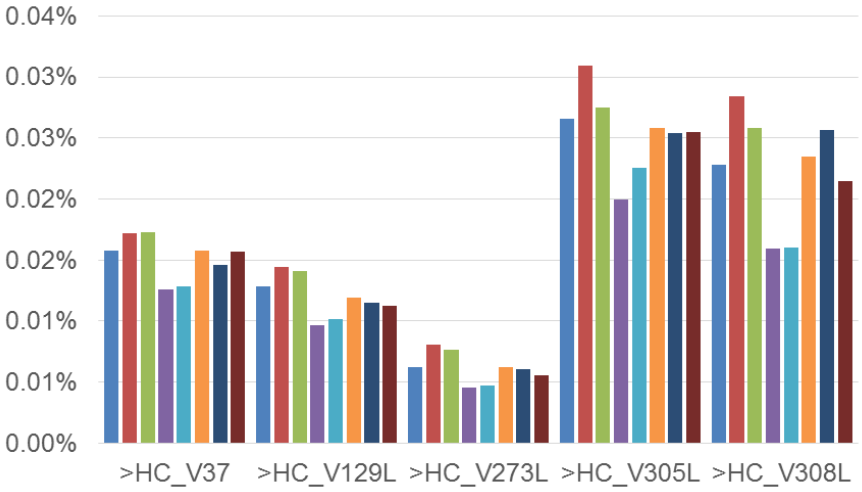


Figure 4.4. Sequence variant abundance distributions by codons from Antibody B, in calculated UV peak area percentage. Each color stands for a clone of Antibody B.

Antibody B, GTC(Val) to CTC(Leu)



Antibody B, GGC(Gly) to GAC(Asp)

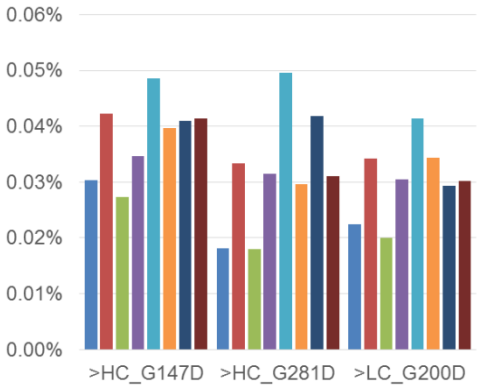


Figure 4.4 Continued.

4.3.3 DNA Error Determination Strategy

As discussed in chapter 4.3.2, when a DNA error occurs at codons for which no translational misincorporation happens, or when a DNA error has significantly higher abundance than the abundance values of translational misincorporation at that codon, the DNA error can be easily identified due to its distinct sequence variant abundance distribution pattern. However, in some cases, DNA mutations and translational misincorporation may occur with the same codon change at a comparable abundance. In such cases, with sequence variant abundance values

across all clones and amino acid residues, we can quantitatively describe the likelihood of a sequence variant in a clone to be a DNA mutant. Assuming that sequence variant level for multiple expression of the same clone under identical conditions follows a Gaussian distribution, z score ($z=(x-\mu)/\sigma$), its corresponding p-value can be used to describe the probability of having a DNA error rather than a translational misincorporation.

For a matrix of sequence variant abundances, we can use average distribution to find an estimated expected value μ. From our observation of sequence variant abundance distribution patterns, we can summarize that sequence variants sharing the same codon change have the same abundance distribution across different clones at each amino acid residue. Thus for any sequence variant belonging to a given clone and amino acid residue, it is possible to use the relative intensity data of other sequence variants with the same codon to estimate the relative intensity of this given sequence variant when it is a translational misincorporation. We estimated residue-to-residue ratio by computing the median of corresponding clones in each residue, then using residue-to-residue ratios and abundance values of the same clone to derive μ for each point on the matrix. This μ is the sequence variant abundance of a specific amino acid residue/clone combination if this sequence variant is a translational misincorporation, which follows exactly the same distribution pattern across clones at different amino acid residues with a given codon. In other words, distributions of μ across different clones are the regression of the sequence variant abundance distribution at different amino acid residues. Example of μ values and sequence variant abundance distribution patterns of a translational misincorporation are shown in Figure 4.5. Distributions of μ at different amino acid residues of the same codon change are linear to each other. In the case of this given sequence variant being a DNA mutant, we should be observing a positive deviation from this estimated relative intensity.

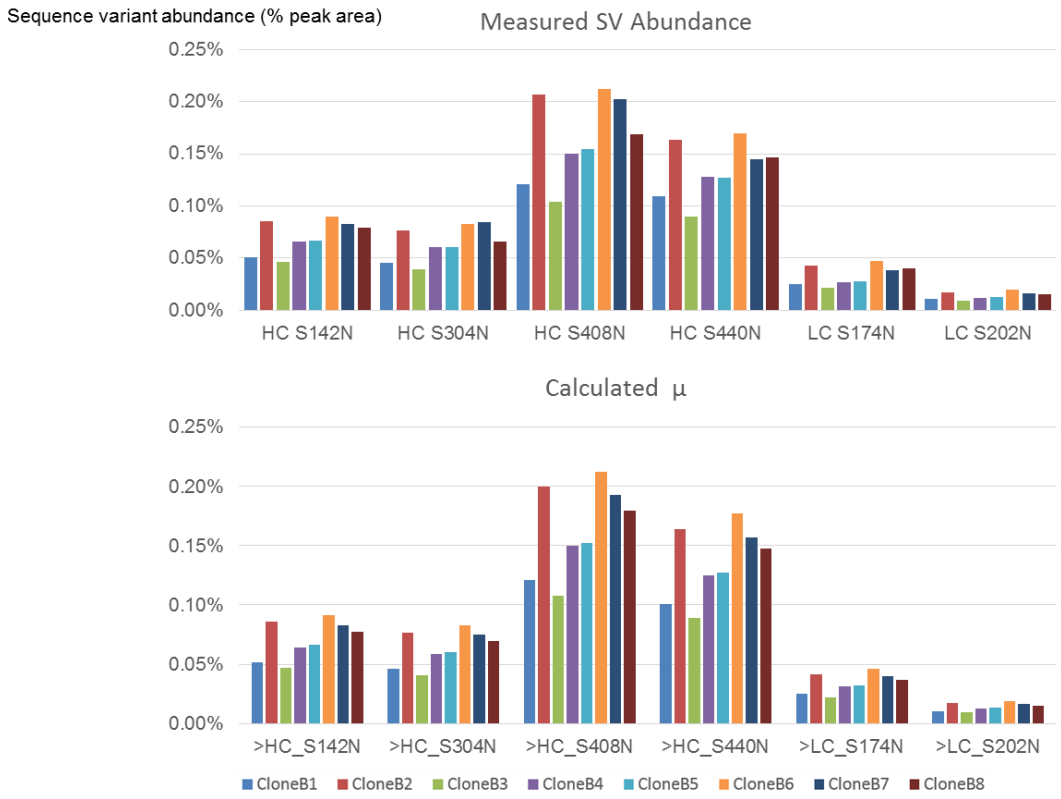


Figure 4.5. Measured sequence variant abundance distribution pattern and μ values of translational misincorporation with codon change AGC(S) to AAC(N) in Antibody B.

To correlate observed positive deviation and probability of having a DNA mutation, we need to estimate the level of deviation of sequence variant abundance (σ) when a monoclonal antibody is expressed. In order to measure this deviation, ideally we need to measure sequence variant abundance of multiple cultures under the same condition of the same cell line, using an identical analytical protocol. This approach will significantly increase the workload of protein expression and purification, but return limited extra information. Also, when we report a DNA mutation through mass spectrometry methods, reducing false positive detections has higher priority than including more potential DNA mutations. Particularly at a very low level where deep DNA sequencing cannot provide external confirmation, such false positives cannot be easily removed.

Considering these facts, two approaches can be applied to estimate the deviation mentioned above, and the greater one should be used as the final deviation value to reduce false positives. For the first approach, we used relative standard deviation of sequence variant abundance at each amino acid residue to represent the maximum value of deviation we could have when expressing the same cell line multiple times. Standard error for sequence variant abundance of each residue/clone combination is thus the product of this RSD and measured sequence variant abundance at each residue/clone combination. This error is therefore considered to be maximum expression error. In addition to random error in expressing one cell line, this error includes differences caused by individuality of cell lines, which means they should be even higher than the actual error.

Second, we used an artificial sequence variant to estimate analytical error during the peptide mapping workflow. Two antibodies (Antibody 1 and Antibody 2, see Table 4.3), with six of one protein's tryptic peptides differing by one amino acid residue to a tryptic peptide in the other protein, were employed. When these two proteins are mixed in different ratios, sequence variants of different ratios can be mimicked. Mixture ratios tested were 0.02%, 0.05%, 0.1%, 0.2%, and 0.5%; each one was analyzed in six replicas, and measured independently. Detected sequence variants relative standard deviation (RSD) versus measured "sequence variant abundance" was plotted, as in Figure 4.6. Median RSD of "sequence variants" whose detected abundance were in four ranges: 0-0.5%, 0.5%-1%, 1%-2%, and 2% and above, were used on other proteins to estimate analytical error when detected sequence variant abundance fell into the corresponding range. This error is considered to be maximum analytical error. In addition to random error in sample processing and analytical methods, this maximum error contains random error originating from manual mixing.

