DISULFIDE ANALYSIS WITH MASS SPECTROMETRY: TOWARD BIOTHERAPEUTIC CHARACTERIZATION

BY

DANIEL FORREST CLARK

Submitted to the graduate degree program in Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson: Dr. Heather Desaire

_____

Dr. Robert Dunn

_____

Dr. Mario Rivera

_____

Dr. Michael Johnson

_____

Dr. Roberto De Guzman

Date Defended: June 16, 2015

The Dissertation Committee for DANIEL FORREST CLARK
Certifies that this is the Approved Version of the Following Dissertation:

**DISULFIDE ANALYSIS WITH MASS SPECTROMETRY: TOWARD
BIOTHERAPEUTIC CHARACTERIZATION**

_____
Chairperson: Dr. Heather Desaire

Date Approved: June 16, 2015

Abstract:

Proteins continue to increase in use as biotherapeutic agents but need to be analyzed to verify efficacy and safety. For instance, disulfide bonds are a protein posttranslational modification (PTM) that are critically important to protein stability and function; therefore, they also impact safety and efficacy. Proteins can assume non-native bonds during recombinant expression, sample handling, or sample purification, leading to undesirable products. Currently, mapping of disulfide bonds is problematic due to sample requirements and data analysis difficulties. In this dissertation, many aspects of disulfide bonding analysis are improved. First, the fragmentation products of disulfide bonded peptides are analyzed to assist efforts in the automated analyses of these species when the analytical workflow involves LC-MS and collision-induced dissociation. Additionally, a new analysis method is introduced to more easily assign disulfide bonding partner peptides. This method utilizes a recent innovation in fragmentation, electron transfer dissociation. The method can also be used to easily detect alternate disulfide bonding patterns and confirmed the findings that there are multiple disulfide structures in a recombinant HIV glycoprotein, gp120. The method is then expanded to incorporate reporter peptides to ensure the sample preparation is not causing these alternate structures. Finally, a software tool capable of automated assignments is also outlined to expedite the data analysis. In sum, this dissertation advances the field of disulfide analysis of proteins, by introducing new methods to expedite the work and new techniques to ensure the reliability of the analyses.

Acknowledgements

First and foremost I would like to thank the Chemistry faculty of the University of Kansas, especially Dr. Heather Desaire. She is a brilliant scientist that always pushes graduate students to constantly improve themselves. Her energy and enthusiasm for both research and teaching were refreshing to even the weariest student. I thank all past and present members of the Desaire group. I greatly appreciate those that laid the foundation for the group as well as the stimulating conversations that took place in and around the laboratory. Dr. Eden Go, I cannot thank you enough for the help you provided with research, data analysis, personal problems, and all the conversations about the basketball game(s) the night before. You always offered time to listen, even when deadlines were looming.

Dr. David McCurdy, thank you for giving me a chance in your undergraduate research lab. I was ecstatic when you offered me a position on one of your projects and it was this opportunity that launched my interest in research.

Thank you to my mother and father, Peggy and Danny; sisters, Margaret and Melissa; and brother, Jonathon; aunt and uncle, Martha and Allan; for the loving support that I needed through graduate school. Thank you to all the friends I've made while attending the University of Kansas. Your support and feedback is very much appreciated.

Table of Contents   **Page**

Chapter 1: Introduction

The work in this dissertation describes a dramatic improvement in the analysis of disulfide bonds in proteins.

1.1 Importance of Disulfide Bonds:

The disulfide bond is a post-translational modification (PTM) that imposes a strict distance constraint via a covalent bond between two distant cysteine residues in the protein, and it is this nature that makes the disulfide bond unique among most other PTMs. Anfinsen first showed the important role disulfide bonds had in ribonuclease through a series of scientific experiments where he denatured, reduced, oxidized, and renatured the protein.[1] He was able to show that when cysteines are oxidized to disulfides under denaturing conditions the protein has extremely diminished activity compared to cysteines oxidized under protein renaturing conditions.[1] This led to the theory that native, or correct, disulfides must be present for a protein to function properly as non-native disulfides force an entirely different tertiary structure.[2-5] A major complication to the use of biotherapeutic agents in medicine is shuffling or changing of disulfide bonds to non-native.[6] Because disulfide bonds play such a key role in stabilizing and maintaining protein tertiary structure, incorrect disulfide bonds often lead to decreased efficacy or possible immune responses.[6] When a free cysteine is present, whether native or due to incomplete oxidation to a disulfide bond, it can attack other disulfides and lead to shuffling (**Figure 1**). Disulfides can shuffle at many points during

protein production such as during expression, purification, and storage. Disulfide mapping should be performed before products are used in human or animal studies to be sure correct disulfide bonds are in place in the final product.
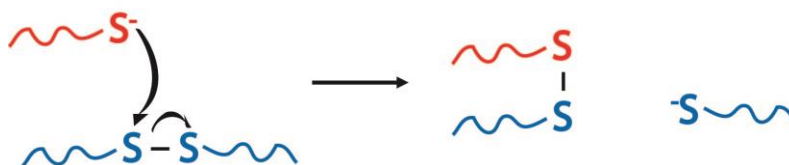
Figure 1: Free thiols may react with disulfide bonds to produce a shuffled, or scrambled disulfide.

1.1.1 Sample Preparation for Disulfide Mapping

Proper sample preparation is a necessity to analyze disulfides because it must both decrease the complexity of the sample for analysis but also not alter the sample in such a way to produce artifacts that may cause incorrect conclusions.[3,7-9] Analyzing disulfide bonds is not trivial, particularly due to the fact that they serve to hold portions of the protein together and can cause steric concerns when proteases are used for part of sample preparation. Chaotropic reagents are commonly used to denature the protein for better protease access to the peptide backbone but must always remain in a non-reducing environment to keep the disulfides intact. In the analysis of the disulfide bonds of heavily glycosylated proteins, such as the HIV Envelope protein, gp120, the main concerns are the extensive glycosylation present and the preservation of native disulfide bonds.[7] If the glycans were not removed prior to disulfide analysis on these proteins, the masses of each of the disulfide bonded peptides would be variable, and these species would be nearly impossible to assign using a mass spectrometry-based analysis. Fortunately there is an enzyme that releases *N*-linked glycans and does so

rapidly and with excellent selectivity. Peptide-N-glycosidase F, PNGase F for short, can be used to reliably and efficiently remove N-linked glycans.[10-12] Deglycosylation of *N*-linked sites with PNGase F, depicted in **Figure 2**, also causes deamidation of the asparagine residue which causes a +0.9840 Dalton mass shift from the amide being converted to a carboxylic acid (aspartate) group.[10-12]
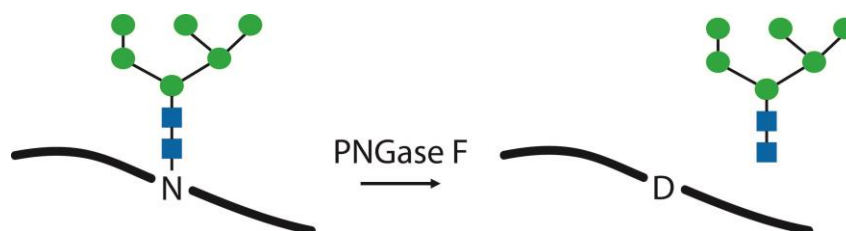
Figure 2: PNGase F reacts with *N*-linked glycosylation sites to release the glycan and deamidate the asparagine.

Other sample preparation for the analysis of disulfide bonds is dependent upon the analysis technique. Often the use of a protease, such as trypsin, is employed to decrease the size of peptides so they may be analyzed.[13-16] Also common is the introduction of an alkylating reagent, such as 4-vinylpyridine, iodoacetamide, or n-ethylmaleimide, during sample preparation. These alkylating reagents react with free cysteine residues to decrease or eliminate disulfide shuffling. Time, temperature, and pH of any deglycosylating or alkylating reactions are also important factors to consider when performing sample preparation for disulfide mapping as they could introduce the possibility of disulfide shuffling.

1.2 Methods for Disulfide Mapping:

1.2.1 SDS-PAGE, NMR, and X-ray

Disulfide bonded peptides have been studied for many decades and the methods of analysis have increased in speed and sensitivity in that time. One of the first methods used to map disulfide bonds was two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (2-D SDS-PAGE).[17,18] This method requires a step after running the first dimension of gel electrophoresis to break the disulfide bond and then capitalizes on the changed electrophoretic mobility of peptides that were disulfide bonded. Peptides that were not disulfide bonded have the same mobility and form a diagonal line across the gel but peptides that were disulfide bonded will lay off the diagonal because breaking the disulfide effectively changes the peptide such that the mobility will change. Analysis with 2-D SDS-PAGE has significant limitations due to poor detection limits and thus large sample requirements. The method also requires an extremely pure protein or else the gel would be complicated with extra spots from different proteins. Additionally, SDS-PAGE has no direct means of identifying what the spots contain. In order for identification to take place, spots from the gel must be excised and analyzed, usually with mass spectrometry.[19] Nuclear magnetic resonance (NMR) spectroscopy and X-ray crystal diffraction spectroscopy also represent two methods commonly used to map disulfide bonds. The two methods provide excellent three-dimensional information and provide the entire structure of a protein. However, NMR and X-ray crystallography suffer from drawbacks such as requiring large quantities of very pure sample. Data analysis with NMR and determining optimal crystallization

conditions necessitate copious amounts of time and work, and these methods, particularly NMR, are not amenable to large, glycosylated proteins.

1.2.2 Liquid Chromatography Mass Spectrometry

Modern disulfide analyses most commonly employ liquid chromatography mass spectrometry (LC-MS) due to readily accessible instrumentation, minimal chemical requirements, low detection limits, excellent selectivity, small sample amounts, and the option for fragmentation of the analyte. Despite all the benefits of using LC-MS, there are some problems with its current state. Fragmentation of disulfides within the mass spectrometer can provide unique structural details and be used as a way to confirm the disulfide bonded peptide. However, analyzing these fragments by hand is tedious and time-consuming. While there are software programs available to automate assignments of disulfides, these were typically developed for single, linear peptides and do not take into account fragmentation that is different in disulfide bonded peptides. In order to assign as many fragments as possible, the fragmentation patterns of disulfides must also be analyzed and the unique fragments utilized in the automated assignment software. This work represents efforts to solve these problems and in order to completely understand these issues, an introduction to the technique of mass spectrometry is required.

1.3 Liquid Chromatography with Mass Spectrometry:

1.3.1 Electrospray Ionization

The first hurdle of an analysis with mass spectrometry is ionization of the analyte. If the analyte is not already an ion, ionization is typically achieved either through the loss or addition of a proton. Peptides and disulfides have amine and carboxylic acid functional groups that allow the molecules to ionize with relative ease. The charge carried by the analyte is of utmost importance in mass spectrometry because it enables the use of electric fields to manipulate it and neutral species will not respond to the same forces. The advent of electrospray ionization (ESI) has had one of the most profound impacts of any one technology to mass spectrometry, and it opened the door to analyses that would have otherwise been impractical or impossible.[20] ESI allows the online introduction of liquid chromatography (LC) effluent to analysis by mass spectrometry at typical LC flow rates. LC effluent cannot be directly injected into a mass spectrometer because the evaporation of solvent would cause an extremely large gas load that the vacuum pumps could not evacuate. Equally important to the coupling of LC and MS was the fact that ESI opened new avenues in LC-MS analysis because, as a soft ionization source, it could ionize non-volatile and thermally labile analytes.[20] Many biologically-important analytes are either non-volatile or thermally labile, and a suitable ionization technique for these did not exist prior to ESI.
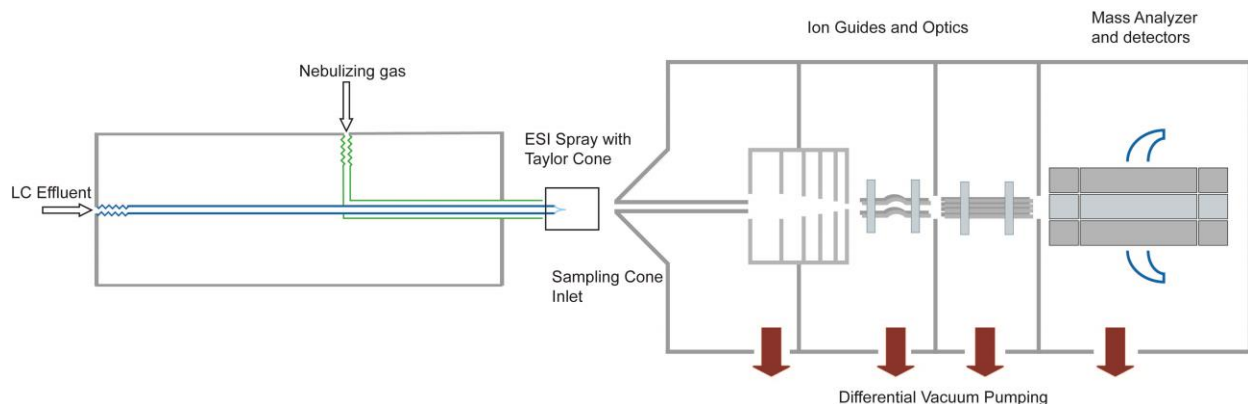
Figure 3: ESI allows coupling liquid chromatography to mass spectrometry.

ESI works by applying a high potential (~4 kV) to a capillary where the sample or LC effluent is exiting (**Figure 3**). A Taylor cone forms at the tip of the spray needle due to the high potential applied (**Figure 4**). ESI typically also employs the use of dry nitrogen gas that flows concentric with the spray needle for two reasons: To help nebulize the droplets formed at the tip of the Taylor cone and to aid the evaporation of the solvent. There are two competing theories on the mechanism of ionization in ESI. One theory focuses on the large electric field gradient that exists between spray needle and MS inlet. The inlet of the mass spectrometer is held at approximately 0 V and the electric field gradient is strong enough (~4 kV/cm) that the droplets that form at the tip of the Taylor cone are essentially ripped apart.
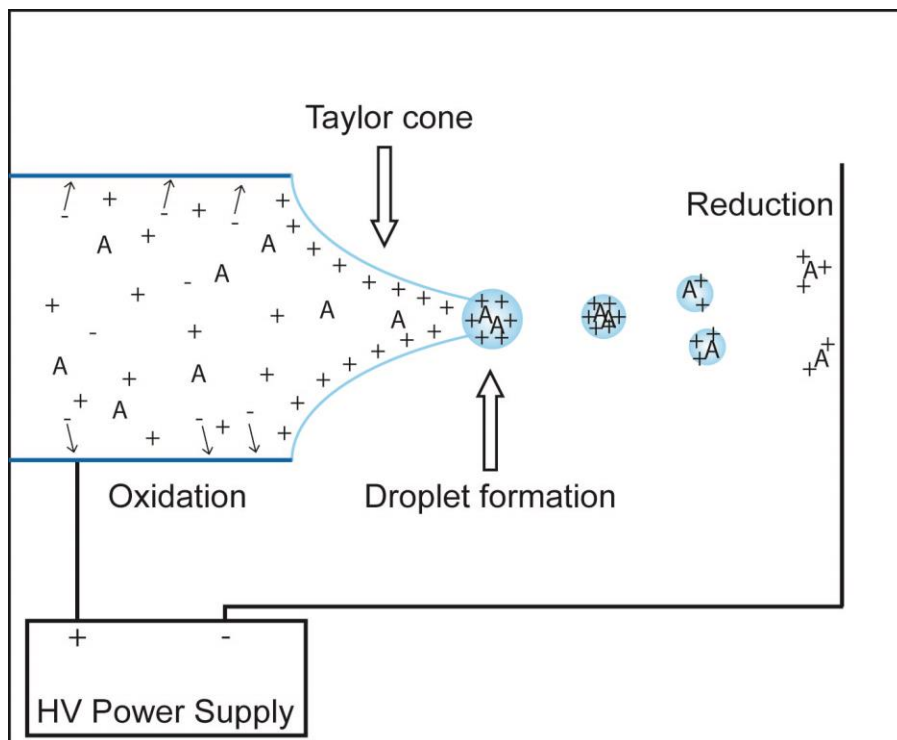
Figure 4: A Taylor cone forms at the ESI needle and ionic droplets are produced from the high potential applied.

The other theory focuses on evaporation of solvent from each droplet and the idea that droplets formed will have a specific amount of charge. If the droplet size is decreased by evaporating solvent, repulsion of charges will occur. If the droplet size is reduced to the point the Rayleigh limit is reached, a droplet explosion results due to repulsion. Once all the solvent has been evaporated the charges will remain on the now gas phase analyte molecules. This leads to both one of the biggest drawbacks and benefits of using ESI: Analyte molecules can take on multiple charge states.

Multiple charging can be a drawback because it clutters the resulting spectrum with a range of *m/z* values, but it is also beneficial because very large molecules, such as proteins, can be analyzed with low cost ion trap instruments when the analyte *m/z* is in the instrument's scan range. Since the ESI process also evaporates solvent around

the analyte molecules, there is substantial adduct formation where the charge carriers are something other than protons in the sample. Common adducts other than protons for positive mode include sodium, potassium, and ammonium. Adducts will change the mass of the analyte and the respective *m/z* and should be minimized if at all possible.