Antibody 1	Antibody 2
VEPK	VESK
ATGIPDR	ATGIPAR
SLSLSPG	SLSLSLG
EPQVYTLPPSR	EPQVYTLPPSQEEMTK
EIVLTQSPGTLSLSPGER	EIVLTQSPATLSLSPGER
TTPPVLDSDGSFFLYSK	TTPPVLDSDGSFFLYSR

Table 4.3. Six pairs of tryptic peptides from Antibody 1 and Antibody 2 that were selected for analytical error study.

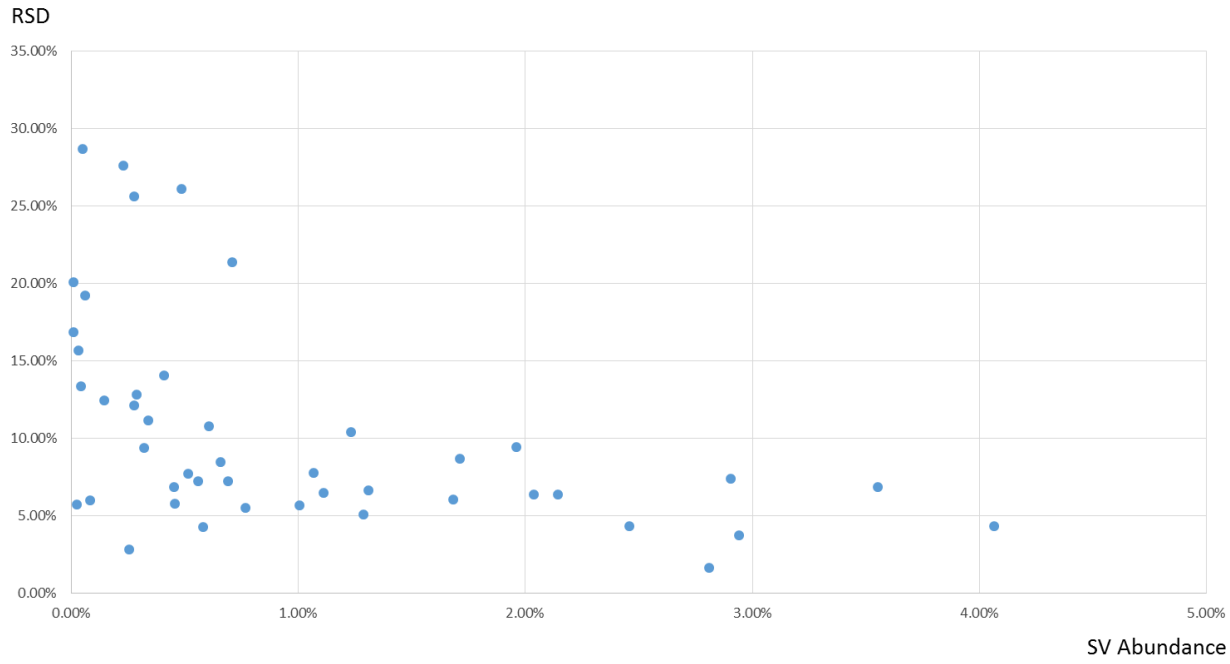


Figure 4.6. Relative standard deviation of detected sequence variant relative abundance plotted against detected sequence variant abundance.

If observed sequence variant abundance has a positive deviation from μ that is greater than the confidence interval derived from the greater one of maximum analytical error and maximum expression error, it is very likely to be a DNA mutation sharing a codon change with misincorporation. From our investigation, as demonstrated in Figure 4.7, some sequence variant

levels have positive deviation greater than maximum analytical error, but none of these deviations is greater than the maximum expression error, thus no contribution of DNA mutation in these sequence variants can be concluded. For this one-tail hypothesis, when significance level is set to 0.01, sequence variants with z scores greater than 2.33 can be considered as DNA mutation. From Figure 4.8 we can see that no sequence variant in antibody A, B, or C reached this level, except for the confirmed DNA error in antibody B.

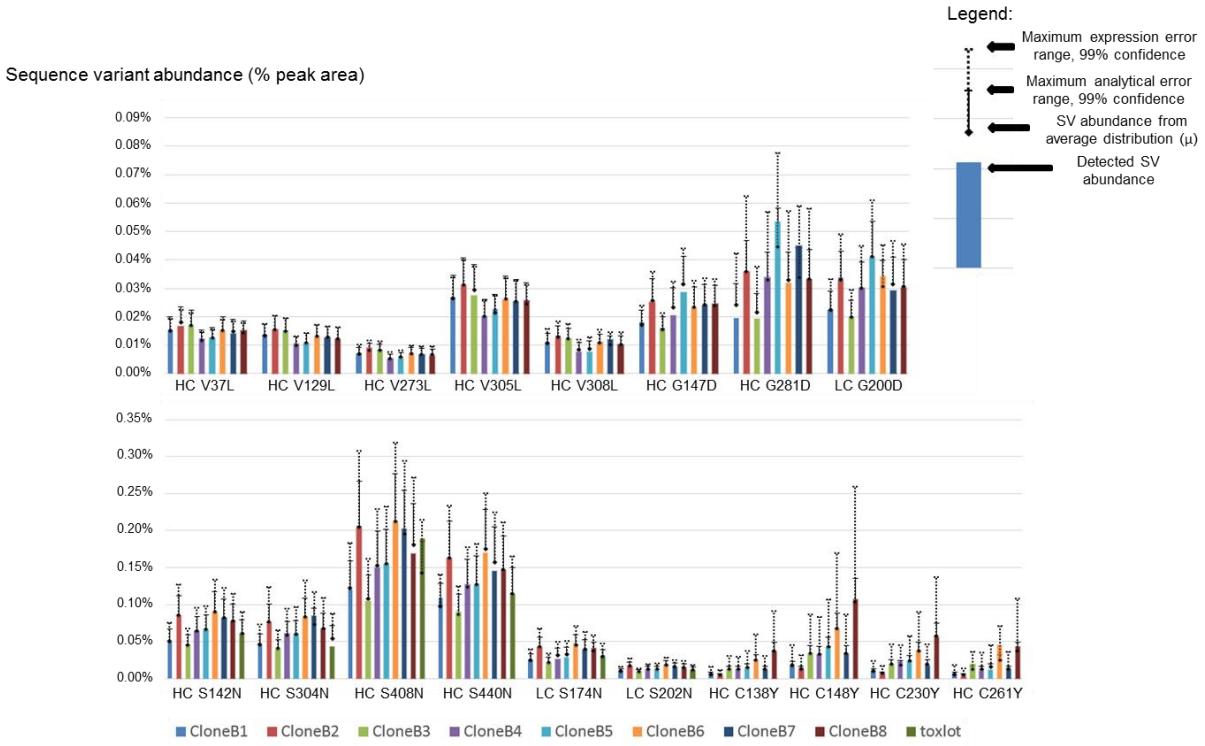


Figure 4.7. Detected sequence variant abundance distribution of four types of misincorporation detected in Antibody B. >99% confidence DNA mutation if detected sequence variant abundance is greater than all error ranges.

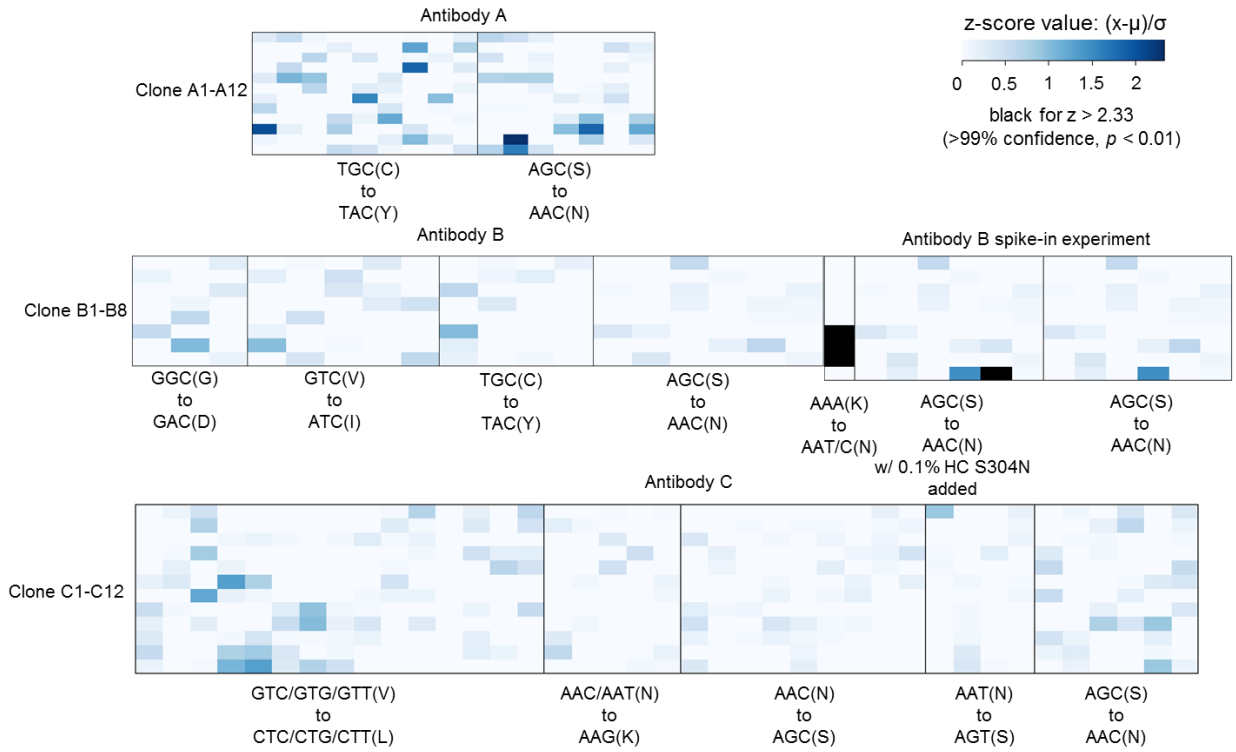


Figure 4.8. Heat maps of z-scores derived by calculated μ values and corresponding maximum error (the greater one of analytical or expression error). Horizontal axes are amino acid position of sequence variants, grouped in codons. Minus values are labeled as zeroes. 0.1% spiked-in HC S304N has a z-score of 4.43. HC K370N is a 62% DNA error caused by vector error.