The analyte molecules have been charged through the ESI process at atmospheric pressure and are not ready to be analyzed with MS because mass spectrometry requires vacuum ($<10^{-5}$ torr) for analysis. The vacuum serves to increase ion transmission through decreased gas-ion collisions. To transfer ions from atmospheric pressure into a reduced pressure environment, hardware adaptations are used to limit the amount of gas that enters the inlet. The sampling cone is the first barrier limiting what is able to enter the low vacuum side of the instrument (**Figure 3**). It greatly reduces the amount of drying nitrogen allowed to enter and leads to the skimmer cone. The skimmer further reduces the amount of nitrogen present and leads to ion optics to guide ions into the analyzer region. Each section of the instrument is differentially pumped such that the analyzer region has the highest vacuum. Once in the mass analyzer, the *m/z* of the ions can be measured.

1.3.2 Mass Spectrometry Analyzer Fundamentals

Mass spectrometry is an analytical technique that uses electric and/or magnetic fields to determine the mass to charge ratio of ions by exploiting fundamental physics. Mass spectrometry applies electric and/or magnetic fields to ionized molecules and the corresponding effect is indicative of the mass to charge ratio, or *m/z*. For instance, the magnetic sector mass analyzer represents one of the simplest designs. As the name

implies, magnetic sector instruments use a strong magnetic field to apply a force to ions that enter the analyzing region. The force, F_B, felt by the ions is related to the magnitude of the magnetic field, B, the charge of the ion, q, and the velocity, v, by the equation

$$F_B = Bqv \qquad \text{Eq. 1}$$

The force applied to the ions is perpendicular to the magnetic field and ion motion resulting in a curved path through the magnetic sector with a defined radius, r. The competing force, centrifugal force, must counterbalance the magnetic force such that

$$Bqv = \frac{mv^2}{r} \qquad \text{Eq. 2}$$

and solving for Br provides that magnetic sector instruments separate ions based on a momentum to charge (mv/q) basis.

$$\frac{mv}{q} = Br \qquad \text{Eq. 3}$$

Because the radius of the instrument is a constant and the magnetic field can be changed, magnetic sector instruments can measure the momentum to charge ratio of ions based on the magnitude of the magnetic field when the ions have a trajectory that reaches the exit slit and detector. Similarly, electric fields can be used to apply a similar force by which ions are separated according to momentum or mass to charge ratios. Many different mass analyzers exist in addition to the magnetic sector, and each has its benefits and limitations. For the purpose of the majority of work presented in the rest of this document, linear ion trap and Fourier transform analyzers were used.

1.3.2.1 Ion Traps

The ion trap represents a very common mass analyzer, due to the cost, ease of maintenance, and the capability of performing tandem mass spectrometry. It is employed for low resolution and low mass accuracy measurements. Ion traps in general operate by exploiting electric potentials and specific geometries to produce an electric field that effectively holds ions within the trapping region.[21] A linear, or 2D, trap is typically constructed of four metal rods that ideally have hyperbolic geometry (**Figure 5**) and two smaller sections at the ends. Opposite rods are connected electrically in each section and three separate DC voltages are applied to the three sections of the trap. Ions are guided into the ion trap by ion optics while the DC potential applied to the front section is momentarily shut off.

Figure 5: Linear ion traps consist of four metal rods and opposite rods are electrically connected. Ions are trapped by applying DC potential and RF to the main section and a DC potential to the end sections.

Once the ion trap is "filled" the DC potential to the front section is reapplied and the ions are imprisoned using the electric field generated by AC and DC potentials of the four rods and two end sections. When the applied potentials hold a given ion, the ion is considered to have a stable trajectory within the trap. Potentials can be manipulated such that the trajectory of an ion or large set of ions is no longer stable and these ions are ejected. A set of differential equations, the Mathieu equations, are used to define what electrical potential values represent situations where the ions will be stable inside the trap. The dimensionless values a and q are represented by

$$a_x = -a_y = \frac{8eU}{m\Omega^2 r_0^2} \qquad\qquad q_x = -q_y = \frac{4eV_{RF}}{m\Omega^2 r_0^2} \qquad\qquad \text{Eq. 4}$$

where e is the magnitude of the electron charge, U is the DC potential, V_{RF} is the zero to peak amplitude of the RF potential, m is the ion mass, Ω is the frequency of the alternating component, and r is the radius of the circle produced at the center of the four metal rods. A brief investigation of the equations reveals that values a and q are most easily modified by changing either the DC potential or the RF potential respectively. The stability diagram in **Figure 6** represents a region where ions are stable both in the X and Y axes of the trap.
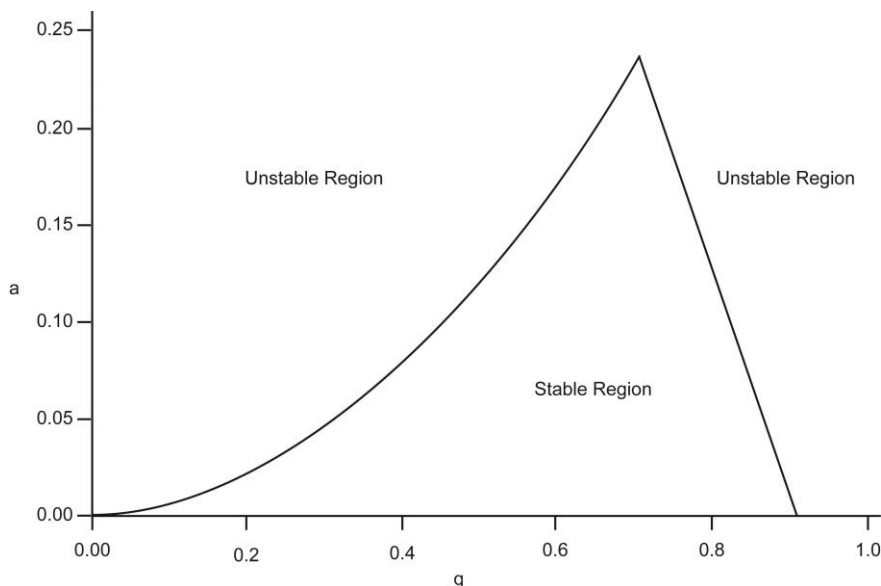
Figure 6: The Mathieu equations dictate regions where ion trajectories are stable in an ion trap.

Knowledge of this stability diagram is advantageous to the user in several ways. First, the ions can selectively be ejected with mass-selective instability mode from the trap by simply adjusting the potentials applied. Assuming the DC potential is zero, then the value of *a* is equal to zero and all ions are trapped by the field created by the RF potential. The *m/z* values of the ions can be selectively measured using a mass selective instability mode when the main RF voltage is increased in a linear fashion. The increasing V_{RF} has the effect of changing where ions are located on the stability diagram. Take for instance three theoretical ions in **Figure 7**.
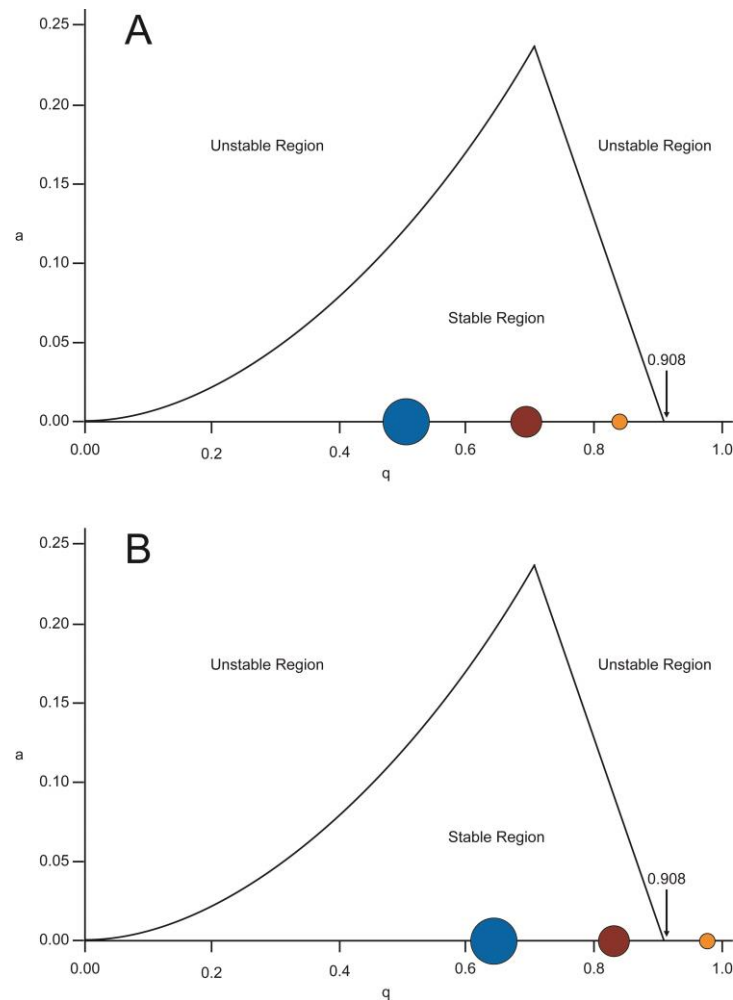
Figure 7: A representation of mass-selective instability mode. A) Three ions have stable trajectories inside an ion trap. B) The RF potential is increased to a point where the smallest ion is no longer stable and is ejected.

Figure 7A represents the three ions that are trapped in an ion trap at their relative natural resonant frequencies which can be solved with Equation 4. When the RF potential is increased, each of the ions' q value increases and results in a shift on the stability diagram. When the q value exceeds 0.908 the ions are ejected from the trap. **Figure 7**B indicates that the smallest ion has been ejected while the other two still have stable trajectories within the ion trap. Often this mass-selective instability mode is

augmented by utilizing resonance ejection because very high potentials (several kilovolts) are required to scan a large mass range and arcing between metal surfaces is possible. Resonance ejection essentially creates a "hole" on the stability diagram by taking advantage of the fact that all ions have a tendency to oscillate when a secular frequency is felt. Applying such a supplementary RF potential to the end caps forces any ions at that *q* value to resonate with enough energy to eject them from the trap. Resonance ejection makes it possible to have a larger scan range with the use of lower potentials to perform the same separation. Detection of the ions is accomplished when the ions are ejected through slits in two opposite rods of the main section when they become unstable. Upon ejection from the trap, the ions collide with the conversion dynode of an electron multiplier that produces electrons in place of the analyte ion. These electrons are sent down a cascade of dynodes that are held at sequentially higher potentials which accelerate the electrons.   Each dynode produces more electrons than the previous one due to being accelerated by the potentials and the electrons are then detectable as a current and the potentials used to eject the ions are used to calculate an *m/z* value and displayed as a mass spectral peak.[22-27]

1.3.2.2 Fourier-Transform Ion-Cyclotron Resonance

Fourier-transform ion-cyclotron resonance mass spectrometry (FT-ICR-MS) is another common ion trapping technique, and it provides much higher resolution and mass accuracy but at higher initial cost and more intense maintenance. It also requires a longer scanning time. This instrument design utilizes a Penning trap which has three sets of parallel plates: One set for trapping, one set for transmitting, and one set for

detecting arranged in a cube-like geometry (**Figure 8**). The FT-ICR combines the trapping ability of electric fields while exploiting the physical phenomenon that when a very strong magnetic field is applied in the Penning trap every, *m/z* will have a unique frequency at which it cycles the trap, termed the cyclotron frequency.[28] The cyclotron frequency, ω_c, can be incorporated assuming a circular trajectory which has a velocity, v

$$v = 2\pi r\omega_c \qquad \text{Eq. 5}$$

where r is the radius of the Penning trap. When this is substituted into Equation 3 and solved for ω_c, the result is

$$\omega_c = \frac{qB}{2\pi m} \qquad \text{Eq. 6}$$

where *q* is the charge, B is the magnetic field, and m is the mass of the ion. This equation clearly provides that every ion with a unique *m/z* (represented by m/*q* in the inverse of Equation 6) will also have a unique cyclotron frequency. Ions are introduced into the trap and remain toward the center until an excitation chirp of a range of frequencies to excite all ions is transmitted, and the ions begin to cycle the trap. As the ions continue to cycle around the Penning trap at their unique cyclotron frequencies, they pass by the detector plates, and the ions induce an image current. This image current is used to construct the data file with the initial data being of the time-domain, representing the frequencies that were observed. A Fourier transform can be performed with knowledge of the magnetic field, and the frequencies are correlated to an ion's *m/z* value.
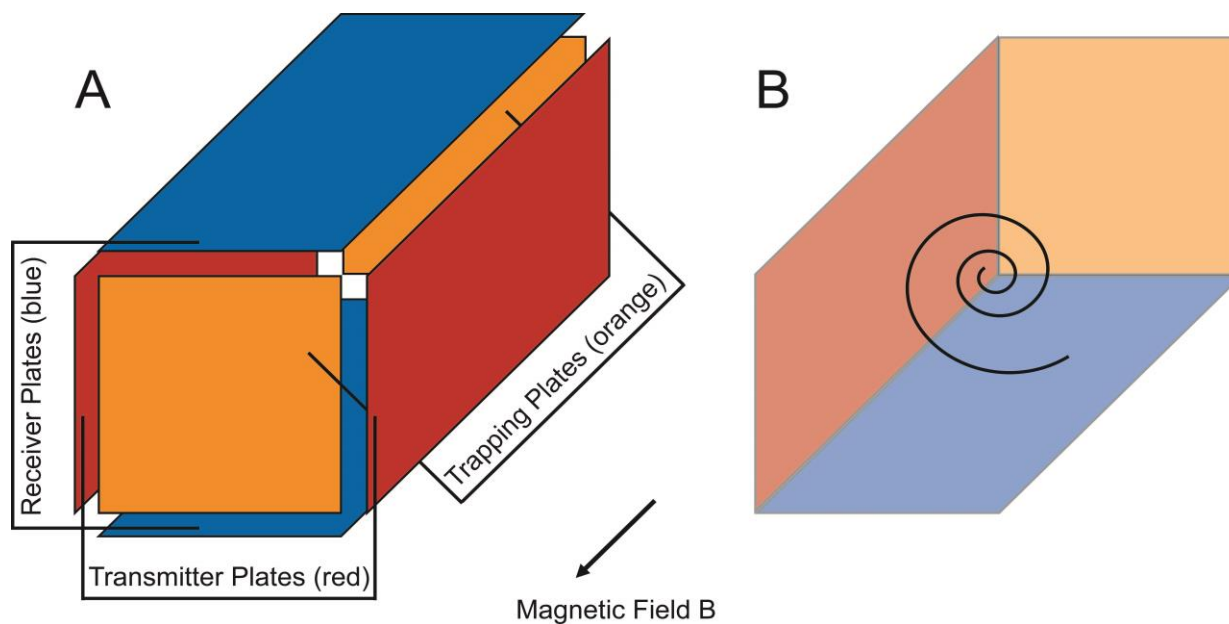
Figure 8: The Penning trap. A) It consists of trapping, transmitting, and receiving plates with a magnetic field present. B) Three plates have been removed to show the ion motion shortly after the excitation chirp and the characteristic relaxation back toward the center of the Penning trap.

The data gathered from the receiver plate represents the time-domain signal from the cyclotron frequencies of all ions in the Penning trap. In order to have any useful meaning, the data must be put through a Fourier transform algorithm. A Fourier transform effectively separates the very complex time-domain signal (**Figure 9**A) into the mass spectrum that is commonly viewed (**Figure 9**B). It does so by teasing out the various cyclotron frequencies and associating an *m/z* with a certain cyclotron frequency. Since the amount of image current on the plates is proportional to the number of ions that produced it, the intensities of the peaks may also be generated.
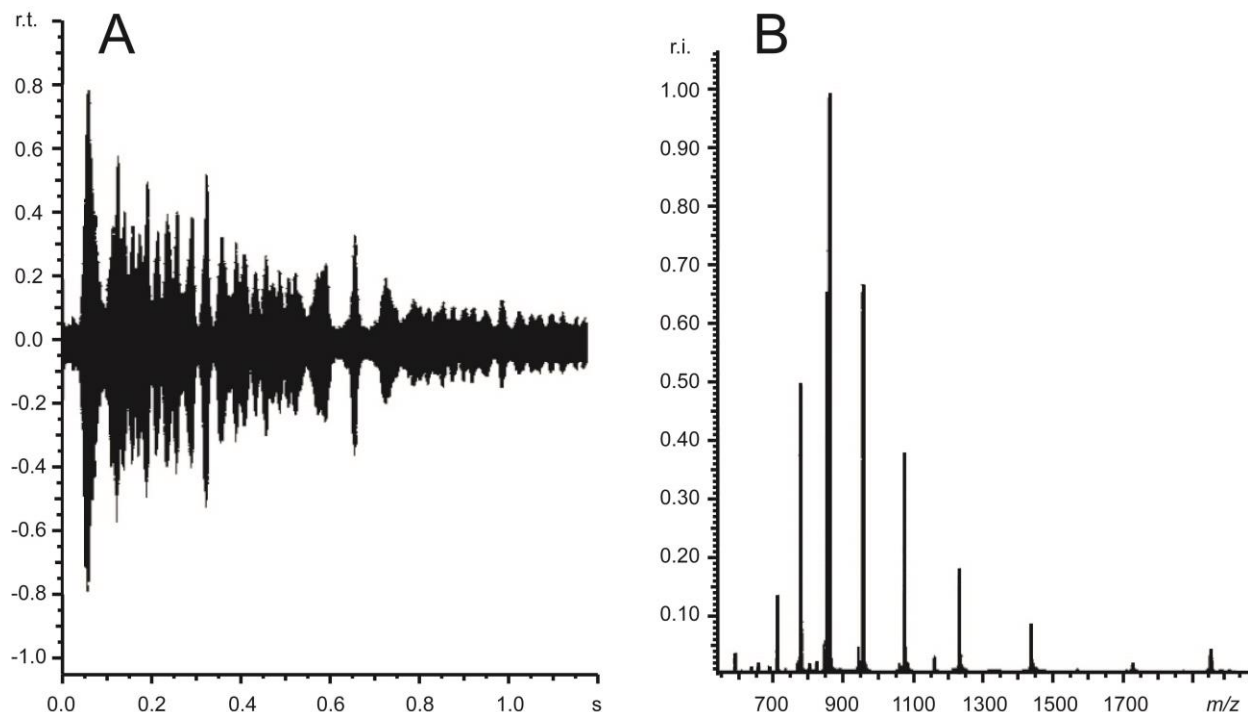
Figure 9: Fourier transform of the transient signal (A) produces the corresponding mass spectrum (B). (Reproduced from Amster[28])

1.3.2.3 Tandem mass spectrometry

The ability to accurately measure a mass to charge ratio (and deduce the mass from knowledge of the charge state) is very useful and has led to the application of mass spectrometry as an analytical technique in almost every area of scientific research. However, as the mass of an ion increases, so do the number of possible chemical formulae that mass can represent. This represents a formidable problem when attempting to either verifiably identify a certain analyte or differentiate isomers. To solve this problem a precursor ion can be isolated and subjected to tandem mass spectrometry (MS/MS) where a type of fragmentation or dissociation is used in order to obtain more structural information. The basic steps for tandem mass spectrometry

include isolation of the precursor ion, fragmentation, and measurement of the

corresponding fragments' *m/z* values (**Figure 10**).
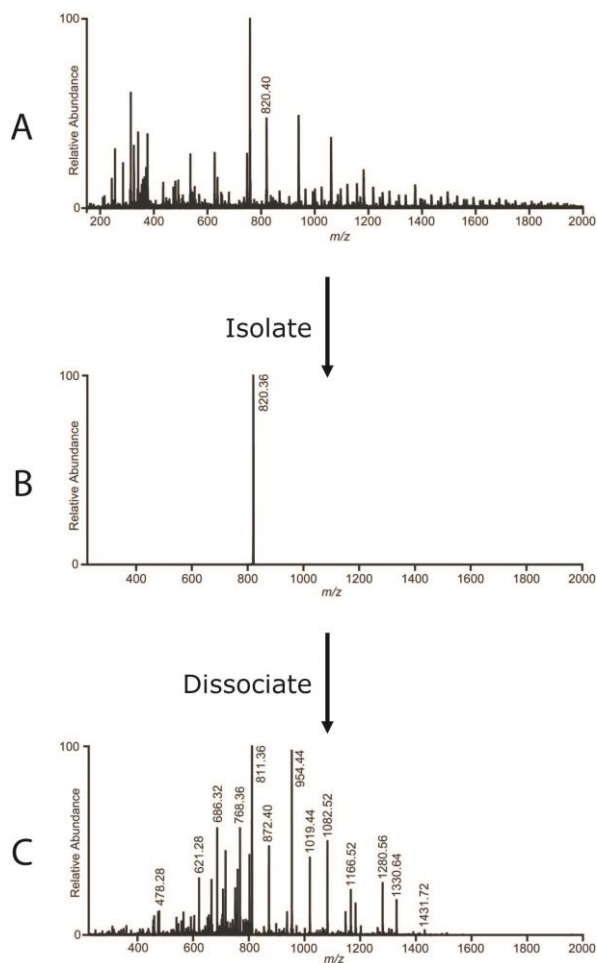
Figure 10: Tandem mass spectrometry can easily be performed on ion traps by (A) choosing an ion from the full scan spectrum, (B) isolating the ion by its *m/z*, and (C) performing dissociation followed by measuring the fragments *m/z* values.

1.3.2.4 Ion isolation

The capability to selectively eject ions from an ion trap becomes very important

when tandem mass spectrometry is wanted or required. Selective isolation of ions for

tandem mass spectrometry can be accomplished in a couple of different ways. From

viewing the stability diagram in **Figure 6**, it can be surmised that a sufficiently narrow packet of ions can be retained in an ion trap by setting that *m/z* to a *q* value of about 0.7 and increasing the DC potential to make the *a* value about 0.24 (the apex of the stability diagram). This leaves only this packet of ions with stable trajectories. Resonance ejection eliminated the need for the DC potential because unstable "holes" could be introduced at particular *q* values, and manipulating the RF potential allowed preferential ejection of unwanted ions (**Figure 11**).
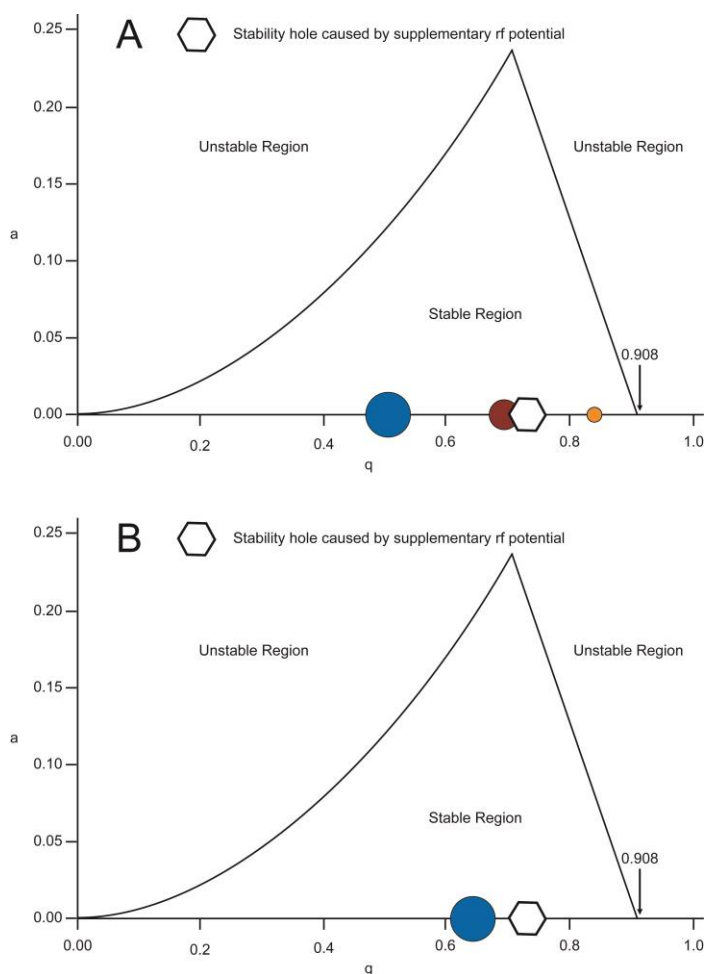
Figure 11: Using resonant excitation for ion isolation. A) The larges ion is selected and a supplementary RF potential is applied to excite ions with a certain q value. B) The VRF is ramped and the smallest ion is ejected after reaching the 0.908 stability boundary while the middle ion was ejected due to resonant excitation.

1.3.2.5 Fragmentation

The two most common types of fragmentation on peptide and protein analytes are collision induced dissociation and electron transfer dissociation.

Collision induced dissociation (CID) represents one of the most widely used fragmentation techniques in peptide analyses.[15,29-32] CID can easily be performed in several instruments, including linear and quadrupole ion traps, by manipulating potentials applied to different parts of the trap. First the precursor ion must be isolated by applying a wide band RF potential that makes all other *m/z* values unstable and are consequently ejected from the trap. Once isolated, increased RF and DC potentials increase the energy of the precursor ions to the point they collide with some of the inert gas atoms, which are purposely included as a bath gas in the trap. Collisions with the inert gas convert the kinetic energy of the precursor ions into internal energy such that they fragment. The resulting product ions from peptides are typically b and y type ions (Figures 12 and 13) but a small amount of a and x type ions are also produced. Fragments can be scanned with the standard mass selective instability mode of the ion trap.
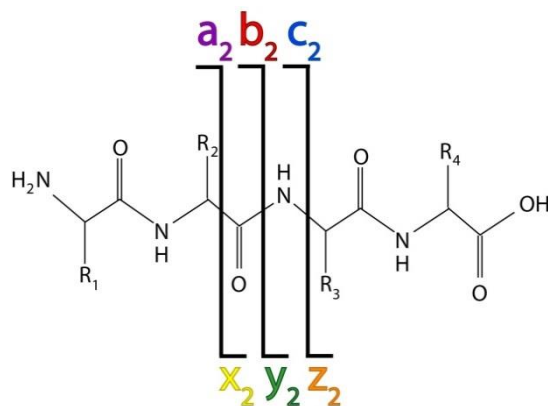
Figure 12: Peptide backbone fragmentation nomenclature put forth by Roepstorff in 1984 is still commonly used today. Nomenclature for peptide fragmentation indicates both type of ion and number of residues represented by letter and number respectively.

CID can be used to solve to a variety of problems, as it can fragment both small molecules and larger peptides, providing a wealth of information about the structure of the analyte. CID is routinely applied to analyze post-translational modifications (PTMs) such as phosphorylation[29,33,34], glycosylation[35-37], and disulfide bonds.[15,16,38] Both phosphorylation and some types of glycosylation are considered labile PTMs, meaning a substantial proportion of the product ions are neutral losses of the modification when CID is performed on the precursor ion. In these cases CID, provides very little attachment information about the location of the PTM on the peptide. Disulfide bonds however, are not labile and are not typically broken during CID. Application of CID to disulfide-bonded peptides provides amino acid sequence information about the two or more peptides, but analyzing these data is more complicated because at least two peptides are undergoing fragmentation, which results in many possible product ions.
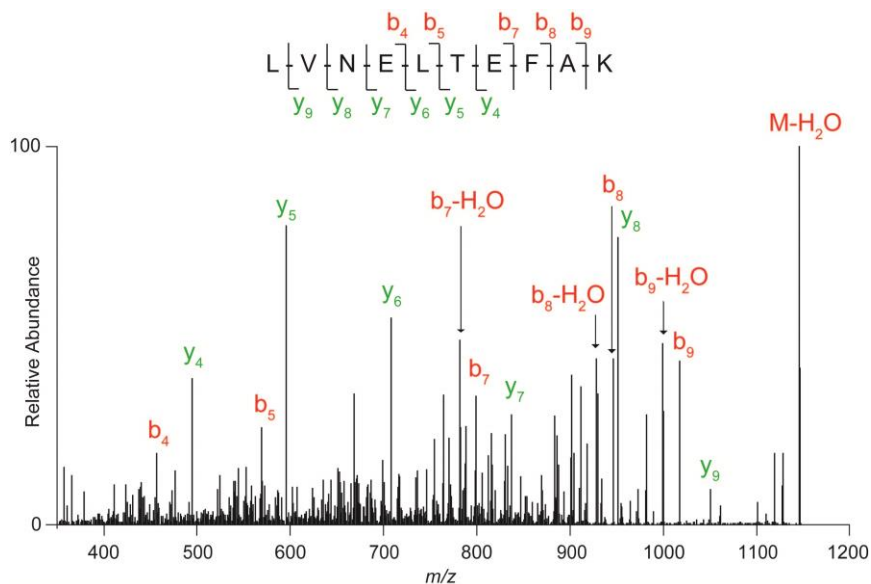
Figure 13: A representative spectrum of a peptide from bovine serum albumin fragmented with CID. Fragments are dominated by b and y type ions.

Electron transfer dissociation (ETD) is one of the newest dissociation techniques that is commercially available. Modeled after electron capture dissociation (ECD), ETD was designed to bring electron dissociation to those who could not afford the FT-ICR-MS instrument, which is required for ECD. Instead of directly injecting electrons into the mass analyzer to react with precursor ions as in ECD, ETD uses a reagent molecule that is injected into the trapping region. Several suitable reagent molecules have been noted and Thermo Scientific currently includes fluoranthene as the reagent molecule on their instruments. The internal fluoranthene vial is heated to evaporate a small amount which is transferred via carrier gas to the ion volume. In the ion volume the reagent gas (which is typically nitrogen) is bombarded by electrons from a filament. Electrons are ejected from nitrogen molecules and accepted by fluoranthene molecules, thereby ionizing them. From the ion volume, the fluoranthene (now negatively charged from the

extra electron and called the reagent ion) is transferred via ion optics to the ion trap where it is allowed to react with precursor ions. The resulting product ions from peptides and proteins are typically c and z type ions (**Figure 14**).[39-43]
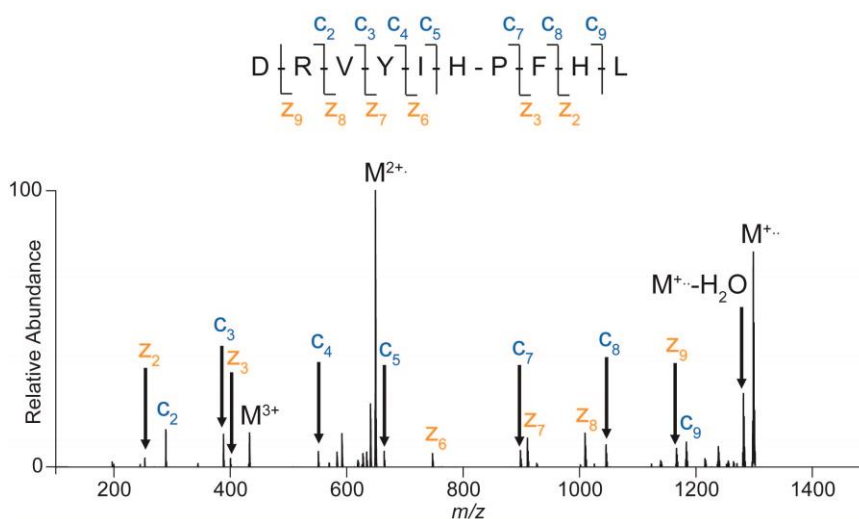
Figure 14: A representative spectrum of a peptide fragmented with ETD. Fragments are dominated by c and z type ions.

Quite contrary to CID, ETD does not affect the labile PTMs of phosphorylation and glycosylation[33-36,40,42,44-48]. Instead, the radical electron attacks the N-C_α bond of the peptide backbone to yield c and z ions that can be used to sequence the peptide.[49-51] Several recent studies have used ETD as a utility for determining site of phosphorylation or glycosylation because the product ions generated indicate a mass shift at the amino acid that is modified[45,51-55]. However, when disulfide bonded peptides are subjected to ETD a very different result occurs: Disulfide bonded peptides preferentially homolytically cleave the disulfide bond.[49] Systematic experimental studies as well as molecular dynamic simulations have attributed this phenomenon to the

electronegativity of the two sulfur atoms. The transferred electron is able to occupy the antibonding σ* orbital between the two sulfur atoms.[56] The antibonding nature of this orbital causes an increased propensity for breaking this bond. The corresponding effect is that very intense product ions at the *m/z* values expected for the linear peptides are detected (See **Figure 15**). ETD can serve as a complementary dissociation technique to CID because both the mechanism of action and resulting fragments are very different. As a result, many researchers apply the two dissociation techniques in an orthogonal manner to verify assignments made with one or the other because the structure will be the ultimate predictor of what fragments will result.[32,57,58]
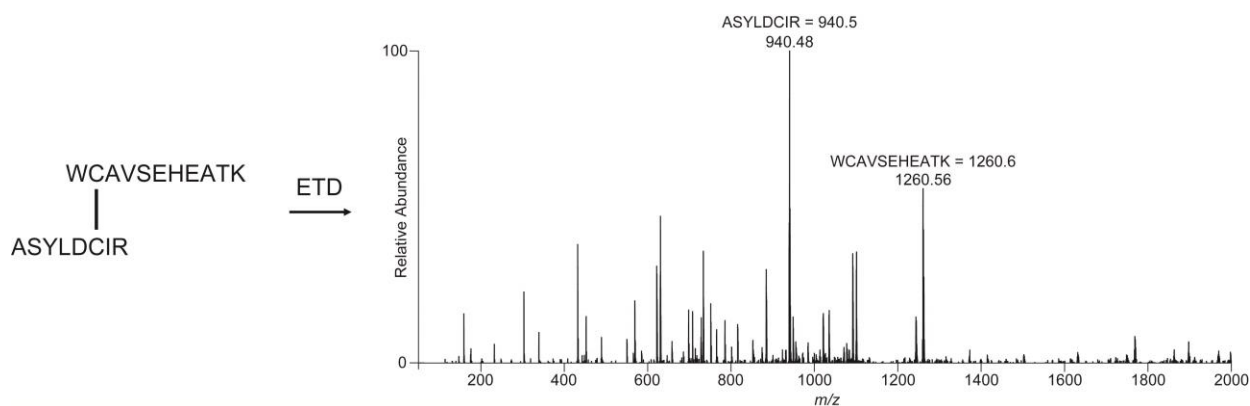
Figure 15: A representative spectrum of a disulfide-bonded peptide from transferrin fragmented with ETD. The spectrum is dominated by the homolytic scission of the disulfide bond to yield fragment ions consisting of the two intact peptides.