A synthetic spike-in experiment was conducted to examine the evaluation system above, by simulating a DNA mutation sharing a codon change with a common translational misincorporation. A synthetic peptide corresponding to a S304N sequence variant from the heavy chain was spiked into the toxicity lot material of protein B, with 0.025%, 0.05%, 0.075%, and 0.1% molar concentration of protein samples. As shown in Figure 4.9, in the calculation of μ, the median was used for estimating expectation, thus a changing sequence variant abundance in a single residue/clone combination have minor influence on all μ values, including μ at the changed residue/clone combination. As for error bars, maximum analytical error is a function of μ, thus it changed little. Maximum expression error is correlated with the RSD of detected

sequence variant abundances, therefore as measured sequence variant abundance increases deviate from the mean across clones, maximum expression error also increased. Starting sequence variant abundance was ~0.04%, and when another ~0.05% of synthetic peptide was spiked in, the measured sequence variant abundance was outside the 95% confidence interval, for which the sequence variant can be a translational misincorporation. When spiked-in molar ratio of synthetic peptide increased to 0.075%, the measured sequence variant abundance fell out of the 99% confidence interval, in which case the method told us that at least 99% of such observations should be considered to be DNA errors. Z score of HC S304N with 0.1% spiked-in synthetic peptide in toxicity lot is plotted in Figure 4.8 for Antibody B.

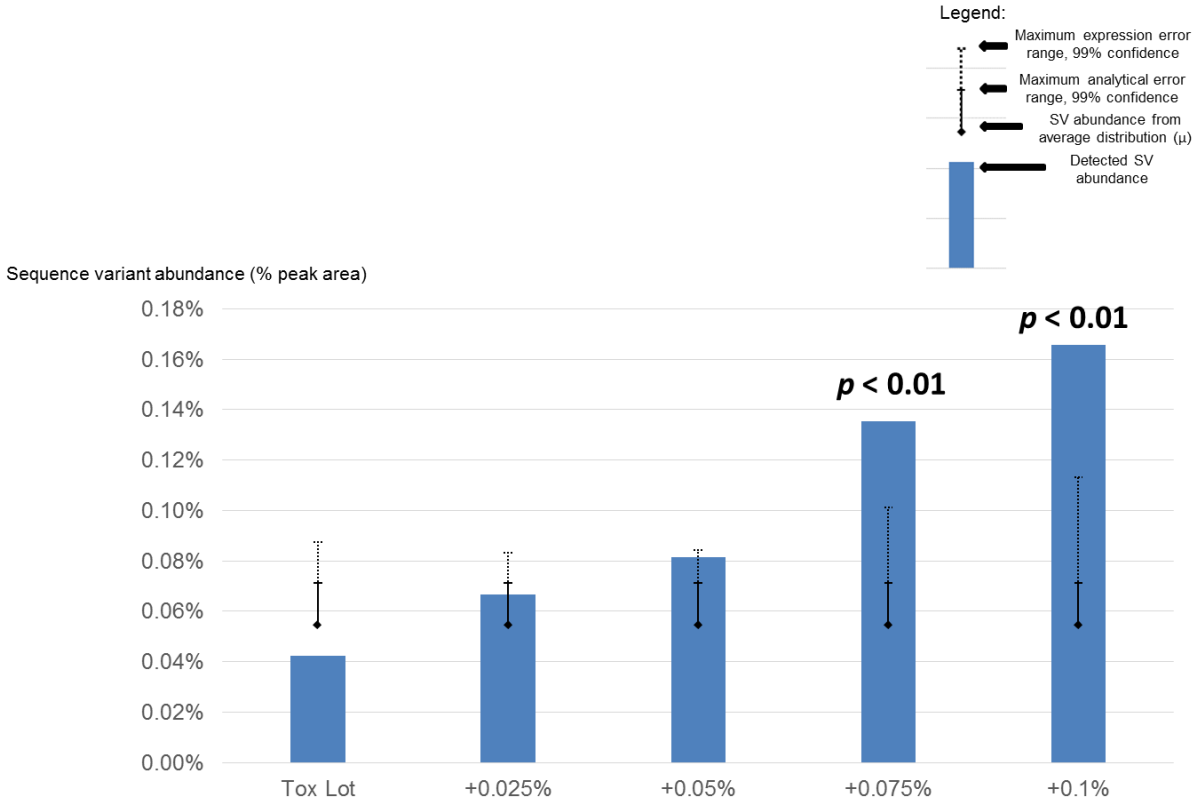


Figure 4.9. HC S304N spike-in experimental results. Note that, as measured sequence variant ratio goes up, thereby simulating a DNA mutation, calculated SV percentage of an ideal misincorporation remains the same.

4.3.4 Implications of DNA Error

In cell line selection for protein expression in mammalian cells, translational misincorporation is possible to reduce or eliminate by adjusting cell culture conditions, while DNA mutations cannot be easily removed. Sequence variants further addressed by our targeted workflow can be confirmed to be DNA mutations or not, therefore no effort will be wasted on optimizing cell culture conditions to remove DNA mutations in the cell. In addition, although cutting-edge DNA sequencing methods, such as duplex sequencing, provide as low as 10^{-7} level mutation detection capability[18], commercially available deep sequencing method, such as Illumina MiSeq, identifies mutation only as low as 1% of all DNA molecules, which prevents us from high-throughput verification of DNA mutation below 1%. The targeted workflow introduced above, however, achieves DNA mutation identification down to 0.1%, and can be implemented on commercially available orbitrap mass spectrometry instruments. It can serve as a complementary method for DNA sequencing from downstream of the central dogma.

4.3.5 False Positive Screening

Another application of the sequence variant abundance distribution pattern is to facilitate false positive removal during sequence variant identification. In some cases, we may observe that a sequence variant abundance distribution pattern at some amino acid residues does not consent with shared patterns of other amino acid residues with the same codon. The possibility of having a false positive should be considered when such inconsistency is observed on multiple clones and one amino acid position. During the sequence variant search for antibody B, an MS peak that appeared to be HC G100D demonstrated drastically different abundance distribution pattern across clones compared to other GGC(G) to GAC(D) changes. The MS/MS spectrum

and retention time shift of the presumed sequence variant appeared plausible (see Figure 4.10). However, when we injected samples for each clone in a reversed order, the abundance distribution of HC G100D became reversed as well, while other sequence variants remained the same (see Figure 4.11). Such phenomenon implies that this MS/MS search hit is more likely to be a false positive from auto-sampler degradation. Further synthetic peptide experiments showed that the HC G100D synthetic peptide has a different retention time than the MS/MS hit.