1.4 Summary of Subsequent Chapters:

Chapter 2 explores the fragmentation products of tryptic disulfide bonded peptides when subjected to low energy collision induced dissociation. These studies are important because computer programs being used for automated assignment of disulfide bonded peptides only consider only single peptide backbone cleavage events in the analysis logic. However, these single cleavage ions typically assign a mere fraction of the product ions that are formed from disulfide bonded peptides. The possibility of double cleavage fragmentation events was considered and investigated with several model disulfide bonded peptides from lysozyme, albumin, and fetuin. Utilizing double cleavage product ions yielded a drastic increase in the number of peaks identified in the CID spectra of almost all disulfide bonded peptides analyzed. In one example, double cleavage product ions also yielded more information about the bonding locations of a peptide with three cysteine residues.

Chapter 3 describes a method of assigning disulfide bonded pairs of peptides using electron transfer dissociation, and these studies have particular relevance to the biopharmaceutical industry. The number of biotherapeutic proteins used in medicine increases each year; they have shown a nearly exponential growth in the last decade. Often these biotherapeutics must remain in native form with correct disulfide bonds in place to be both safe and efficacious. A method is proposed that uses the unique homolytic cleavage of the disulfide that occurs due to a transferred radical electron, yielding the masses of the two or more peptides that were once bonded. Searching through the tandem mass spectra allows identification of peaks where two or more

expected tryptic peptide masses appear and can be assigned as a disulfide bonding pair. This method was set up to assign both free cysteine residues and disulfide bonded peptides and was applied to a recombinant HIV envelope gp120 protein. The method was able to detect an aberrant disulfide bonding pattern in one region of the protein, and it accurately identified a tetra disulfide bonded peptide (four peptide chains).

Chapter 4 introduces an experimental approach for identifying any potential disulfide shuffling that may occur during sample preparation. This work is significant because it provides validation that the disulfide bonding pattern elucidated by mass spectrometry is equivalent to the disulfide bonding pattern in the original protein. This project was initiated after other researchers in the Desaire research group had recently identified highly unexpected, heterogeneous disulfide bonding patterns in an important potential HIV vaccine. To demonstrate that these unexpected disulfides were present in the protein as it was received, and not introduced during sample preparation, several experiments with different buffer conditions, enzymatic conditions, and variant sequences had been conducted, and they all yielded similar results of multiple disulfide bonding configurations for recombinant monomers of gp120. Yet, because the outcomes of the studies were surprising to researchers, we sought an orthogonal approach for verifying that sample preparation did not contribute to the unusual disulfide bonding pattern. In order to eliminate any possibility that sample preparation was a source of detectable disulfide shuffling, a method of introducing a small reporter disulfide bonded peptide at the beginning of the experiment is outlined. In theory, if

disulfide shuffling occurred during sample preparation, then the protein of interest would also shuffle with the reporter peptides and produce a characteristic mass spectral peak. This method was used in conjunction with model proteins lysozyme and transferrin to show lack of shuffling during a standard enzymatic deglycosylation and tryptic digest. However, when non-optimal conditions were used during sample preparation shuffling between the target protein and reporter peptides was detected. The method was applied to the gp120 protein of interest, and this work verified that sample preparation was not modifying the disulfide bonding pattern. This method could detect the presence of disulfide shuffling at many steps along the protein production and formulation process for biotherapeutics.

Chapter 5 outlines a potential future direction of disulfide mapping studies performed herein. A computer program, Disulfide Hunter, could provide automated disulfide assignments when used in conjunction with the ETD method from Chapter 3. For automation to occur, the analysis must be reduced to essentially mathematical calculations so the software can make assignments. Disulfide Hunter is arranged to eliminate the human contribution to data analysis, which also eliminates subtle differences that may occur between two researchers. It would include an interface for the user to input enough information about the experiment that the computer could take into account many modifications, especially those that may affect the mass of the peptides. Disulfide Hunter would output a list of disulfide bonded peptides including retention time and other information from the spectra assigned as such. The user could then easily quality-check the assignments.

1.5 References:

(1) Anfinsen, C. B.; Haber, E. *J. Biol. Chem.* **1961**, *236*, 1361-1363.

(2) Anfinsen, C. B. *Science* **1973**, *181*, 223-230.

(3) Creighton, T. E. *Methods Enzymol.* **1984**, *107*, 305-329.

(4) Jones, D. T.; Taylor, W. R.; Thornton, J. M. *Nature* **1992**, *358*, 86-89.

(5) *Journal of Cellular Biochemistry* **1991**, *47*, 157-221.

(6) Trivedi, M. V.; Laurence, J. S.; Siahaan, T. J. *Curr. Protein Pept. Sci.* **2009**, *10*, 614-625.

(7) Leonard, C. K.; Spellman, M. W.; Riddle, L.; Harris, R. J.; Thomas, J. N.; Gregory, T. J. *J. Biol. Chem.* **1990**, *265*, 10373-10382.

(8) Kóródi, I.; Asbóth, B.; Polgár, L. *Biochemistry* **1986**, *25*, 6895-6900.

(9) Haniu, M.; Arakawa, T.; Bures, E. J.; Young, Y.; Hui, J. O.; Rohde, M. F.; Welcher, A. A.; Horan, T. *J. Biol. Chem.* **1998**, *273*, 28691-28699.

(10) Norris, G. E.; Stillman, T. J.; Anderson, B. F.; Baker, E. N. *Structure (London, England : 1993)* **1994**, *2*, 1049-1059.

(11) Maley, F.; Trimble, R. B.; Tarentino, A. L.; Plummer, T. H. *Anal. Biochem.* **1989**, *180*, 195-204.

(12) Plummer, T. H. J.; Tarentino, A. L. *Glycobiology* **1991**, *1*, 257-264.

(13) Givol, D.; De Lorenzo, F.; Goldberger, R. F.; Anfinsen, C. B. *Proc. Natl. Acad. Sci. U. S. A.* **1965**, *53*, 676-684.

(14) Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 327-332.

(15) Tsarbopoulos, A.; Varnerin, J.; Cannon-Carlson, S.; Wylie, D.; Pramanik, B.; Tang, J.; Nagabhushan, T. L. *J. Mass Spectrom.* **2000**, *35*, 446-453.

(16) Yen, T. Y.; Joshi, R. K.; Yan, H.; Seto, N. O. L.; Palcic, M. M.; Macher, B. A. *J. Mass Spectrom.* **2000**, *35*, 990-1002.

(17) Rodbard, D.; Levitov, C.; Chrambac.A *Separation Science* **1972**, *7*, 705-723.

(18) Sommer, A.; Traut, R. R. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 3946-3950.

(19) Wittmann-Liebold, B.; Graack, H. R.; Pohl, T. *Proteomics* **2006**, *6*, 4688-4703.

(20) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. *Science* **1989**, *246*, 64-71.

(21) March, R. E. *Int. J. Mass Spectrom. Ion Process.* **1992**, *118*, 71-135.

(22) Schlunegger, U. P.; Stoeckli, M.; Caprioli, R. M. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1792-1796.

(23) Jonscher, K. R.; Yates, J. R. *Anal. Biochem.* **1997**, *244*, 1-15.

(24) March, R. E. *J. Mass Spectrom.* **1997**, *32*, 351-369.

(25) Stafford, G. C.; Kelley, P. E.; Syka, J. E. P.; Reynolds, W. E.; Todd, J. F. J. *Int. J. Mass Spectrom. Ion Process.* **1984**, *60*, 85-98.

(26) Paul, W.; Reinhard, H. P.; Zahn, O. *Z. Phys.* **1958**, *152*, 143-182.

(27) Cooks, R. G.; Glish, G. L.; McLuckey, S. A.; Kaiser, R. E. *Chem. Eng. News* **1991**, *69*, 26-&.

(28) Amster, I. J. *Journal of Mass Spectrometry* **1996**, *31*, 1325-1337.

(29) Jonsson, A. P. *Cellular and Molecular Life Sciences* **2001**, *58*, 868-884.

(30) Hogan, J. M.; McLuckey, S. A. *J. Mass Spectrom.* **2003**, *38*, 245-256.

(31) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255-261.

(32) Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. In *Annual Review of Biomedical Engineering* 2009; Vol. 11, p 49-79.

(33) McLachlin, D. T.; Chait, B. T. *Current Opinion in Chemical Biology* **2001**, *5*, 591-602.

(34) Mann, M.; Ong, S. E.; Gronborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. *Trends in Biotechnology* **2002**, *20*, 261-268.

(35) Mann, M.; Jensen, O. N. *Nature Biotechnology* **2003**, *21*, 255-261.

(36) Zaia, J. *Mass Spectrometry Reviews* **2004**, *23*, 161-227.

(37) Hogan, J. M.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *Journal of Proteome Research* **2005**, *4*, 628-632.

(38) Angata, K.; Yen, T. Y.; El-Battari, A.; Macher, B. A.; Fukuda, M. *J. Biol. Chem.* **2001**, *276*, 15369-15377.

(39) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *Journal of the American Chemical Society* **1998**, *120*, 3265-3266.

(40) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. *Analytical Chemistry* **1999**, *71*, 4431-4436.

(41) Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. *Journal of the American Chemical Society* **1999**, *121*, 2857-2862.

(42) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 1793-1800.

(43) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Analytical Chemistry* **2000**, *72*, 563-573.

(44) Zubarev, R. A. *Mass Spectrometry Reviews* **2003**, *22*, 57-77.

(45) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. *Molecular & Cellular Proteomics* **2005**, *4*, 310-327.

(46) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 880-882.

(47) Cooper, H. J.; Hakansson, K.; Marshall, A. G. *Mass Spectrometry Reviews* **2005**, *24*, 201-222.

(48) Catalina, M. I.; Koeleman, C. A. M.; Deelder, A. M.; Wuhrer, M. *Rapid Communications in Mass Spectrometry* **2007**, *21*, 1053-1061.

(49) Chrisman, P. A.; Pitteri, S. J.; Hogan, J. M.; McLuckey, S. A. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1020-1030.

(50) Syrstad, E. A.; Turecek, F. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 208-224.

(51) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2006**, *1764*, 1811-1822.

(52) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. *Molecular & Cellular Proteomics* **2007**, *6*, 1942-1951.

(53) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. *Analytical Chemistry* **2007**, *79*, 9243-9252.

(54) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 2199-2204.

(55) Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G. *Nature Methods* **2007**, *4*, 798-806.

(56) Simons, J. *Chemical Physics Letters* **2010**, *484*, 81-95.

(57) Molina, H.; Matthiesen, R.; Kandasamy, K.; Pandey, A. *Analytical Chemistry* **2008**, *80*, 4825-4835.

(58) Alley, W. R.; Mechref, Y.; Novotny, M. V. *Rapid Communications in Mass Spectrometry* **2009**, *23*, 161-170.

Chapter 2: Collision Induced Dissociation Products of Disulfide-bonded Peptides:

Ions Result from the Cleavage of More than One Bond

This work has been published by the Journal of the American Society for Mass Spectrometry and is reproduced with permission from the journal.

Disulfide bonds are a posttranslational modification (PTM) that can be scrambled or shuffled to non-native bonds during recombinant expression, sample handling, or sample purification. Currently, mapping of disulfide bonds is difficult due, to various sample requirements and data analysis difficulties. One step towards facilitating this difficult work is developing a better understanding of how disulfide-bonded peptides fragment during Collision Induced Dissociation (CID). Most automated analysis algorithms function based on the assumption that the preponderance of product ions observed during the dissociation of disulfide-bonded peptides result from the cleavage of just one peptide bond, and in this report we tested that assumption by extensively analyzing the product ions generated when several disulfide-bonded peptides are subjected to CID on a QTOF instrument. We found that one of the most common types of product ions generated resulted from two peptide bond cleavages, or a double cleavage. We found that for several of the disulfide-bonded peptides analyzed, the number of double cleavage product ions outnumbered those of single cleavages. The influence of charge state and precursor ion size was investigated, to determine if those parameters dictated the amount of double cleavage product ions formed. It was found in this sample set that no strong correlation existed between the charge state or peptide size and the portion of product ions assigned as double cleavages. This data shows

that these ions could account for many of the product ions detected in CID data of disulfide bonded peptides. We also showed the utility of double cleavage product ions on a peptide with multiple cysteines present. Double cleavage products were able to fully characterize the bonding pattern of each cysteine where typical single *b*/*y* cleavage products could not.

2.1 Introduction:

Verification of protein three-dimensional structure is often complicated, yet necessary for both research purposes and industrial production. Disulfide bonding is one very important protein feature that can affect protein structure and functionality.[1-6] The ability of two amino acid residues in close proximity to undergo an oxidation to form a bond is unique to cysteine and has multiple spatial requirements.[7] These intra- and possibly inter-protein bridges play a large part in stabilizing the overall structure of a protein.[1,3-6] Anfinsen et al. showed that in the absence of native disulfide bonds, some proteins cannot achieve the correct three-dimensional structure to perform their proper function.[1,5] The determination of the disulfide bonding network of a protein is therefore one important tool for characterizing isolated and recombinantly-expressed proteins, and techniques to facilitate disulfide bonding analysis are presented herein.

There are several ways to go about determining the disulfide bonding network of a protein with mass spectrometry. Many groups have developed methods to break the disulfide bond, generating characteristic mass shifts or a reporter ion to monitor.[8-14] However, this approach is complicated because the sample preparation is somewhat complex, often requiring the addition of reagents that will cause either a characteristic mass shift (such as isotope labeling) or scission of the disulfide bond.[10,15-22] Electron capture dissociation and electron transfer dissociation have also shown promise to potentially map disulfide bond networks through a top down approach[23,24] as well as a bottom up approach.[25] A simpler work flow involves subjecting tryptic peptides to CID and using *de novo* sequencing to assign a peptide sequence.[26] However, the key limitation of this approach with disulfide-bonded peptides is data analysis. Identification

of MS[n] data for disulfide-bonded peptides is tedious, time-consuming, and difficult-especially when more than two peptides are involved. As a result, computer algorithms are relied upon heavily to aid in the analysis.

Currently, there are several programs that assign MS[n] data for disulfide-bonded peptides.[27-31] Four of the most notable programs include: xComb[28], MS2Assign[29], SearchXLinks[30], and MassMatrix.[31] All of these programs attempt to match product ions in the MS/MS data to a theoretical spectrum of product ions from candidate disulfide-bonded peptide sequences. Each program has its merits and limitations; however, they each share a potential weakness: in all the programs, the assumption is made that the most common fragmentation observed during CID will be the cleavage of one bond (typically a peptide bond). This assumption is likely based on the early observations from Schilling et al, who reported that the abundance of double cleavage ions was less than 10% with a QTOF instrument.[29] (The terminology "double cleavage" is adopted from references 28 and 30.) Studies using electrospray ionization (ESI) have found that the number and type of cleavages can be dependent on charge.[24,25,32-34] If, in fact, disulfide bonded peptide ions, produced during ESI, typically undergo multiple bond cleavages under CID conditions, software that attempts to assign disulfide bonded peptides could be optimized by incorporating this information.

The present investigation was carried out to provide additional insight into the fragmentation trends of disulfide linked peptides, so that this fundamental knowledge can provide guidance to developers of disulfide mapping software. The key question asked here is: When can one assume that disulfide linked peptides predominantly undergo just a single peptide cleavage, after ionization by ESI-MS and CID? To answer

this question, a variety of disulfide-bonded peptides are subjected to CID in an ESI-QTOF mass analyzer, and a diligent attempt is made to identify each of the product ions, by considering many possible fragmentation pathways that the peptides could undergo. The product ions are assigned manually, using a rigorous set of rules for verifying the assignment. These studies ultimately indicate that cleavage of two peptide bonds is an important and common fragmentation pathway for disulfide-bonded peptides undergoing CID.

2.2 Experimental:

2.2.1 Materials and Reagents

Lysozyme from chicken egg white, acetic acid, and formic acid were purchased from Sigma-Aldrich (St. Louis, MO). Ammonium bicarbonate was purchased from Fluka (Milwaukee, WI). Ammonium carbonate and HPLC grade acetonitrile were purchased from Fisher (Pittsburgh, PA). Sequencing-grade modified trypsin was purchased from Promega (Madison, WI). The 4-vinylpyridine used was purchased from Acros Organics (Morris Plains, NJ). Water was purified by a Millipore Direct-Q3 Water Purification System (Billerica, MA).