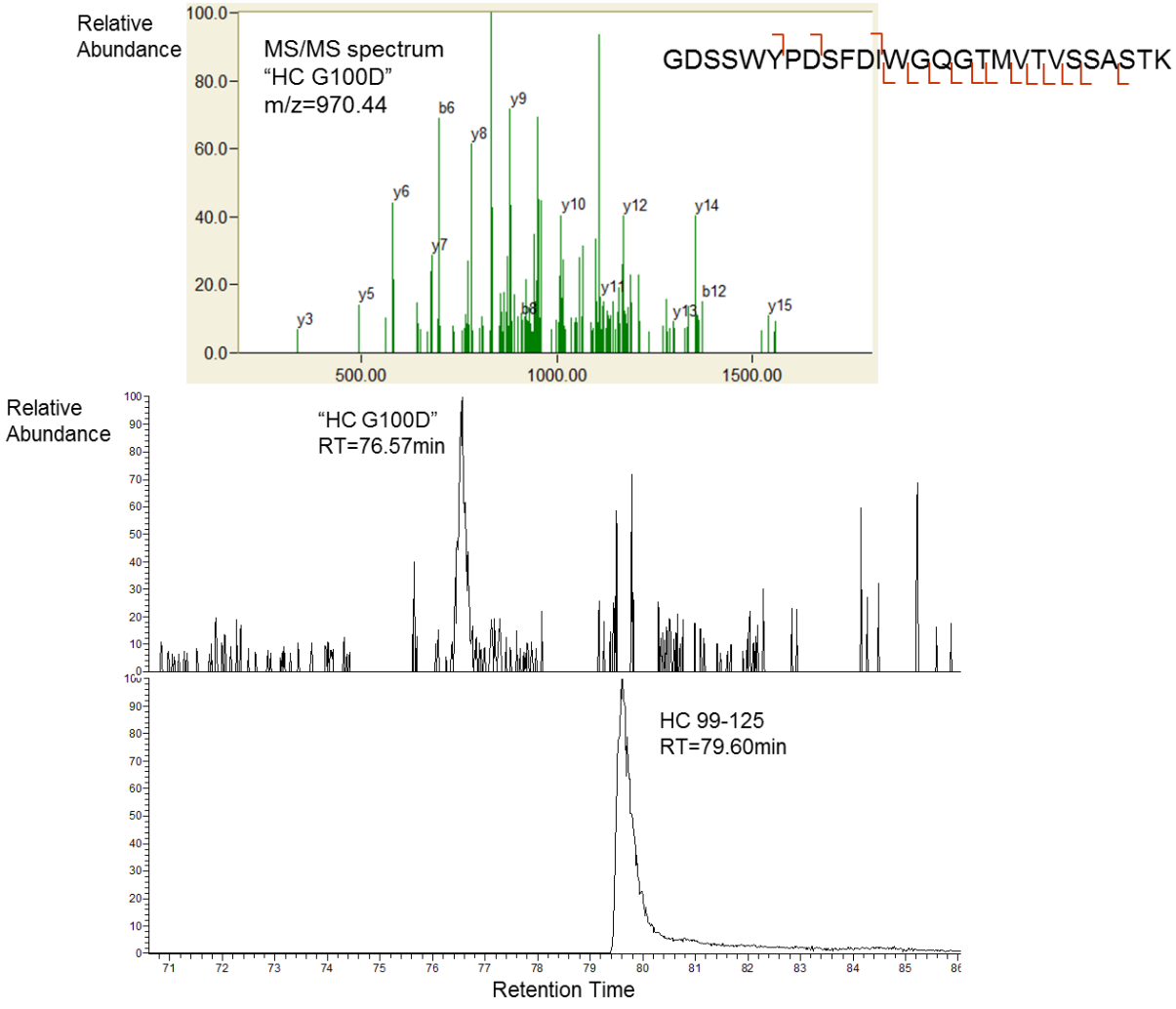


Figure 4.10. MS/MS and retention time shift of the "HC G100D" search hit from antibody B. The retention time shifted earlier than what can be explained by increased polarity.

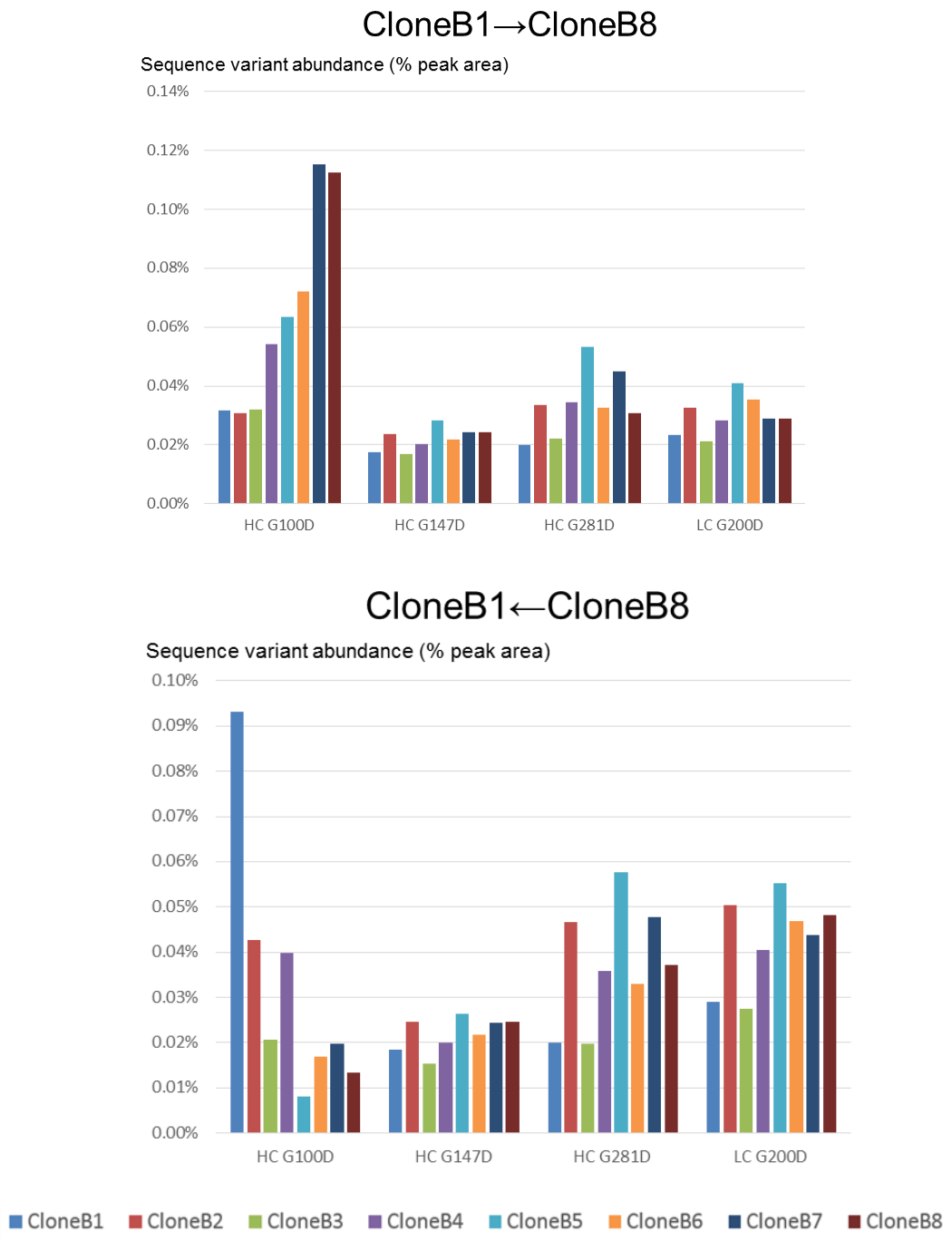


Figure 4.11. Sequence variant abundance distribution pattern of codon GGC(G) to GAC(D), antibody B. The distribution of HC G100D significantly differs from the other three sequence variants, and later was found to change when the injection order was reversed.

4.4 Conclusion

In this work, we presented a workflow to identify low level sequence variants sharing the same codon changes. For translational misincorporation of each type of codon change, similar distribution pattern across multiple clones was found at all amino acid residues. Such distribution pattern can be utilized to discriminate low level DNA mutations, no matter whether they have the same codon change with a common translational misincorporation or not. The pattern can also be used to screen out false positives, particularly those that cannot be overruled by tandem mass spectrometry or retention time shifts.

4.5 References

1. Scott, A. M.; Wolchok, J. D.; Old, L. J., Antibody therapy of cancer. *Nat. Rev. Cancer* **2012,** *12*, 278-287.
2. Marasco, W. A.; Sui, J., The growth and potential of human antiviral monoclonal antibody therapeutics. *Nat. Biotechnol.* **2007,** *25*, 14.
3. Tan, J. J.; Ma, X. T.; Liu, C.; Zhang, X. Y.; Wang, C. X., The Current Status and Challenges in the Development of Fusion Inhibitors as Therapeutics for HIV-1 Infection. *Curr. Pharm. Design* **2013,** *19*, 1810-1817.
4. Lebkowski, J. S.; Miller, J. H.; Calos, M. P., Determination of DNA-Sequence Changes Induced by Ethyl Methanesulfonate in Human Cells, Using a Shuttle Vector System. *Molecular and Cellular Biology* **1986,** *6*, 1838-1842.
5. Parker, J., Errors and alternatives in reading the universal genetic code. *Microbiological Reviews* **1989,** *53*, 273.
6. Dietrich, A.; Kern, D.; Bonnet, J.; Giege, R.; Ebel, J. P., Interpretation of Transfer-RNA-Mischarging Kinetics. *Eur. J. Biochem.* **1976,** *70*, 147-158.
7. Kunkel, T. A.; Bebenek, R., DNA replication fidelity. *Annu. Rev. Biochem.* **2000,** *69*, 497-529.
8. Kramer, E. B.; Farabaugh, P. J., The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. *Rna* **2007,** *13*, 87-96.
9. Guo, D. L.; Gao, A.; Michels, D. A.; Feeney, L.; Eng, M.; Chan, B.; Laird, M. W.; Zhang, B. Y.; Yu, X. C.; Joly, J.; Snedecor, B.; Shen, A., Mechanisms of Unintended Amino Acid Sequence Changes in Recombinant Monoclonal Antibodies Expressed in Chinese Hamster Ovary (CHO) Cells. *Biotechnology and Bioengineering* **2010,** *107*, 163-171.
10. Huang, Y.; O'Mara, B.; Conover, M.; Ludwig, R.; Fu, J.; Tao, L.; Li, Z. J.; Rieble, S.; Grace, M. J.; Russell, R. J., Glycine to glutamic acid misincorporation observed in a recombinant protein expressed by Escherichia coli cells. *Protein Science* **2012,** *21*, 625-632.
11. Ibba, M.; Soll, D., Quality control mechanisms during translation. *Science* **1999,** *286*, 1893-1897.

12. Tanaka, K.; Takenaka, S.; Tsuyama, S.; Wada, Y., Determination of Unique Amino Acid Substitutions in Protein Variants by Peptide Mass Mapping with FT-ICR MS. *J. Am. Soc. Mass Spectrom.* **2006,** *17*, 508-513.
13. Zhang, Z. Q.; Shah, B.; Bondarenko, P. V., G/U and Certain Wobble Position Mismatches as Possible Main Causes of Amino Acid Misincorporations. *Biochemistry* **2013,** *52*, 8165-8176.
14. Makarov, A.; Denisov, E.; Kholomeev, A.; Baischun, W.; Lange, O.; Strupat, K.; Horning, S., Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Anal. Chem.* **2006,** *78*, 2113-2120.
15. Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Farrah, T.; Lam, H.; Tasman, N.; Sun, Z.; Nilsson, E.; Pratt, B.; Prazen, B.; Eng, J. K.; Martin, D. B.; Nesvizhskii, A. I.; Aebersold, R., A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **2010,** *10*, 1150-1159.
16. Bern, M.; Cai, Y. H.; Goldberg, D., Lookup peaks: A hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal. Chem.* **2007,** *79*, 1393-1400.
17. Ruan, Q.; Ji, Q. C.; Arnold, M. E.; Humphreys, W. G.; Zhu, M. S., Strategy and Its Implications of Protein Bioanalysis Utilizing High-Resolution Mass Spectrometric Detection of Intact Protein. *Anal. Chem.* **2011,** *83*, 8937-8944.
18. Kennedy, S. R.; Schmitt, M. W.; Fox, E. J.; Kohrn, B. F.; Salk, J. J.; Ahn, E. H.; Prindle, M. J.; Kuong, K. J.; Shen, J.-C.; Risques, R.-A.; Loeb, L. A., Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat. Protocols* **2014,** *9*, 2586-2606.

Chapter 5

Conclusions and Future Directions

5.1 Summary of Results

Ion-electron reactions such as ECD and EDD provide substantial information for the structural analysis of biomolecules. The underlying mechanisms of these reactions constitute an intriguing research topic, and also provide insights for improving ECD/EDD performance as MS/MS methods in proteomics research. In Chapter 2, we conducted a systematic study of how metal-peptide interactions can affect ECD of metallopeptides. For the first time, the effect of the metal coordination spheres provided by peptides was systematically evaluated for metallopeptide ECD performance. Our data showed that certain peptide moieties can serve as key contributors to metal-peptide interactions. Removal of such moieties undermines ECD efficiency for metallopeptides. More interestingly, different metal ions have different preference towards their interacting moieties, rendering ECD performance of each metal-peptide combination unique. We also observed a novel phenomenon in that metal ion-assisted binding of complementary

peptide fragments can prevent them from being released and observed. Some of the previous research showing the absence of backbone fragments upon metal adduction may be explained by this effect.

In Chapter 3, we attempted to improve EDD by introducing anion adduction. From our results, adduction of salt anions can facilitate electron detachment, but not necessarily improve the abundance of backbone fragment ions. However there is still space for future development of this strategy (see Chapter 5.2). In Chapter 4, we demonstrated that LC/MS/MS can be a powerful tool in pharmaceutical protein production. With targeted LC/MS/MS and customized data processing programs, pharmaceutical protein sequence variants can be detected more sensitively, and quantified more accurately. Additional information provided by this workflow also enabled us to determine whether protein sequence variants originate from translational misincorporation, or from DNA mutation. Such information is valuable for cell line development and clone selection.

During the quantitative data processing of bulk MS/MS spectra with unique adducts and/or isotopic distribution patterns, customized programs contributed significantly in reducing the labor for MS/MS spectral interpretation. An additional customized program was developed, for IR-reactive compound searching in data-independent IRMPD LC/MS/MS as well as a program dedicated to database searching for protein sequence variants. Details of these programs are described in the Appendix and results from data analysis assisted by these programs are demonstrated throughout the thesis.

5.2 Future Outlook

5.2.1 AI-EDD and niECD of Chloride and Bromide Adducted Peptides

In Chapter 3, we attempted to use anion adduction to improve EDD efficiency. Although the results were not ideal, the effect of such adduction remains of interest to examine with AI-EDD. There are two possible hypotheses about why only charge reduced precursor ions are observed to increase in abundance upon bromide adduction: First, salt anion adducts may introduce an additional electron detachment pathway, with such a pathway not generating any backbone cleavages. In this case, it is unlikely that salt anion adduction would benefit the performance of EDD. Alternatively salt anions may be involved in salt bridges. Such salt bridges stabilize the intact precursor ions, thus preventing complementary fragment pairs from being released and observed. IR laser activation may provide additional internal energy to such complexes, thus revealing more backbone fragments. In the future, it would be worthwhile to perform AI-EDD experiments on the same complexes, and confirm whether such activation would have a positive effect on EDD performance. The experimental workflow and data interpretation steps should be the same as described in Chapter 3 for consistency. In addition, the change in relative abundance for different types of fragment ions should be monitored individually. If the distribution of fragment ion types changes, we should consider the possibility of invoking an alternative backbone fragmentation mechanism in the presence of salt anions.

Some peptides cannot undergo niECD, possibly due to inability of forming gaseous zwitterions.[1] Zwitterions formed via a salt bridge are known to benefit electron capture in positive ion mode,[2-3] thus targeted zwitterion generation could have potential for broadening the application of niECD by generating zwitterions in the gas phase for peptides that otherwise do not adopt such structures.

5.2.2 AI-ECD of Metal Supercharged Proteins

During the ionization process, proteins adopt higher charge states if certain reagents are added into the spray solution.[4] This phenomenon is called supercharging.[4-5] For organic additives, supercharging is usually realized by altering the Rayleigh limiting charge of ESI droplets, allowing droplets with higher charge state to exist, and undergo the ESI process described in the charge residue model.[6] More recently, trivalent metal ions were also found to have similar effects on the charge state distribution of proteins in ESI.[7] Such supercharging is particularly significant when aqueous spray solution is used.

From our results in Chapter 2, we found that, with proper IR laser activation, ECD of trivalent metal-adducted peptides is possible even with multiple metal ions present on the same peptide. ECD is known to show enhanced performance at higher charge states,[8-9] so using metal-supercharged protein ions as precursors instead of protonated proteins can potentially improve ECD performance. Such experiments should be conducted by comparing the ECD sequence coverage and fragmentation efficiency of the same protein samples with and without supercharging metal ions. Similar to other published top-down ECD experiments,[8, 10-11] multiple charge states of the protein precursor should be isolated and fragmented at the same time to ensure adquate S/N ratio.

5.2.3 Negative Mode Peptide Mapping for Protein Sequence Variant Analysis

In Chapter 4, targeted LC/MS/MS revealed superior peptide detection sensitivity compared with top ten DDA LC/MS/MS. However, innovation in LCMS/MS experimental strategy will not necessarily improve protein sequence coverage if the same protease is applied in both cases.

Depending on the sequence of a protein and the preferential cleavage site of a protease(s) applied, certain regions of the protein may be improperly cleaved. Peptides from such digestion may be too long or too short for MS/MS identification. In such cases, an orthogonal protease is often applied to ensure MS/MS identification confidence and to improve sequence coverage. The most common protease used in peptide mapping is trypsin, which produces peptides ending in lysine or arginine residues. Both of these residues are basic, thus positive ion mode MS is preferred for the analysis of trypsin digests. When trypsin digestion cannot cleave proteins to the desired degree, some of the common alternative enzymes, such as AspN and GluC,[12] produce peptides ending in aspartic acid or glutamic acid. Potentially, negative mode MS is more preferable for the analysis of the corresponding peptides.

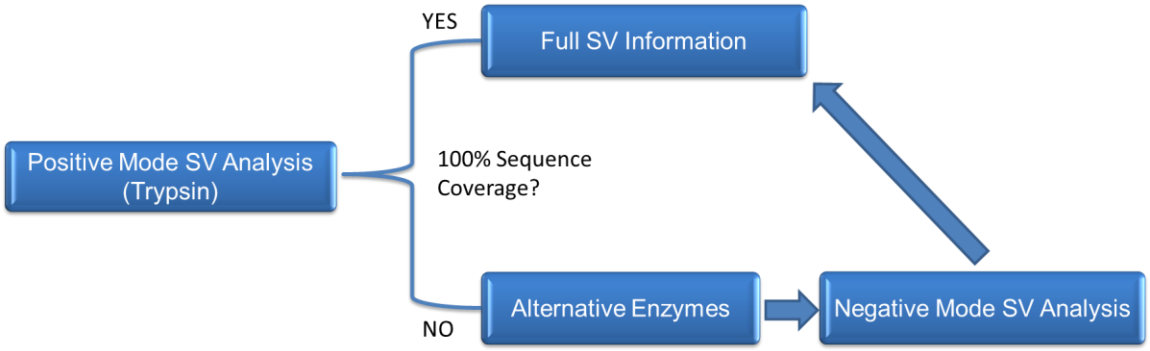


Figure 5.1. Proposed workflow for sequence variant analysis, involving negative ion mode MS analysis.