2.2.2 Sample Preparation and Proteolysis

Samples of ~300 µg of protein were dissolved in two different ammonium carbonate 15 mM buffers, one at pH 6.0 (H_2CO_3/NH_4HCO_3) and the other at pH 9.0 ($NH_4HCO_3/(NH_4)_2CO_3$) to approximately 3.0 µg/µL. Both were incubated at 37˚ C for one hour. Free cysteines were alkylated by the subsequent addition of 4-vinylpyridine

and allowed to incubate at room temperature for one hour in the dark. Two trypsin additions were made, each being 1:15 enzyme:substrate, eight hours apart. After 24 hours from the first trypsin addition, 1 µL concentrated acetic acid per 20 µL of solution was added. Samples were analyzed immediately following acetic acid addition.

2.2.3 QTOF LC-MS

Liquid chromatography was performed on a Waters Acquity UPLC using a Micro-Tech Scientific capillary column (500 µm i.d., 10 cm, 3 µm, 300 Å) (Vista CA) equipped with a 5 µL injection loop. A flow rate of 5 µL/min and a linear gradient beginning at 98%:2% A:B with increasing organic concentration over 50 minutes was used. Solvent A consisted of 99.9%/0.08% H_2O/formic acid and solvent B consisted of 99.9%/0.08% acetonitrile/formic acid. Positive ion mode was used, and the ESI needle was held at 3.5 kV. MassLynx software was set up to perform MS^2 on selected m/z ratios with a Q-TOF-2 (Micromass Ltd, Manchester UK). The collision energy varied, depending on the m/z of the precursor ion. Between m/z 350 and 600, 25% normalized collision energy was used; between m/z 601 and 900, 30% was used; and between m/z 901 and 2000, 35% normalized collision energy was used. In the case of peptides D, E, and F these collision energies did not adequately dissociate the precursor ion, so an additional 5% normalized collision energy was added, which yielded sufficient fragmentation.

2.2.4 Data Analysis

A "look-up" table of possible fragment ions was prepared for each peptide. (All look-up tables used are provided in Supplementary Material). The table contained the

m/z's and compositions of all possible *b* and *y* ions, ions resulting from disulfide bond scission, and ions resulting from the combination of any two single cleavages (including 2 *b* ions, 2 *y* ions, all possible combinations of 2 *b*/*y* ions, ions containing a single *b* or *y* cleavage and a disulfide scission, and any of the above single cleavage ions accompanied by loss of water or loss of NH_3). Only monoisotopic masses were used in the table and in peak assignment. For each MS^2 spectrum of disulfide bonded peptides, the 25 to 40 monoisotopic peaks of highest relative abundance were assigned by matching the monoisotopic mass to the masses in the custom-prepared look-up table. If the monoisotopic peak could not be clearly identified, the peak was discarded. If the ions did not match any mass in the look-up table resulting from combinations of *b* or *y* ion cleavage and/or disulfide cleavage, then neutral losses of water or NH_3 in combination with *b* or *y* ions were considered as a possible fragmentation pathway. (Assignment of any neutral losses required the necessary residues or termini capable of producing those neutral losses present in that peptide fragment.) Finally, for ions that remained unassigned, *a* and *x* ions were also considered as potential product ions. For MS^2 data, the assigned composition for each product ion was further verified by assuring that the charge state for the experimental data matched the charge state for the ion's assigned composition. Peaks were required to have no more than a ±0.15 Th mass error to be verified as "correctly assigned". Peaks in spectra are labeled either as "1", "2", "U", "M-H_2O/ M-NH_3" or "A". These designations refer to single cleavage (1), double cleavage (2), unknown (U), precursor ion – water or ammonia (M-H_2O/ M-NH_3), or ambiguous (A), respectively. An ambiguous ion was an ion whose *m/z* (and charge

state) was within the mass tolerance of both a possible single and double cleavage product ion on the look-up table for that particular peptide.

2.3 Results and Discussion:

To develop a better fundamental understanding of fragmentation of multiply-charged disulfide-linked peptides undergoing CID, a variety of analytes were generated by tryptic digestion of the protein lysozyme. The disulfide-linked peptides investigated in this study are shown in **Table 1**. Of these, A, B, and D are native disulfide linkages in lysozyme, and C was generated through intentional disulfide scrambling, by adjusting the pH of the solution to 9.0, prior to digestion. In addition, the fragmentation trends of two of these peptides (C and D) were assessed in various charge states. Using this data set, the dissociation characteristics of the peptides are described below, and the effects of peptide size and peptide charge state on the propensity of the peptide ion to undergo multiple b/y cleavages are determined.

Table 1: Disulfide-bonded peptides

Peptide Label	Protein and Uniprot Accession	Peptide Disulfide Bond Map
A	Chicken egg white lysozyme P00698	C-E-L-A-A-A-M-K G-C-R
B	Chicken egg white lysozyme P00698	G-Y-S-L-G-N-W-V-C-A-A-K C-K
C	Chicken egg white lysozyme (non-native) P00698	G-Y-S-L-G-N-W-V-C-A-A-K W-W-C-N-D-G-R
D	Chicken egg white lysozyme P00698	N-L-C-N-I-P-C-S-A-L-L-S-S-D-I-T-A-S-V-N-C-A-K W-W-C-N-D-G-R
E	Bovine serum albumin P02769	T-C-V-A-D-E-S-H-A-G-C-E-K
F	Bovine fetuin P12763	A-Q-F-V-P-L-P-V-S-V-S-V-E-F-A-V-A-A-T-D-C-I-A-K C-N-L-L-A-E-K

2.3.1 Small peptides

Do the fragmentation characteristics of disulfide-bonded peptides depend on the overall size of the peptide? To investigate this question, the fragmentation characteristics of two small disulfide-bonded peptides (A and B in **Table 1**) containing one inter-peptide bond were investigated. Both contained a large alpha peptide and a much smaller beta peptide. Both species were subjected to CID in a QTOF mass spectrometer in the 2+ charge state (**Figure 1**). After assigning all of the most abundant product ions for these two species, by matching *m/z* and charge state of the observed product ions to an exhaustive list of theoretical product ions (as described in the experimental section) it was apparent that the size of the peptide did not dictate the amount of single cleavage versus double cleavage ions observed (**Table 2**). Peptide A produced all single cleavage product ions while peptide B produced a mix of both single cleavage and double cleavage product ions. Approximately one third of the ions assigned for peptide B were the result of double peptide backbone cleavages. These data demonstrate that, even though the relative size of alpha and beta peptides can be very similar, the propensity to form double cleavages is very different. **Table 2** shows that if double cleavages were not included, only about half the peaks that could have been assigned in peptide B would have been assigned.
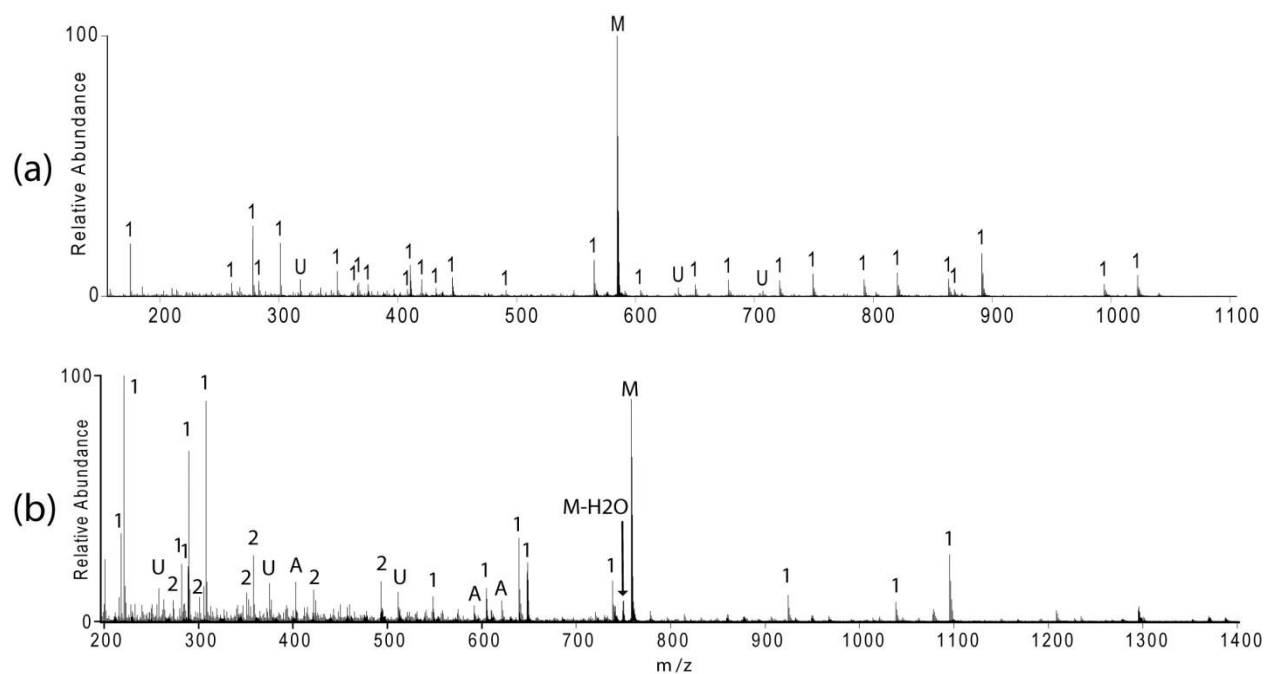
Figure 1: QTOF MS^2 spectra for peptides A and B in the 2+ charge state, at *m/z* (**a**) 584 and *m/z* (**b**) 758. Each "1" refers to a single cleavage, "2" refers to a double cleavage, "U" refers to an unknown ion, and "A" refers to an ambiguous assignment. Some assignment labels have been omitted due to space constraints.

Table 2: Comparison of the product ions generated from disulfide-bonded peptides that vary in composition and charge state. The peptide compositions are listed in **Table 1**. Peptides E and F were used to validate results. All assignments are provided in Supplementary Material.

Peptide Label	Residues in Alpha	Residues in Beta	Mass	Charge State Analyzed	Single Cleavages	Double Cleavages	Only Neutral Loss	Unknown/ Ambiguous
A	8	3	1167.52	2+	28	0	0	2
B	12	2	1514.70	2+	14	11	1	13
C_1	12	7	2200.96	3+	24	6	1	5
C_2	12	7	2200.96	2+	12	8	0	7
D_1	23	7	3267.46	4+	6	30	0	4
D_2	23	7	3267.46	3+	6	14	2	4
E	13	N/A	1346.52	2+	3	22	2	5
F_1	24	7	3248.68	4+	17	9	0	4
F_2	24	7	3248.68	3+	9	14	0	6

2.3.2 Medium peptide

After investigating the fragmentation patterns of two small disulfide-bonded peptides, the next step in our comprehensive fragmentation analysis of these types of ions is to contrast the fragmentation pattern of small disulfide-bonded peptides to medium and large species. To generate a "medium" sized peptide, a non-native lysozyme disulfide-bonded peptide containing one inter-peptide bond was formed, through pH and temperature control. This peptide is labeled C shown in **Table 1**, and it exemplifies a peptide that contains a medium-sized peptide for both the alpha and beta chains. When fragmented in two different charge states (2+ and 3+), peptide C formed a mix of single and double cleavage product ions (**Table 2**). One fifth of the ions assigned in the higher charge state were double cleavages and approximately one half of the ions assigned in the lower charge state were double cleavages. Comparing peptides B and C, one can see that even though they have quite different compositions, concerning alpha and beta peptide relative size, they fragment in similar quantities of single and double cleavage product ions. These data support the conclusion that the mass of the alpha/beta chains cannot be used to predict whether or not the peptide will undergo multiple cleavages during CID.

2.3.3 Large peptide

The fragmentation of a "large peptide" was also investigated for these studies. Tryptic digestion of lysozyme yielded one disulfide-bonded peptide containing one intra-peptide bond as well as one inter-peptide bond. This peptide is labeled D, shown in

Table 1, and the MS/MS data for this peptide (in 2 different charge states) is shown in Figures 2a and 2b. When subjected to CID, the peptide formed numerous double cleavages and few single cleavage product ions (**Table 2**). This peptide had the largest proportion of peaks assigned as double cleavages of the four analyzed. The product ions for this species were assigned from precursors of two different charge states. Both charge states analyzed (3+ and 4+) showed a large proportion of ions resulting from cleavage of at least two bonds. However, many more high-abundance peaks were assigned as double cleavages in the higher charge state.

2.3.4 Effect of charge state

Peptides C and D gave a sufficient signal in the CID spectra to warrant analysis in multiple charge states, as summarized in **Table 2**. These two peptides provided conflicting data in the attempt to find a way to associate relative charge state with the amount of ions formed as a result of multiple bond cleavages. Peptide C was analyzed in both the 2+ and 3+ charge states. The lower charge state had several more peak assignments that consisted of double cleavage products. Peptide D was analyzed in both the 3+ and 4+ charge states. However, this example showed the higher charge state to have a much higher population of double cleavage product ions (**Figure 2**). While this data set is small, it is large enough to indicate that there is not a strong correlation between the charge state and the type of ions formed.
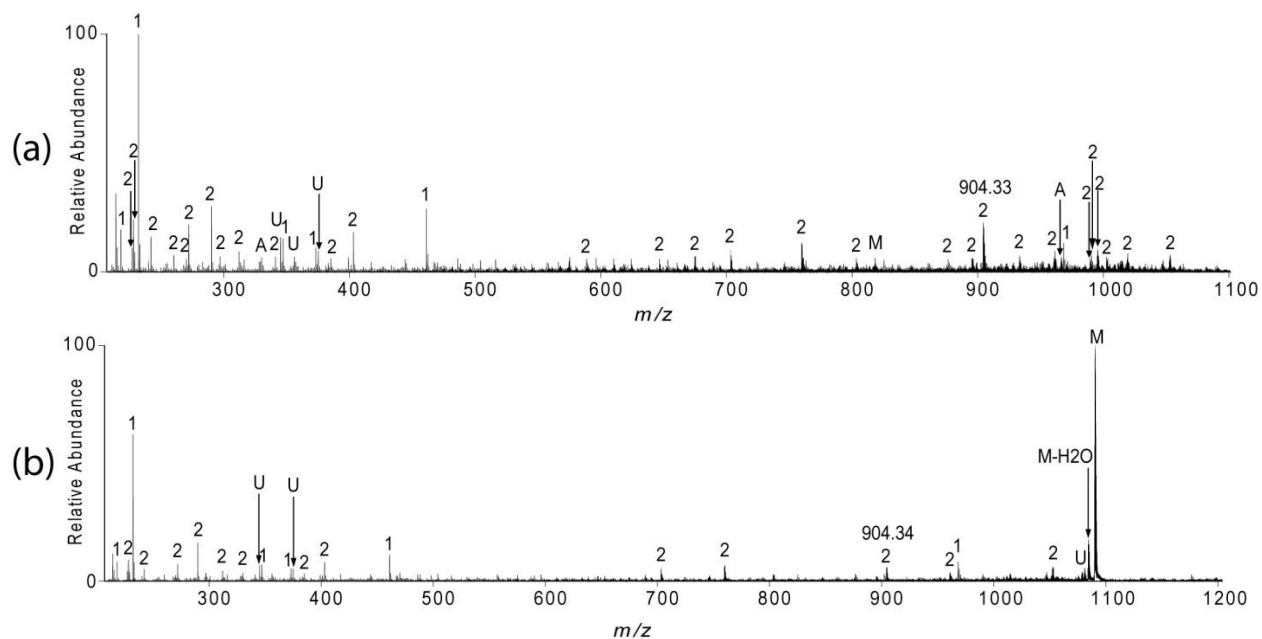
Figure 2: (**a**) QTOF MS^2 spectrum for peptide D in the 4+ charge state at *m/z* 817. Each "1" refers to a single cleavage and "2" refers to a double cleavage. (**b**) QTOF MS^2 spectrum for peptide D at *m/z* 1090, an ion in the 3+ charge state. Each "1" refers to a single cleavage, "2" refers to a double cleavage, "U" refers to an unknown ion, and "A" refers to an ambiguous assignment. Some assignment labels have been omitted due to space constraints.

2.3.5 Double cleavages are significant, but cannot be predicted

The presence of double cleavage product ions in low energy collision induced dissociation spectra has been noticed previously, but the general perception is that these ions are typically in relatively low abundance.[27,29] Our results show that the formation of double cleavage product ions in disulfide-bonded peptide CID spectra is significantly more abundant than previously reported. As can be seen with the peptides C and D, the majority of the high abundance peaks can consist of double peptide backbone cleavages. Likewise, extensive fragmentation was also observed for a small disulfide-bonded peptide, and the charge state of the ion did not appear to dictate the propensity for these ions to form. Since these ions cannot be suppressed or readily predicted to be present or absent, one should hence-forth consider the likely possibility

that a significant number of product ions could be from the cleavage of at least two bonds when an unknown sample is subjected to CID for the purpose of disulfide mapping. The addition of these possible fragmentation pathways in peak assignment algorithms will allow for a more thorough analysis of peaks and may generate more confident peptide assignments, however including these additional ions may not guarantee a lower false-positive discovery rate.