Because our current mass spectrometry based workflows for protein sequence variant analysis are all based on peptide mapping, issues that affect peptide mapping also affect sequence variant analysis. Peptides with undesirable length produce low quality MS/MS spectra, from which no accurate identification can be obtained. Therefore, use of alternative enzymes will help in identifying protein sequence variants located on peptides that are not extensively

analyzed by trypsin, and a negative ion mode MS-based workflow for protein sequence variant detection would be valuable.

A key step in such a workflow is the selection of an appropriate MS/MS method. Because sequence variant peptide ions are often of low abundance, electron-based methods may not be good options due to their relatively low fragmentation efficiency. Low-energy CAD is also ruled out for the same reason. HCD may be a better candidate, however for most commercially available instruments, pulse programing need to be edited to ensure that HCD fragments are detected in ion traps to ensure speed of detection. Photon-based activation may be a better choice.[13]

5.3 References

1. Yoo, H. J.; Wang, N.; Zhuang, S.; Song, H.; Hakansson, K., Negative Ion Electron Capture Dissociation: Radical-Driven Fragmentation of Charge-Increased Gaseous Peptide Anions. *J. Am. Chem. Soc.* **2011,** *133*, 16790-16793.
2. Creese, A. J.; Cooper, H. J., The Effect of Phosphorylation on the Electron Capture Dissociation of Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2008,** *19*, 1263-1274.
3. Woods, A. S., The Mighty Arginine, the Stable Quaternary Amines, the Powerful Aromatics, and the Aggressive Phosphate: Their Role in the Noncovalent Minuet *J. Proteome Res.* **2004,** *3*, 478-484.
4. Iavarone, A. T.; Williams, E. R., Supercharging in Electrospray Ionization: Effects on Signal and Charge *Int. J. Mass Spectrom.* **2002,** *219*, 63-72.
5. Iavarone, A. T.; Jurchen, J. C.; Williams, E. R., Supercharged protein and peptide lone formed by electrospray ionization. *Anal. Chem.* **2001,** *73*, 1455-1460.
6. Iavarone, A. T.; Williams, E. R., Mechanism of Charging and Supercharging Molecules in Electrospray Ionization *J. Am. Chem. Soc.* **2003,** *125*, 2319-2327.
7. Flick, T. G.; Williams, E. R., Supercharging with Trivalent Metal Ions in Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1885-1895.
8. Iavarone, A. T.; Paech, K.; Williams, E. R., Effects of Charge State and Cationizing Agent on the Electron Capture Dissociation of a Peptide. *Anal. Chem.* **2004,** *76*, 2231-2238.
9. Kalli, A.; Hakansson, K., Comparison of the Electron Capture Dissociation Fragmentation Behavior of Doubly and Triply Protonated Peptides from Trypsin, Glu-C, and Chymotrypsin Digestion *J. Proteome Res.* **2008,** *7*, 2834-2844.
10. Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W., Plasma ECD for the Characterization of Large Proteins by Top-Down Mass Spectrometry. *Anal. Chem.* **2003,** *75*, 1599-1603.

11. Ge, Y.; Lawhorn, B. G.; El Naggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W., Top Down Characterization of Larger Proteins (45 kDa) by Electron Capture Dissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2002,** *124*, 672-678.
12. Swaney, D. L.; Wenger, C. D.; Coon, J. J., Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2010,** *9*, 1323-1329.
13. Brodbelt, J. S., Shedding Light on the Frontier of Photodissociation *J. Am. Soc. Mass Spectrom.* **2011,** *22*, 197-206.

Appendix

Automation of Data Processing for MS/MS and Protein Sequence Variant Database Searching

A.1 Introduction

In biomolecule analysis using mass spectrometry, data processing is often the most time consuming step. Numerous commercial or open source tools have been developed for automated mass spectrometry data processing. However, during the research in this thesis, some data analysis tasks still highly rely on manual data interpretation. Most MS/MS analysis tools do not support analysis of MS/MS with customized adducted ions, such as metal ions with unique isotopic distribution that we have applied in Chapter 2. Also, database searching for low-abundance protein sequence variants in Chapter 4 produce large numbers of false positive search hits, many of which can only be removed through manual verification of MS/MS spectra, EIC peak retention time shift, or even synthetic peptide verification. Therefore, false positive removal is the most time consuming step in peptide mapping.

In order to improve data analysis efficiency, some data analysis tools were developed or customized for data analysis contributing to this thesis. This appendix will explain how these tools function, and their potential application in other data analysis tasks.

Many instrument vendors have their unique data format, such as .baf data index of Bruker instruments, .raw data package of Thermo instruments, and .pkl data package of Waters instrument, etc. There are also open mass spectrometry formats, such as mzXML[1] and mzML.[2] To extract biomolecule ion information from datasets such as their masses and charge states, typical steps include peak detection, feature extraction/deconvolution, charge determination and monoisotopic peak determination.[3] Raw data from ion detector or Fourier transform are in the form of x-y scatter plots. By comparing y-values of points with other points having close x-values, raw peaks can be determined. By algorithms based on isotopic abundance prediction or fitting,[4] isotopic peaks of the same ion can be grouped. For peptides, *m/z* resolving power usually allows isotopic level resolution, thus charge states can be easily determined by isotopic distribution. Monoisotopic peak of peptides is also usually easily discernable. However for larger proteins or protein complexes, isotopic level resolution is often not accessible, and then charge states have to be calculated by neighboring "peak clusters".

A.2 Methods

Sample preparation and data acquisition steps are listed in previous chapters. Mass spectrometry data of different formats are processed using corresponding software, including Bruker DataAnalysis (Bruker Daltonics, Billerica, MA), Thermo Xcalibur QualBrower or QuanBrowser (Thermo Fisher Scientific, Waltham, MA), and Waters MassLynx (Waters

Corporation, Milford, MA). Thermo .raw data file is also accessible using its C#/Visual Basic library.

Most of the customized data processing programs were written in R (version 3.0.2 or above), using RStudio (RStudio, Inc., Boston, MA). False-positive removing macro was written in Visual Basic, built in Microsoft Excel (Microsoft Corporation, Redmond, WA).

A.3 Results and Discussion

A.3.1 MS/MS Matching Tool for Alternative Charge Carriers

Figure A.1 demonstrates the workflow of MS/MS matching tool for alternative charge carriers. In principle, instead using one set of calculated fragment ion mass list, it will calculate multiple fragment ion mass list using different mass adduct. It can also automatically report fragmentation pattern with or without adducted ions, and calculate fragmentation efficiency based on normalized-to-charge monoisotopic mass peak abundance.

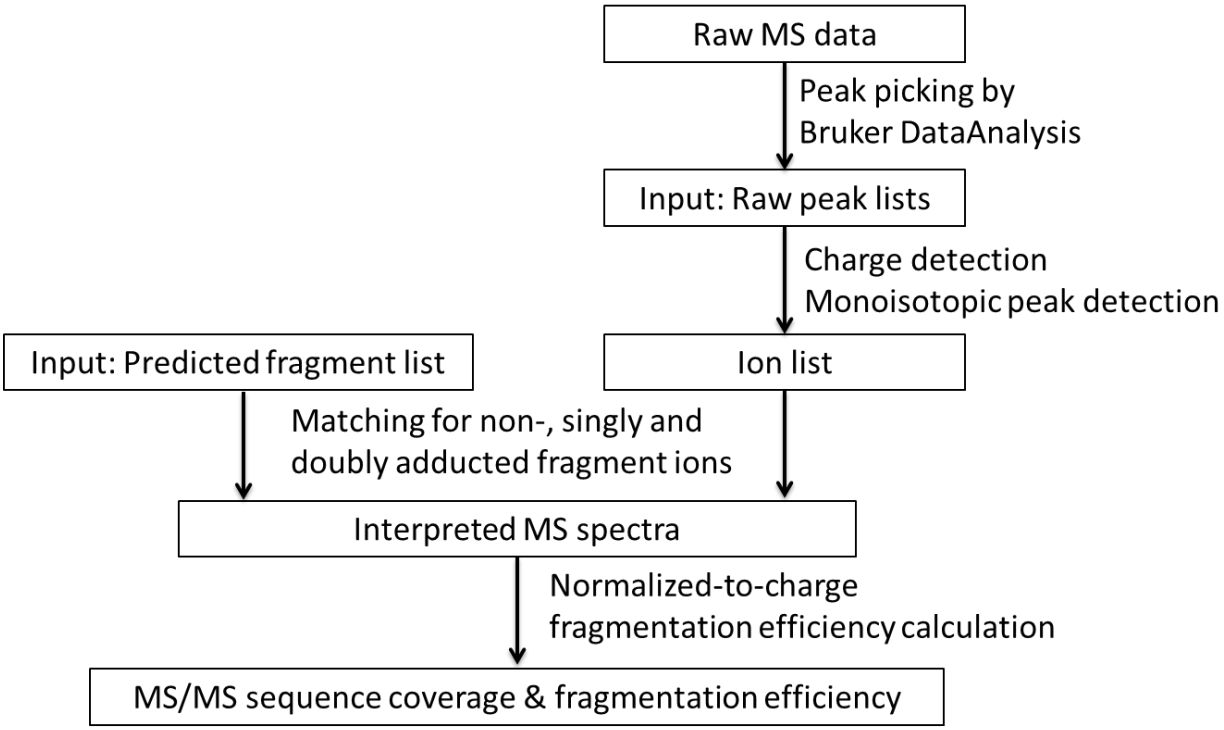


Figure A.1. Flow chart of MS/MS matching tool for alternative charge carriers.