2.3.6 Validation

In an effort to validate the above findings, another data set that was 50% the size of the original set was analyzed. A peptide with an intrapeptide disulfide bond from bovine serum albumin was fragmented in one charge state (labeled peptide E) and a peptide with an interpeptide disulfide bond from bovine fetuin was fragmented in two different charge states (labeled peptide F). See Tables 1 and 2. The data (included in supplementary material and **Table 2**) show that, again, the number of double cleavage product ions can be as plentiful if not more so than single cleavage products, for various types of disulfide-linked peptides.

Additionally, a recent analysis of the disulfide linkage pattern of a recombinant HIV-Env protein used CID data to assign disulfide bonds [35], and the study further supports our findings that many double cleavage product ions are generated upon CID of disulfide linked peptides. A total of 12 different disulfide-linked peptides of a variety of sizes and disulfide-linkage types were characterized using CID data in the 2+, 3+, and 4+ charge states.   About 340 product ions were assigned in the MS/MS spectra. Of those product ions, about 48% (164 ions) were the result of cleavage of at least 2

peptide bonds. (See supplemental data in reference 35). This study further validates the findings described here, that double cleavage ions can be abundant, regardless of the peptide size or charge state.

2.3.7 Utility of Double Cleavages for Disulfide Bond Elucidation

Double cleavage product ions can provide valuable insight into the disulfide network, even when multiple cysteine residues reside on a single peptide. (As an example, see peptide D in **Table 1**.) The presence of more than two cysteines in a disulfide-linked peptide complicates disulfide analysis, because even if one knows which peptides are bonded to each other, the disulfide connectivity is impossible to infer. In peptide D in **Table 1**, three cysteines are present in one tryptic peptide, and this peptide is disulfide-bonded to a peptide containing one cysteine. This particular peptide presents an interesting case as single cleavage product ions will not provide significant information about the disulfide networking due to the intra-peptide bond. Double cleavage assignments were able to verify the bonding of each separate cysteine residue. The MS/MS data for this peptide is in Figures 2a and 2b, and these data show a double cleavage peak at 904.33 Th. (This ion appeared in both the 3+ and 4+ charge states.) The only possibility for this mass and charge state is that of a double cleavage ion involving only the center of the three cysteine residues and the other peptide it is disulfide-bonded to, as shown in **Figure 3**. Thus, using this ion allowed for the determination of the bonding structure of all the cysteines: Two must be intra-peptide bonded and the center is bonded to another peptide. If double cleavages had been

ignored, full characterization of the bonding structure of the peptide would have been impossible, without further studies.
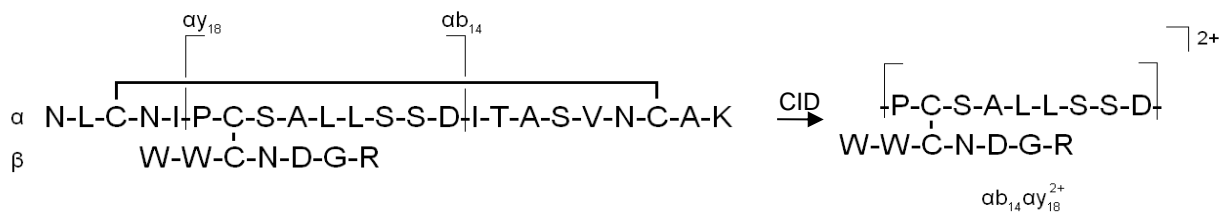

Figure 3: Inter-peptide disulfide bond can be assigned as the center cysteine residue due to this peak corresponding to a double cleavage product at *m/z* 904.33.

2.4 Conclusion:

Acquiring a better fundamental understanding of disulfide-linked peptides' fragmentation is an important first step in developing better automated approaches to assigning disulfide bonding using MS/MS data.   We contribute to this field by asking the question: How prevalent are multiple bond cleavages of disulfide-linked peptides, when the species undergo collision induced dissociation? After thoroughly analyzing MS/MS data for several different types of disulfide linked peptides, we unequivocally showed that product ions resulting from cleavage of at least two peptide backbone bonds (double cleavages) are prevalent in many different disulfide linked peptides' MS/MS data. Furthermore, the precise amount of double cleavages cannot be linked to charge state or the size of the peptide. This information, about the abundant presence of double cleavages in MS/MS analysis of disulfide-linked peptides, is important because it could be incorporated into disulfide analysis algorithms to increase the number of product ions assigned in the MS/MS data.   Additionally, we showed that assigning

double cleavage ions is useful when peptides contain more than two cysteines, because

these product ions can provide information about the connectivity of the cysteines.

2.5 References:

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223-230.

(2) Arnesano, F.; Banci, L.; Bertini, I.; Martinelli, M.; Furukawa, Y.; O'Halloran, T. V. *J. Biol. Chem.* **2004**, *279*, 47998-48003.

(3) Yano, H.; Kuroda, S.; Buchanan, B. B. *Proteomics* **2002**, *2*, 1090-1096.

(4) Angata, K.; Yen, T. Y.; El-Battari, A.; Macher, B. A.; Fukuda, M. *J. Biol. Chem.* **2001**, *276*, 15369-15377.

(5) Anfinsen, C. B.; Haber, E. *J. Biol. Chem.* **1961**, *236*, 1361-1363.

(6) Givol, D.; De Lorenzo, F.; Goldberger, R. F.; Anfinsen, C. B. *Proc. Natl. Acad. Sci. U. S. A.* **1965**, *53*, 676-684.

(7) Creighton, T. E. *Methods Enzymol.* **1984**, *107*, 305-329.

(8) Kim, H. I.; Beauchamp, J. L. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 157-166.

(9) King, G. J.; Jones, A.; Kobe, B.; Huber, T.; Mouradov, D.; Hume, D. A.; Ross, I. L. *Anal. Chem.* **2008**, *80*, 5036-5043.

(10) Sinz, A. *J. Mass Spectrom.* **2003**, *38*, 1225-1237.

(11) Soderblom, E. J.; Bobay, B. G.; Cavanagh, J.; Goshe, M. B. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 3395-3408.

(12) Soderblom, E. J.; Goshe, M. B. *Anal. Chem.* **2006**, *78*, 8059-8068.

(13) Thakur, S. S.; Balaram, P. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 358-366.

(14) Chrisman, P. A.; McLuckey, S. A. *J. Proteome Res.* **2002**, *1*, 549-557.

(15) Pearson, K. M.; Pannell, L. K.; Fales, H. M. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 149-159.

(16) Yen, T. Y.; Macher, B. A. *Methods Enzymol.* **2006**, *415*, 103-113.

(17) Yen, T. Y.; Yan, H.; Macher, B. A. *J. Mass Spectrom.* **2002**, *37*, 15-30.

(18) Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. *Nat. Methods* **2008**, *5*, 315-318.

(19) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255-261.

(20) Qin, J.; Chait, B. T. *Anal. Chem.* **1997**, *69*, 4002-4009.

(21) Yen, T. Y.; Joshi, R. K.; Yan, H.; Seto, N. O. L.; Palcic, M. M.; Macher, B. A. *J. Mass Spectrom.* **2000**, *35*, 990-1002.

(22) Mentinova, M.; Han, H.; McLuckey, S. A. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 2647-2655.

(23) Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J.-H.; Begley, T. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 672-678.

(24) Liu, J.; Gunawardena, H. P.; Huang, T. Y.; McLuckey, S. A. *Int. J. Mass Spectrom.* **2008**, *276*, 160-170.

(25) Gunawardena, H. P.; Gorenstein, L.; Erickson, D. E.; Xia, Y.; McLuckey, S. A. *Int. J. Mass Spectrom.* **2007**, *265*, 130-138.

(26) Standing, K. G. *Curr. Opin. Struct. Biol.* **2003**, *13*, 595-601.

(27) Gao, Q. X.; Xue, S.; Doneanu, C. E.; Shaffer, S. A.; Goodlett, D. R.; Nelson, S. D. *Anal. Chem.* **2006**, *78*, 2145-2149.

(28) Panchaud, A.; Singh, P.; Shaffer, S. A.; Goodlett, D. R. *J. Proteome Res.* **2010**, *9*, 2508-2515.

(29) Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 834-850.

(30) Wefing, S.; Schnaible, V.; Hoffmann, D. *Anal. Chem.* **2006**, *78*, 1235-1241.

(31) Xu, H.; Zhang, L. W.; Freitas, M. A. *J. Proteome Res.* **2008**, *7*, 138-144.

(32) Wells, J. M.; Stephenson, J. L.; McLuckey, S. A. *Int. J. Mass Spectrom.* **2000**, *203*, A1-A9.

(33) Hogan, J. M.; McLuckey, S. A. *J. Mass Spectrom.* **2003**, *38*, 245-256.

(34) Lioe, H.; O'Hair, R. A. J. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1109-1123.

(35) Go, E. P.; Zhang, Y.; Menon, S.; Desaire, H. *J. Proteome Res.* **2011**, *10*, 578-591.

Chapter 3: A Simple Approach to Assign Disulfide Connectivity Using Extracted Ion Chromatograms of Electron Transfer Dissociation Spectra

This work has been published by the journal Analytical Chemistry, and it is reproduced with permission from the journal.

Increasing interest in production of protein-based pharmaceuticals (biotherapeutics) is accompanied by an increased need for verification of protein folding and correct disulfide bonding. Recombinant protein expression may produce aberrant disulfide bonds and could result in safety concerns or decreased efficacy. Thus, the thorough analysis of disulfide bonding is a necessity for protein therapeutics. The use of ETD facilitates this analysis because disulfide bonds are preferentially cleaved when subjected to ETD. Here, we make use of this well-characterized reaction to assign disulfide bonding networks by coupling the use of extracted ion chromatograms (XICs) of cysteine-containing peptides with ETD analysis to produce an efficient assignment approach for disulfide bonding. This method can be used to assign a disulfide pattern in a *de novo* fashion, to detect disulfide shuffling, and to provide information on heterogeneity, when more than one disulfide bonding pattern is present. The method was applied for assigning the disulfide-bonding network of a recombinant monomer of the HIV envelope protein gp120. It was found that one region of the protein, the V1/V2 loops, had significant heterogeneity in the disulfide bonds.

3.1 Introduction:

Correct protein conformation is often vital to the protein's activity, and post-translational modifications such as disulfide bonds have substantial roles in maintaining the native fold. Disulfide bonds play a major structural role in proteins by introducing three-dimensional constraints on distant portions of protein. Disulfide bonds have been shown to have a variety of functions in proteins including facilitating viral entry[1], maintaining activity[2-4], and thermostability.[5] Maintaining and verifying correct disulfide bonds becomes increasingly important as more recombinantly-expressed proteins are used in vaccines and therapeutics.[6,7] Variation between batches could signal a problem with the production, purification, formulation, or storage of the protein therapeutic.[6,7] Non-native disulfide bonds could potentially mean decreased efficacy or cause adverse reactions to the use of the pharmaceutical.

The assignment of disulfide bonds is difficult because proteins usually contain multiple cysteine residues, and every additional cysteine adds another layer of complexity to the analysis. Mass spectrometry is the preferred technique for disulfide mapping because changes in mass due to formation of a disulfide bond can be detected with common commercial instruments, and fragmentation can be performed to verify the assignment. Collision induced dissociation (CID) is the most common dissociation technique for peptides, and has been previously applied for analysis of disulfide-bonded peptides.[8-12] Several new techniques for disulfide bond analysis have been developed recently, and they greatly facilitate disulfide assignment; these methods include electron capture dissociation[13], electron transfer dissociation[14], hydroxyl radical addition with subsequent CID[15], online electrolytic cleavage[16], free radical initiated peptide

sequencing with o-TEMPO-Bz-conjugated peptides[17], and direct ultraviolet photodissociation.[18]

Electron capture dissociation (ECD) and electron transfer dissociation (ETD) have been widely implemented in commercially-available instruments. Normal, linear peptides are known to fragment into c and z ions during ECD and ETD.[19,20] However, when disulfide-bonded peptides are subjected to ECD and ETD, highly abundant peaks for each linear peptide are observed due to homolytic cleavage of the disulfide bond.[13,14,21-24] Recently, Wu et al. has made use of this unique radical reaction to map disulfide-bonded peptides of monoclonal antibodies and other recombinant proteins.[21-23] However, the approach used by Wu requires tandem mass spectrometry using CID MS^2, ETD MS^2, and MS^3 involving ETD followed by CID on each of the fragments. Subsequent CID fragmentation of each ETD fragment provides more information about that ion but at a detrimental cost to the duty cycle. Other areas for improvement are also apparent for the aforementioned approach; it requires an intense precursor signal due to inefficiencies from alternative fragmentation pathways and it does not flag spectra that may be of interest due to specific ions detected. Commercial instruments are capable of taking thousands of spectra for every LC-MS injection, and identifying which of those spectra are relevant to disulfide bond analysis is a daunting task. It would be beneficial to expedite analysis by having a method that alerts the researcher to spectra of interest so that they may be targeted for further investigation.

Here, we demonstrate a unique and highly efficient approach to disulfide mapping using the homolytic reaction that disulfides undergo during ETD. Instead of searching the thousands of tandem mass spectra typically collected, searching

extracted ion chromatograms offers a way to expedite the data analysis. This approach typically reduces the number of spectra searched for a specific peptide to about 1% of the total spectra collected. Making use of extracted ion chromatograms (XICs) of the *m/z* values of cysteine-containing peptides allow a method of searching for disulfide pairs or bonding partners. The method was tested with several model proteins and a recombinant monomer of gp120, the envelope protein on the HIV-1 virus.

3.2 Experimental:

3.2.1 Materials and Reagents

Albumin from bovine serum (BSA), chicken lysozyme, human apo-transferrin, dithiothreitol (DTT), 4-vinylpyridine (4-VP), iodoacetamide (IAM), acetonitrile, formic acid, acetic acid, Tris-HCl, and Tris base were obtained in high purity from Sigma-Aldrich (St. Louis, MO). The HIV-1 Env 1086.C gp120 was expressed and purified in the laboratory of Dr. Barton F. Haynes, Duke Human Vaccine Research Institute (Duke University, Durham, NC), using the method described elsewhere.[25] Peptide N-glycosidase F (PNGase F) purified from *Flavobacterium meningosepticum* was purchased from New England BioLabs (Ipswich, MA). Sequencing-grade trypsin was obtained from Promega (Madison, WI). Ultrapure water was obtained from a Millipore Direct-Q3 filtration system (Billerica, MA).

3.2.2 Reduction and Alkylation

About 300 μg of lysozyme and BSA were reduced with 5 mM DTT for 1 hour in the dark at room temperature and alkylated with either 8 mM 4-VP or IAM for 1 hour in

the dark at room temperature. Tryptic digestion was carried out at 37 °C for 18 hours at a 1:30 enzyme-to-protein ratio (w/w). The digestion was stopped by the addition of 1 μL concentrated formic acid per 50 μL of digestion volume, and the resulting protein digest mixture was analyzed by reversed phase HPLC/ESI-LIT MS and ETD MS/MS.

3.2.3 Non-reduced Protein Digestion

About 300 μg of lysozyme and transferrin and 75 μg of 1086.C gp120 were alkylated with 5 mM IAM in the dark at room temperature to cap free cysteine residues prior to enzymatic digestion. Following alkylation with IAM, 1086.C gp120 was deglycosylated with 2 μL of PNGase F enzyme solution (≥10,000 units/mL) and was incubated at 37 °C for 72 hours. All three proteins were digested with trypsin. In-solution trypsin digestion was carried out at 37 °C for 18 h at a 1:30 enzyme-to-protein ratio (w/w). The digestion was stopped by the addition of 1 μL concentrated formic acid per 50 μL of digestion volume and the resulting protein digest mixture was analyzed by reversed phase HPLC/ESI-FTICR MS and HPLC/ESI-LIT-ETD MS/MS.

3.2.4 LC/ESI-FTICR MS Analysis

The tryptic digests were subjected to reversed phase-high performance liquid chromatography (RP-HPLC, Dionex, Sunnyvale, CA) coupled with a hybrid linear ion trap Fourier transform-ion cyclotron resonance (LTQ FTICR) mass spectrometer equipped with a 7 T actively shielded magnet (Thermo Scientific Corp, San Jose, CA). The mobile phase A was 99.9% water with 0.1% formic acid. Mobile phase B was 99.9% acetonitrile with 0.1% formic acid. Approximately 5 μL of the sample was injected

and separated on a C18 PepMap 300 column (150 mm × 300 μm i.d., 5 μM, 300 Å; LC Packings, Sunnyvale, CA) at a flow rate of 5 μL/min. The tryptic peptides were eluted using the following gradient, which was modified from the method described in the literature.[26] The mobile phase initially contained 2% B and the level of B increased linearly to 40% over 30 min, then ramped to 90% B over 20 min. The column was re-equilibrated after holding at 90% B for 10 min.