This program is intended to address only small peptides (<20 amino acid residues) MS/MS experiments, performed on directly infused samples. Compared to LC-MS/MS spectra of digested protein, such offline small peptide MS/MS spectra usually have better signal to noise ratio and mass resolution. Also, precursor charges in such experiments performed in this thesis are less than 5. Therefore, multiple isotopic peaks can be discerned for each ion, and charge determination is possible by counting m/z difference between isotopic peaks. Since the isotopic distributions of adducted ion are different, any form of isotopic distribution cannot be used for monoisotopic peak determination. Monoisotopic peaks are determined solely by using the peak with the lowest *m/z* in each peak cluster.

Testing of this program was done by comparing sequence coverage and fragmentation efficiency of the same MS/MS spectrum from program analysis and manual analysis. ECD

spectra of La- and Eu-adducted Abeta16 were selected for testing. Lanthanum has only one stable isotope: La^{139} (natural abundance: 99.91%), while europium has two: Eu^{151} (natural abundance: 47.81%) and Eu^{153} (natural abundance: 52.19%). Therefore, lanthanum will not alter isotopic distribution of fragment ions, while europium will. Among all alternative charge carriers involved in this thesis, copper, nickel, chloride and bromide can also change the isotopic distribution of fragment ions. Testing results indicate that sequence coverage and fragmentation efficiency result from the program is consistent with manual analysis.

In application of this program in data analysis of data for Chapter 2 and Chapter 3 projects, it significantly improved data analysis throughput. Manual elaboration of a MS/MS spectrum with adducted fragment ions usually takes 10-15 mins, while the program takes less than one minute. Such high throughput enabled us to sensitively trace fragmentation efficiency change caused by alternative charge carrier adduction.

A.3.2 LC-MS Searching Tool for Low-abundance Protein Sequence Variants

In the targeted LC-MS/MS workflow described in Chapter 4, inclusion list generation is a key step that no commercially available software can perform. A customized R program was developed to achieve this function. The flow chart of this program is shown in Figure A.2.

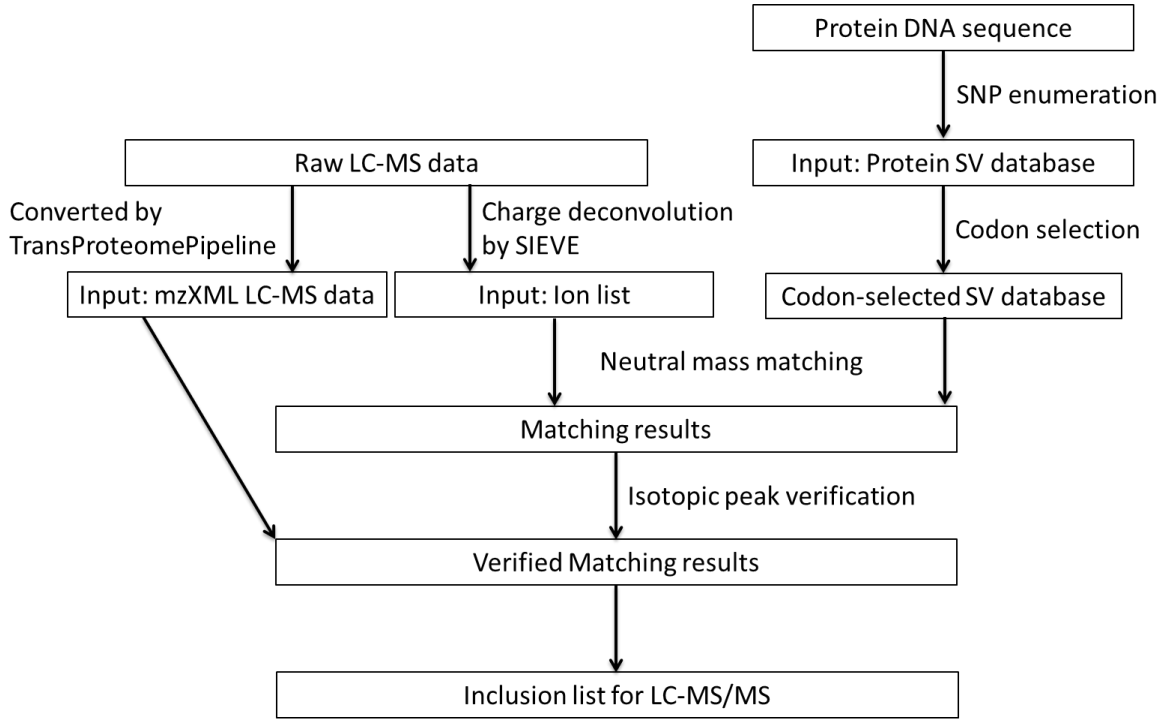


Figure A.2. Flow chart of LC-MS searching tool for protein sequence variants.

This program requires three input files that need to be generated outside the program: a protein sequence variant database, which comes from protein sequence with all single nucleotide polymorphism (SNP) possibilities; an ion list generated by SIEVE (Thermo Fisher Scientific, Waltham, MA), and raw data in mzXML format, converted by TransProteomePipeline or other software. When generating ion list by SIEVE, feature number should be set as high as the computer's memory allows. SIEVE also has a built-in LC-MS matching function, but it cannot correctly recognize monoisotopic peaks at low intensity. The sequence variant database is codon-selected before searching, so that only sequence variants with specific codon changes are retained (see Chapter 4).

The program performs neutral mass matching, based on calculated neutral mass values of peptides in database and neutral mass values from ion list. For each ion that has a SV match, the program will go to raw data (in mzXML format) for confirmation. Since many ions are very low

in abundance (S/N ratio <=100), it may be confused with noise peaks. However, considering the atomic composition of peptides, the m+1 isotopic peak abundance is largely determined by carbon mass percentage of the peptide (M_c/M). Phenylalanine has the highest carbon mass percentage, while cysteine has the lowest carbon mass percentage. Therefore, the m+1 isotopic peak abundance of a peptide composed of phenylalanine only should be the highest abundance value, and the m+1 isotopic peak abundance of a peptide composed of cysteine only should be the lowest abundance value of all peptides. Noise peaks are very unlikely to have isotopic distribution pattern like a peptide, thus peaks with m+1 isotopic peak abundance falling out of this "F-C" zone is likely to be a noise peak. In actual verification process, this window is usually widened for 15%-30% to compensate error brought about by low S/N ratio and instable signal. By ruling out peaks with incorrect isotopic distribution, the program significantly reduces false positives caused by noise peak matching.

A.3.3 Wild Type Decoys for Improved Protein Sequence Variants Searching

Detection of low-level protein sequence variants requires sensitive analytical techniques. Due to that most PSV occur in low level, false positive is common in MS/MS database search. These false positives may include modifications, adducts and in-source fragments. Herein, we introduce a strategy for LC-MS/MS database searching using wild type peptides as decoys to rule out false positives. Search for sequence variant peptides and wild type peptides were conducted separately. MS/MS spectra with search hits from wildtype were removed in search results of sequence variant peptide. We tested this strategy on artificial sequence variants from protein mixture, as well as monoclonal antibody with low level sequence variants.

In order to search for PSV with higher accuracy, wild type peptides are used as decoys in MS/MS database search. Figure A.3 shows the workflow in Thermo Proteome Discoverer. A Microsoft Excel Macro was built to remove search results that hit both databases. (Figure A.4)

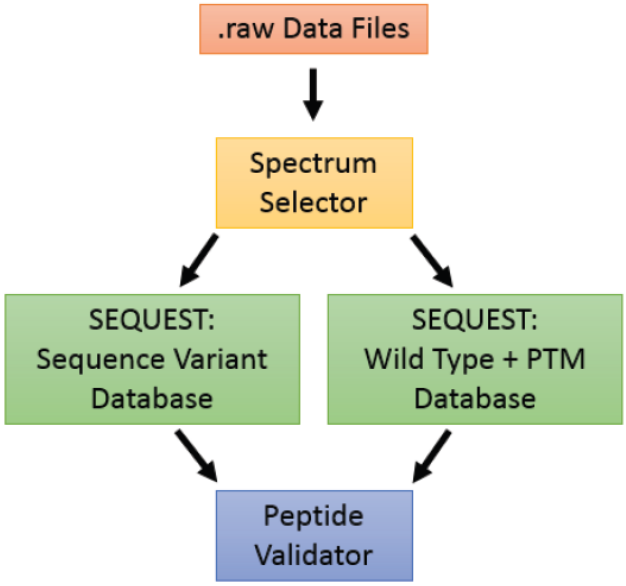


Figure A.3. Thermo Proteome Discoverer workflow for PSV database searching.