The samples were infused into the electrospray ion source, and the hybrid LTQ FTICR data acquisition was performed in a data-dependent scanning mode. Briefly, the MS[1] spectra were recorded in an FT MS scan and the six most abundant peptide ions in the MS scan were sequentially selected for CID performed in the LTQ mass analyzer, with 35% normalized collision energy and a 3 min dynamic exclusion window.

Electrospray ionization was achieved with a spray voltage of ~3.0 kV. Nitrogen was used as a nebulizing gas and set at a pressure of 10 psi. The capillary temperature was maintained at 200 °C. All data were acquired in the positive-ion mode and analyzed using Xcalibur 2.0 software (Thermo Scientific Corp, San Jose, CA).

3.2.5 LC/ESI-LIT-ETD MS/MS Analysis

The tryptic digests were subjected to reversed phase-high performance liquid chromatography (RP-HPLC, Waters Acquity, Milford, MA) coupled with a dual linear ion trap (LTQ Velos) mass spectrometer equipped with electron transfer dissociation (Thermo Scientific Corp, San Jose, CA). The mobile phase A was 99.9% water with 0.1% formic acid. Mobile phase B was 99.9% acetonitrile with 0.1% formic acid.

Approximately 5 μL of the sample was injected and separated on an Aquasil C18 column (150 mm × 300 μm i.d., 5 μM, 300 Å; Thermo Scientific, San Jose, CA) at a flow rate of 5 μL/min. The tryptic peptides were eluted using the following gradient. The mobile phase initially contained 3% B and the level of B increased linearly to 50% over 40 min, then ramped to 80% B over 20 min. The column was re-equilibrated after holding at 80% B for 10 min.

The samples were infused into the electrospray ion source, and the LTQ Velos data acquisition was performed in a data-dependent scanning mode. Briefly, the MS[1] spectra were recorded and the three most abundant ions in the MS scan were sequentially selected for ETD performed in the LTQ Velos mass analyzer with a two minute dynamic exclusion window.   A reaction time of 100 ms was used and the fluoranthene signal was optimized to approximately 1-6x10[7] counts prior to instrument use.

Electrospray ionization was achieved with a spray voltage of ~3.0 kV. Nitrogen was used as a nebulizing gas and set at a pressure of 10 psi. The capillary temperature was maintained at 250 °C. All data were acquired in the positive-ion mode and analyzed using Xcalibur 2.1 software (ThermoElectron Corp, San Jose, CA).

3.3 Results and Discussion:

3.3.1 Method Overview

A pictorial representation of the method used to assign disulfide bonds of dipeptides is shown in **Figure 1**. Briefly, ETD data of proteolyzed proteins is acquired in a data-dependent fashion throughout the chromatographic run, and these data are

searched by plotting extracted ion chromatograms (XICs) for each cysteine-containing peptide in the protein. Because disulfide linked peptides fragment efficiently across their S-S bond in ETD[14,20], plotting the XICs of the individual peptides, as shown in **Figure 1**, readily identifies the tandem mass spectra where each of the cysteine-containing peptides are present. The method we describe here involves interrogating each of these peaks and assigning them in order to assign the complete disulfide bonding in any given protein. The key advantage of this approach, over other competing approaches, is that the masses of the disulfide-linked peptides do not have to be known, or calculated, in advance. Instead, only the masses of the cysteine-containing peptides must be known, and the disulfide bonding pattern can be deduced, without interrogating the data file for precursor ions of a given mass. Additionally, because each peak in the XIC is interrogated, all the disulfide heterogeneity in a protein can be accounted for in an expedient manner.
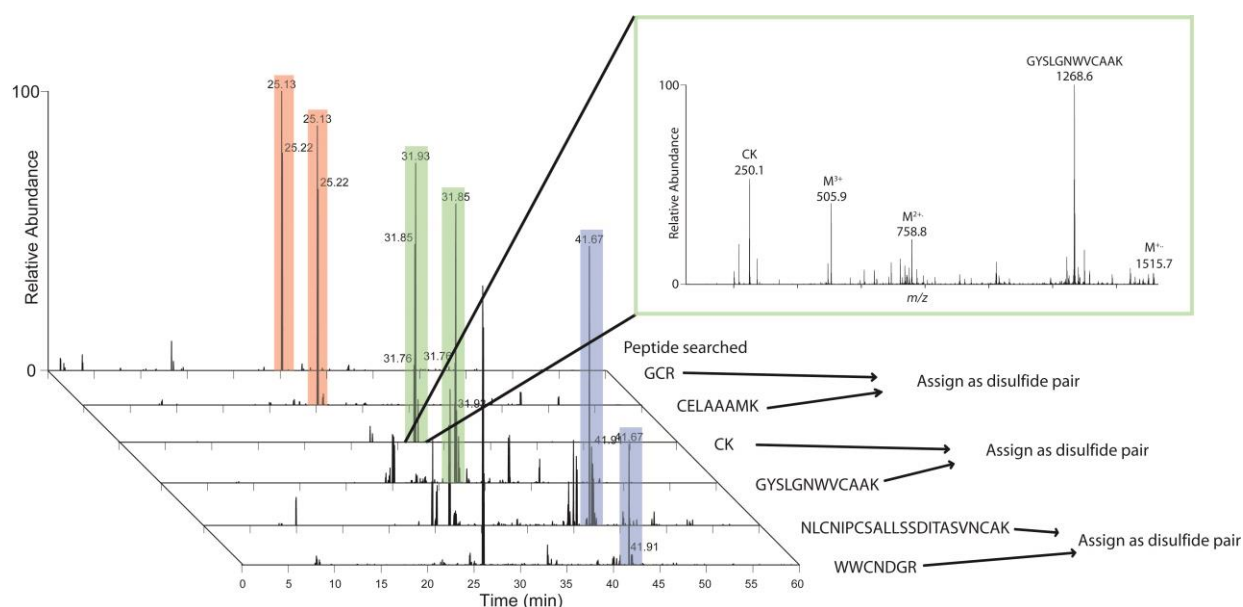
Figure 1: Assigning peptides with one disulfide bond. Extracted ion chromatograms are constructed to show where peptide mass marker ions of cysteine-containing peptides elute. Co-eluting peptides are assigned as disulfide pairs after further scrutiny, as described in the text. The inset shows an example mass spectrum of the peptides GYSLGNWVCAAK and CK that were disulfide bonded and dissociated in ETD. Data from lysozyme are shown.

When these cysteine-containing peptides are disulfide bonded to just one other peptide, the dipeptides are identified as pairs of peaks in the XIC plots. As an example, the top two XICs in **Figure 1** are for the peptides GCR and CELAAAMK. These two XICs each have a peak at 25.13 minutes. The peak corresponds to the ETD data for this disulfide-linked pair. **Figure 1** shows six examples of XICs for six different peptides, and three disulfide pairs are assigned. One can note that in addition to the disulfide pairs, other peaks appear in the XICs. Each of these peaks are also interrogated, and their identity is assigned as either a disulfide linked multipeptide (where more than two peptide chains are connected), an internally linked, disulfide bonded peptide, a peptide capped with alkylating reagent, or an interference, as described in **Figure 2**. Each of these special cases is elaborated upon later.
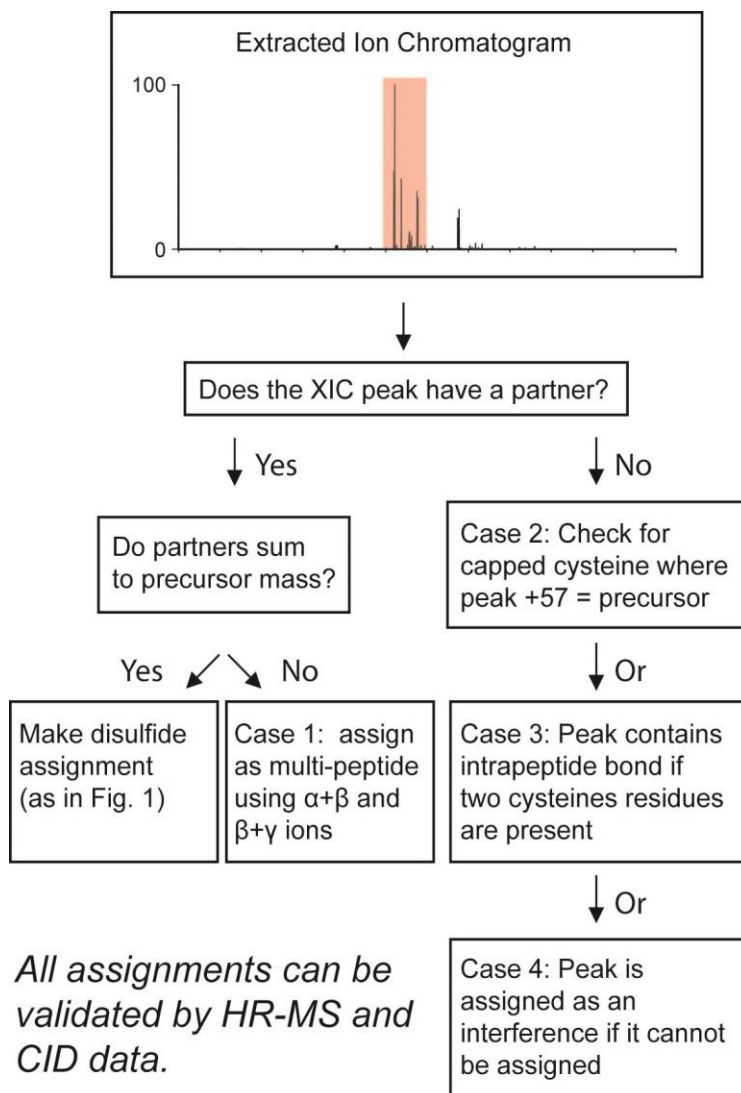
Figure 2: A representative workflow of the method. An XIC is produced for a cysteine-containing peptide and interrogated with the questions in the flowchart until it is assigned.

In addition to using the XIC data to assign the disulfide-bonded peptides, each assignment can be quickly validated in two additional ways: by validating masses of the assigned precursor ions and verifying that CID data supports the assignment. To determine the precursor ion mass, either high resolution data can be obtained or the charge state of the low resolution ETD spectrum can be deduced, so that the precursor's mass can be calculated. To determine the charge state from the ETD

spectrum, the data is queried to identify charge-reduced precursor ions. An example ETD spectrum, inset in **Figure 1**, shows the charged-reduced precursor ions in ETD data of disulfide-linked peptides are generally quite abundant. In this case, the charge reduced precursors at *m/z* 758.8 and 1518 can be used, along with the precursor ion, at *m/z* 505.9, to infer the charge state of the precursor is three. Using this information, the mass of the precursor ion can be calculated to be 1518, and this mass is equal to the sum of the two peptides assigned to be disulfide linked in this spectrum, thereby validating the assignment. For additional certainty, CID spectra can be acquired on the same peptide, and the CID data can be interrogated to verify that the peak is correctly assigned.

3.3.2 Special Cases

Case 1: Assigning Multipeptide Species. Disulfide linked peptides with three or more peptide chains are assignable with minor modifications to the above method. In these cases, generally two of the three peptide chains are detected individually as cleaved peptides in the XICs, but the masses of these two species do not add to the mass of the precursor ion. In a disulfide linked tripeptide, at least one peptide must contain two or more cysteine residues, and this peptide usually remains attached to one of its partner peptides during ETD. These tripeptide disulfides can be readily assigned because when the masses of the peptides with a single cysteine, which are detectable in the XICs, are subtracted from the precursor mass, the resulting mass perfectly matches that of the tryptic peptide with two cysteine residues. So the identity of each of the three peptides is determined.

Figure 3 shows ETD data for two tripeptides from transferrin that could be readily assigned using this strategy. **Figure 3**A was identified as an important ETD spectrum because two peptides, at *m/z* 1057 and 685, each had an intense peak in their XIC at 36.51 minutes. Yet, the ETD spectrum at this retention time (**Figure 3**A) clearly did not correspond to these peptides being linked as a dipeptide. Rather, the mass of the precursor indicated that an additional peptide of 1703 Da was present. Therefore, the tripeptide, as shown in the figure, was assigned, and ions corresponding to the β+γ ion and α+β ion were identified, further supporting the assignment. The α+β ion is detected at *m/z* values 1908 and 1273 in the 2+ and 3+ charge states, respectively, and the β+γ ion is detected at *m/z* 1194 in the 2+ charge state. **Figure 3**B shows a second example of ETD data from a tripeptide that was assigned in the same manner.
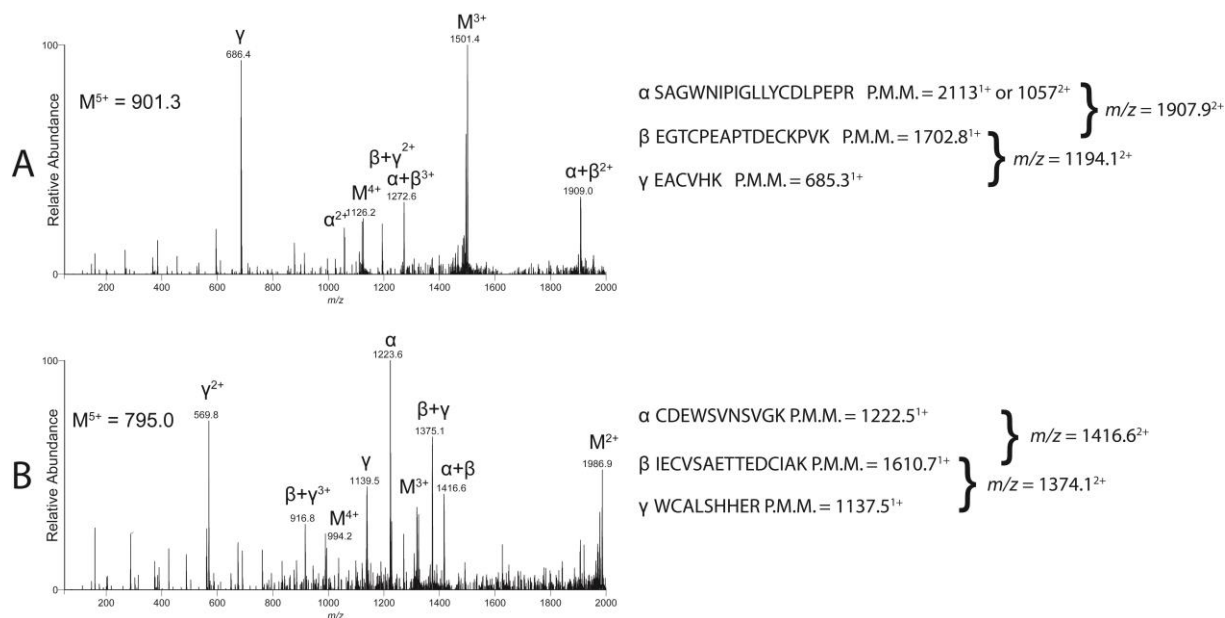

Figure 3: ETD spectra of two different disulfide-bonded peptides (A and B) with three linear peptides from human transferrin are shown. The most abundant ions are charge reduced precursors and combinations of α, α+β, β+γ, and γ peptides which allows the investigator to confirm the β peptide.

Case 2: Assigning Free Cysteine. Any protein may contain free cysteine residues, which do not participate in disulfide bonding. To characterize these non-linked cysteine residues using our XIC/ETD approach, all readily observable peaks present in the XICs that are not assigned as disulfide linked peptides are further evaluated. For these peaks, when the precursor ion mass is equivalent to the peptide mass plus the mass of the capping reagent, (for example, plus 57 Da when IAM is used), the peak is assignable as an alkylated peptide, originating from a protein with a free cysteine in it.