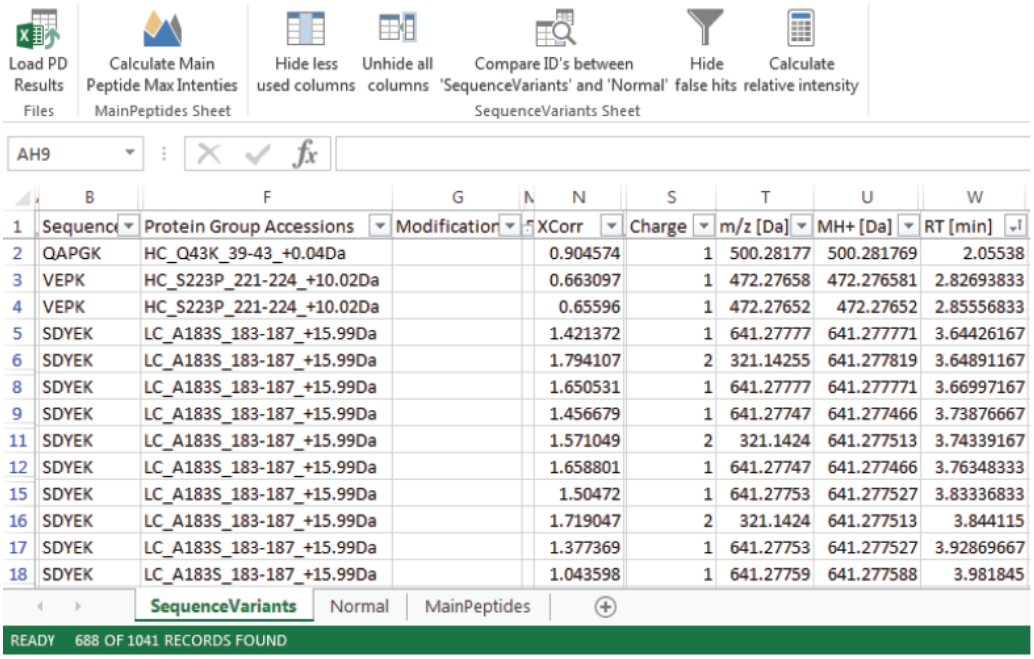
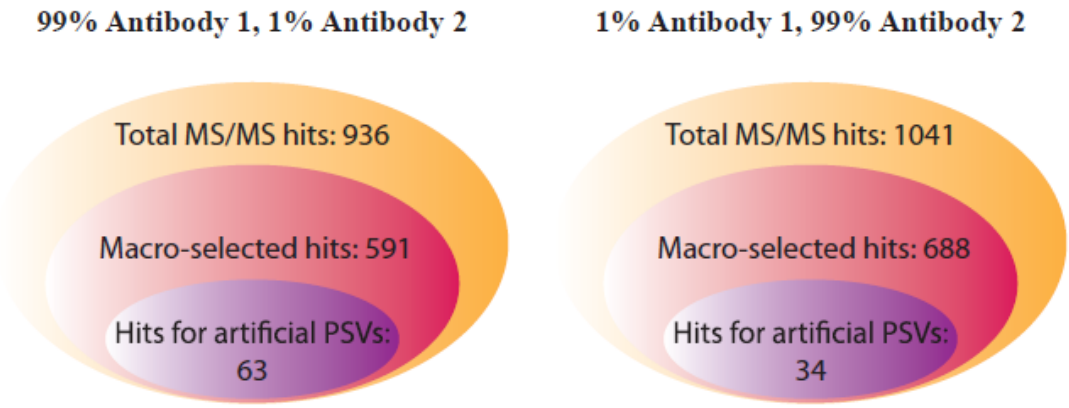


Figure A.4. In-house Microsoft Excel macro user interface for false positive removal.

We validated the method with artificial SVs (antibody mixtures of Antibody 1 and 2, as described in Chapter 4) to ensure that false negatives will not occur in the workflow when detecting 1% level SVs. Such false negatives may originate from erroneous removal of SV search hits. From our result (Figure A.5), 11 out of 12 artificial PSVs out in two antibody mixtures were correctly detected. The missing PSV is due to that its precursor ions were not selected for MS/MS. More than 30% of false positive hits were removed by the macro.

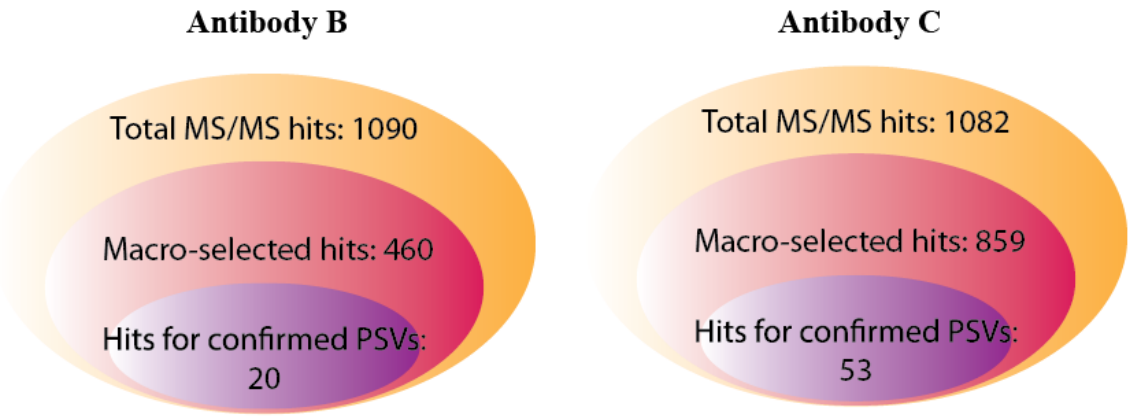


PSV peptides, from 1% Antibody 2	MS/MS hits	EIC peaks	PSV peptides, from 1% Antibody 1	MS/MS hits	EIC peaks
VESK	2	1	VEPK	2	1
ATGIPAR	5	1	ATGIPDR	5	1
SLSLSLG	2	1	SLSLSPG	0	0
EPQVYTLPPSQEEMTK	37	1	EPQVYTLPPSR	11	1
EIVLTQSPATLSLSPGER	8	1	EIVLTQSPGTLSLSPGER	8	1
TTPPVLDSDGSFFLYSR	9	1	TTPPVLDSDGSFFLYSK	8	1

Figure A.5. Numbers of MS/MS hits before/after adding wild type as decoys, and hits representing each artificial PSV peptide. Peptide SLSLSPG was not found before false positive removal.

After false negative test, we moved on to test the program on top 10 DDA LC-MS/MS data acquired on protein digest samples with actual sequence variants. Antibody B and C have already been extensively analyzed in Chapter 4. One high level DNA error in Antibody B and

multiple low level translational errors in both antibodies were confirmed and quantified. From our result (Figure A.6) we can see that: percentage of false positives removed varies by samples. Low level translational error search hits were preserved in false positive removal.



PSVs	MS/MS hits	Relative Abundance
HC K370N	6	62%
HC S408N	2	0.18%
HC S142N	2	0.09%
HC S304N	2	0.07%
HC C230Y	3	0.06%
HC C261Y	3	0.06%
HC V305L	2	0.02%

PSVs	MS/MS hits	Relative Abundance
HC N106K	1	1.29%
LC A60S	7	0.60%
HC N374S	3	0.24%
HC N328K	4	0.23%
HC N403S	6	0.15%
LC N53S	2	0.12%
HC N78S	2	0.12%
HC S421N	2	0.11%
HC N328S	4	0.10%
HC S453N	2	0.09%
LC S12T	5	0.04%
HC N403K	1	0.04%
HC V315L/I	10	0.03%
HC V286L/I	2	0.03%
HC V277L/I	2	0.02%

Figure A.6. Numbers of MS/MS hits before/after adding wild type as decoys, and hits representing each PSV peptide in Antibody B and C.

In general, this program showed that wild type peptides can be used as decoys in database searching for protein sequence variants of LC-MS/MS peptide mapping data. Adding decoys will not cause loss of true positive hits, nor affect sensitivity significantly. Percentage of false positive search hits that can be removed by adding decoys varies by sample and data. The whole decoy adding-false positive removal workflow can be highly automated.

A.4 Conclusions

As shown in this appendix, customized programs can be applied in various aspects of mass spectrometry data analysis. In the analysis of individual MS/MS spectra, our customized program can analyze spectra that are complicated in one degree (having unique isotopic distribution and mass adduction) while simple in some other (low charge, known sequence). In LC-MS and LC-MS/MS database searching, customized programs can reduce human labor involved and fit specific requirements such as high sensitivity. Programing capability enables us to solving unique mass spectrometry questions with higher flexibility than commercial programs.

A.5 References

1. Pedrioli, P. G. A.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian Jr, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R., A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotech* **2004,** *22*, 1459-1466.
2. Deutsch, E., mzML: A single, unifying data format for mass spectrometer output. *Proteomics* **2008,** *8*, 2776-2777.
3. Katajamaa, M.; Oresic, M., Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A* **2007,** *1158*, 318-328.
4. Tautenhahn, R.; Bottcher, C.; Neumann, S., Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **2008,** *9*.