The above-described approach requires that the peptide/alkylating reagent bond is cleaved each time the alkylated peptide is subjected to ETD. Therefore, to identify an ideal alkylating reagent for this work, two commonly used alkylating reagents, iodoacetamide and 4-vinylpyridine, were tested extensively, and ETD data for alkylated peptides are shown in **Figure 4**. Of these two choices, IAM was identified as an ideal capping reagent. When peptides alkylated with iodoacetamide were subjected to ETD, the peptide mass marker ions were among the most intense peaks in the spectrum, and the only peaks with a higher abundance were precursor, charge-reduced precursor, and neutral loss ions (**Figure 4**A and B). In contrast, the peptides alkylated with 4-vinylpyridine did not show a loss of the alkylating reagent to reveal the peptide mass marker ions of *m/z* 1268.7 and 936.5 in **Figure 4**C and D respectively. Therefore, 4-vinylpyridine is not a useful reagent, when the goal is to identify capped cysteine residues by evaluating the XICs of cysteine-containing peptides.
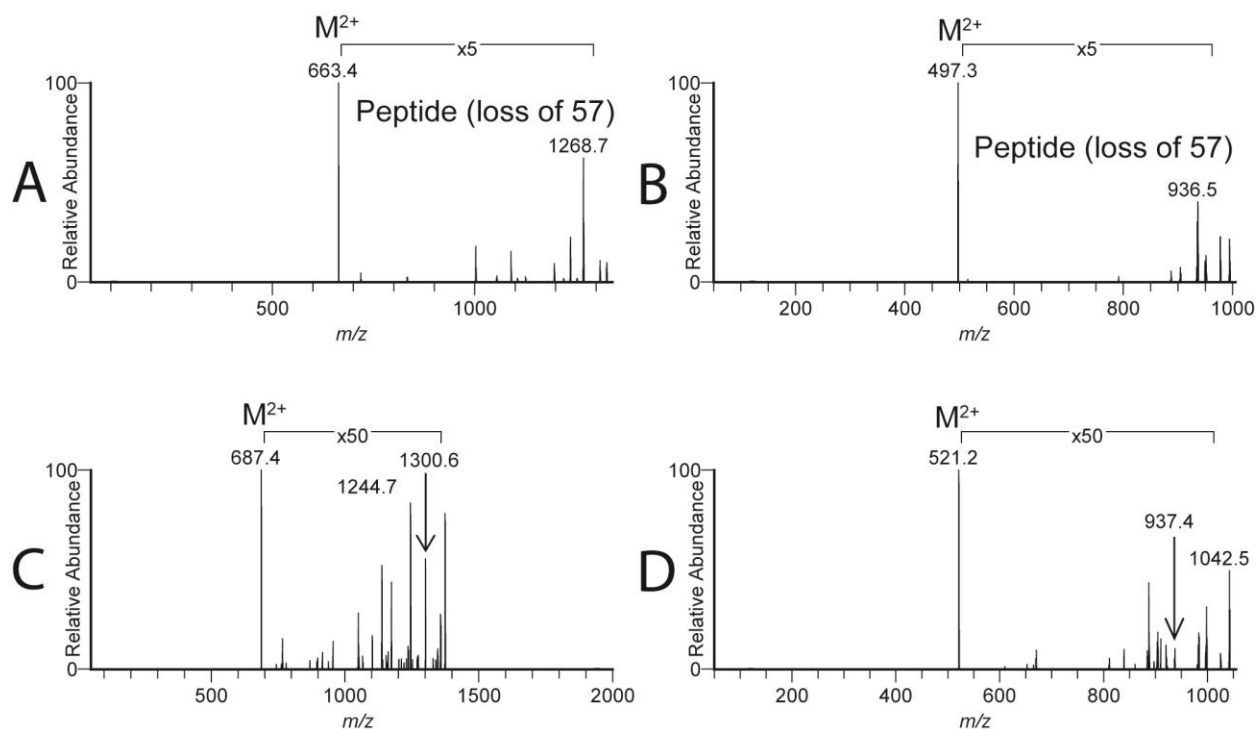
Figure 4. Results from ETD on peptides from lysozyme that were reduced and alkylated. Peptides alkylated with iodoacetamide (A and B) show a prominent loss of 57, corresponding to loss of the alkylating reagent. When the same peptides were alkylated with 4-vinylpyridine (C and D), the peptide mass marker was not detected after ETD. A and C are peptide GYSLGNWVCAAK with the bare peptide expected at 1268.6 *m/z*. B and D are peptide WWCNDGR with the bare peptide expected at 936.4 *m/z*.

Case 3: Internally linked disulfides. The peaks in the XICs may also be due to a single tryptic peptide with two cysteine residues, where those two cysteine residues are connected to each other. The XICs can be used to preliminarily identify the internally linked peptides, because in those cases, the peak in the XIC must correspond to a peptide with two cysteine residues in it; it must not have a "partner", a peak in any other XIC with the same retention time; and the XIC peak must correspond to the mass of the precursor or one of the charge-reduced species. This case is highlighted in **Figure 2**.

Case 4: Interferences. Not every peak in the XICs generated by this method can be assigned to a disulfide linked or capped cysteine-containing peptide. The peaks may also be due to interferences such as linear peptide precursors, charge-reduced precursors, and fragment c or z ions. These peaks are quickly evaluated and can be disregarded when they do not fulfill any of the criteria specified above; namely, they do not have a "partner," an ion in another XIC that has an identical retention time, and they are not assignable as internally linked peptides or peptides capped with an alkylating reagent.

Using the above outlined approach, the disulfide bonding in two model proteins, transferrin and lysozyme, was assigned, and in each case the XIC/ETD method provided assignment information that was consistent with the known disulfide bonding pattern of the proteins. The assignments made for transferrin are presented in **Table 1** and lysozyme assignments are provided in Supplemental **Table 1**. All cysteine-containing peptides were detected in their accepted disulfide map with the exception of one from transferrin. This peptide (shown in red in **Table 1**) consists of three intrachain disulfide bonds and was not detected.

Table 1: Disulfide assignments for transferrin from ETD-XIC data.

Peptide	Peptide (XIC)* m/z Theor.	Peptide (XIC)* m/z Exptl.	Dipeptide** m/z Theor.	Dipeptide** m/z Exptl.
WCAVSEHEATK	1259.6	1259.6		
ASYLDCIR	939.5	939.5		
CQSFR	639.3	639.3		
SVIPSDGPSVACK	1357.7	1357.7		
SCHTGLGR	829.4	829.4		
CLK	362.2	362.3		
SAGWNIPIGLLYCDLPEPR	1057.0^{2+}	1057.1^{2+}	1272.3^{3+}	1272.2^{3+}
EGTCPEAPTDECKPVK	NA	NA		
EACVHK	685.3	685.4	1194.0^{2+}	1194.0^{2+}
AVANFFSGSCAPCADGTDFPQLC-QLCPGCGCSTLNQYFGYSGAFK	NA	NA		
DQYELLCLDNTR	1481.7	1481.7		
DCHLAQVPSHTVVAR	1631.8	1632.0		
WCALSHHER	1137.5	1137.6	1374.1^{2+}	1374.1^{2+}
IECVSAETTEDCIAK	NA	NA		
CDEWSVNSVGK	1222.5	1222.5	1417.1^{2+}	1417.2^{2+}
CGLVPVLAENYD†K	1419.7	1419.7	1418.7^{2+}	1418.7^{2+}
CSTSSLLEACTFR	NA	NA		
INHCR	641.3	641.4	1029.0^{2+}	1029.0^{2+}
SDNCEDTPEAGYFAVAVVK	1007.5^{2+}	1007.5^{2+}		
DDTVCLAK	863.4	863.4		
SCHTAVGR	829.4	829.5		
CLVEK	590.3	590.3		
FDEFFSEGCAPGSK	1519.6	1519.6	1556.2^{2+}	1556.2^{2+}
LCMGSGLNLCEPNNK	NA	NA		
DSSLCK	651.3	651.3	1121.5^{2+}	1121.6^{2+}
DYELLCLDGTR	1296.6	1296.6		
KPVEEYANCHLAR	1528.8	1528.8		
QQQHLFGSD†VTDCSGNFCLFR	1201.0^{2+}	1201.0^{2+}		

ETD-XIC data were used to make assignments by aligning expected peptide ETD product ions. Assignments were checked against accepted disulfide bonds in Uniprot (Accession P02787). One cysteine-containing peptide from transferrin was not assigned (in red) with the method. *Peptide product ion m/z values of single, linear peptides. **Peptide product ion m/z values of dipeptides as a result of a single, linear peptide loss from a tripeptide. †Residues deglycosylated by PNGase F.

The aforementioned method was also used to investigate the disulfide-bonding patterns of a recombinant monomer of the HIV-1 envelope protein 1086.C gp120. This protein has 18 cysteine residues on 16 tryptic peptides. The protein sequence is available in Supplemental **Figure 1**. To do a complete *de novo* analysis of the disulfide bonding of this protein using traditional approaches where the potential precursor ions of all the possible disulfide linked pairs are calculated and searched, 426 precursor masses would need to be searched for, each in multiple charge states. In practice, more than a thousand precursor ions would have to be individually searched for in the MS^1 data. Instead, our method simply uses the 16 XIC plots for each tryptic peptide, and infers the disulfide bonding from these plots.

Two of the XICs from the gp120 data are shown in **Figure 5**, along with the assignments for the peaks. For these two chromatograms, three different types of species are present: a dipeptide disulfide is detected, along with two different multipeptide disulfides. Finally, several spectra that correspond to interferences were also observed. These assignments were based on the approach outlined in **Figure 2**. After preliminary assignment of each peak in the XIC (as dipeptide disulfide, multipeptide disulfide, interference, etc.), each assignment was confirmed with high resolution MS data along with CID data.
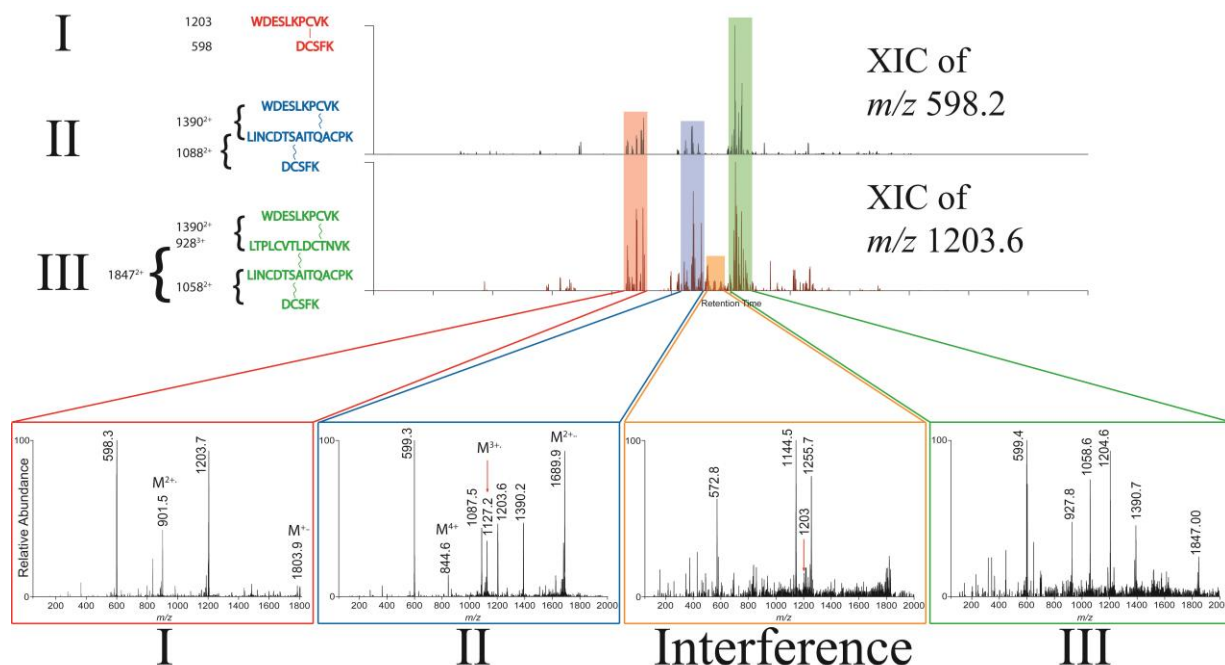
Figure 5. The different forms detected from the V1/V2 region of 1086.C gp120 can be seen when viewing the XICs for *m/z* 598 and 1203 (top and bottom respectively). An interfering species is seen with *m/z* 1203 from 28 to 30 minutes. It was the result of a neutral loss from a charge reduced precursor from a different peptide.

The aggregate results from the disulfide mapping of the gp120 protein are shown in **Table 2**. These data point to several interesting observations. First, substantial heterogeneity is observed in the first and second variable regions, the V1 and V2 regions, of the protein, which includes the cysteine residues on the 2nd -5th tryptic peptides. Yet, outside that part of the protein sequence, all other cysteine residues proved to be exceedingly faithful to their respective partners and were found only bonded to one partner peptide. Additionally, no alkylation of peptides was detected. Aside from the heterogeneity in the V1/V2 region, all the disulfide assignments are consistent with those published by Leonard *et al.* in the early 1990's on a different gp120, where Edman sequencing was used to assign the peptides.[27] While

heterogeneity in the V1/V2 region was not reported in the earlier work, it has been reported by us on a similar protein, an oligomeric gp140.[10]

Table 2: Disulfide assignments for gp120 from ETD-XIC and FTICR-MS data.

Peptide	ETD XIC m/z	Theoretical m/z	Experimental m/z	Mass error (ppm)	Charge state
EVHNVWATCHACVPTDPNPQEMVL	1294^{2+}	910.9277	910.9290	1	4+
TTLFCASDAK	1055.5	1214.2343	1214.2387	4	3+
		1820.8476	1820.8466	<1	2+
WDESLKPCVK	1203.6	844.6492	844.6497	<1	4+
LINCDTSAITQACPK	NA	1125.8630	1125.8674	4	3+
DCSFK	598.2				
WDESLKPCVK	1203.6	816.2292	816.2305	2	6+
LINCDTSAITQACPK	NA	979.2735	979.2782	5	5+
LTPLCVTLDCTNVK	NA	1223.8399	1223.8452	4	4+
DCSFK	598.2				
LTPLCVTLDCTNVK	1518.8	759.3893	759.3886	<1	2+
		1517.7708	1517.7747	3	1+
LINCDTSAITQACPK	1576.8	788.3795	788.3796	<1	2+
		1575.7511	1575.7585	5	1+
VSFDPIPLHYCAPAGFAILK	1080^{2+}	861.6933	861.6937	<1	4+
DVSTVQCTHGIK	1286.6	1148.5884	1148.5928	4	3+
		1722.3787	1722.4146	21	2+
TFDGTGPCR	NA	715.2959	715.2997	5	2+
CNDK	NA				
TIIVHLDESVNIVCTRPNDNTR	1255^{2+}	730.9594	730.9596	<1	5+
QAHCNIDESK	1143.5	913.4472	913.4494	2	4+
		1217.5937	1217.5992	5	3+
TIIVHLDESVNIVCTRPNDNTR	1255^{2+}	727.5541	727.5548	1	5+
Q*AHCNIDESK	1126.5	909.1906	909.1925	2	4+
		1211.9182	1211.9228	4	3+
AIYAPPIEGEITCNSDITGLLLLR	1286^{2+}	799.6604	799.6627	3	4+
SFNCR	625.3	1065.8779	1065.8833	5	3+
		1598.3130	1598.3229	6	2+
GEFFYCDTSDLFDTYR	1018^{2+}	797.0988	797.1004	2	4+
SSDGTITLQCK	1151.5	1062.4625	1062.4672	4	3+
		1593.1898	1593.1987	6	2+

ETD spectra were used for assignments and HR-MS and CID spectra, for validation. Asterisks represent pyroglutamate. No ETD data was acquired for the disulfide bonded peptide in red, but it was assigned using HR-MS and CID data.

The XIC/ETD method presented here was not able to identify all disulfide assignments presented in **Table 2**. One dipeptide, TFDGTGPCR – CNDK, could not be assigned in the ETD data. Searching the high resolution data revealed the cause: The peptide is small, hydrophilic, and has a short retention time. It co-eluted with several other small peptides and unidentified species such that the signal intensity of this species was probably negatively impacted. CID data was acquired on this species, but due to the longer reaction time of ETD, this peak was not selected for fragmentation. Additional efforts in chromatography and/or a re-design of the ETD data-dependent scanning method could alleviate or mitigate this problem in future analyses. This disulfide assignment was made by confirming the mass of the disulfide-bonded peptide in the high resolution FTICR-MS data and by assigning the CID data associated with this peak.

3.4 Conclusion:

We developed a new method to map disulfide bonds using extracted ion chromatograms of ETD tandem mass spectra and implemented it to assign the disulfides of a recombinant HIV-1 envelope protein, 1086.C gp120. Before using this method to analyze gp120, the disulfides of two model proteins, lysozyme and transferrin, were assigned to validate the method. These model proteins have a vast array of different disulfide connectivities including intrapeptide disulfides, dipeptides, and tripeptides. Assignments were in agreement with the accepted disulfide linkages reported in Uniprot. Free cysteine residues were mimicked by purposely reducing and alkylating proteins.

This method was then used to map the disulfide bonds in recombinant HIV-1 Env, 1086.C gp120. The method was able to assign the bonding partners of 16 of the 18 cysteine residues and showed that select cysteine residues had several different bonding configurations. While several different structures resulted from analysis of the peptides in the V1/V2 loops, all other disulfide bonds in the protein proved to be exceedingly homogenous and only bonded with their expected partner peptide. The only disulfide peptide pair not assigned was detected but eluted with several other species and was not selected for ETD due to decreased signal.

The method outlined here has proven to be an expedient way of assigning the disulfide bonds of both model proteins and the various structures formed in a recombinant HIV-1 envelope protein. While the method still requires the logical capacity of a human to sort out some of the more complex cases, implementation into a computer program to automate the analysis of disulfide bonds is feasible. With or without automation, the method described herein, combining ETD and XIC's, is a large step forward for assigning disulfide bonds in any protein.

3.5 References:

(1) Azimi, I.; Matthias, L. J.; Center, R. J.; Wong, J. W. H.; Hogg, P. J. *J. Biol. Chem.* **2010**, *285*, 40072-40080.

(2) Chen, H.; Zhang, G.; Zhang, Y.; Dong, Y.; Yang, K. *Biochemistry* **2000**, *39*, 12140-12148.

(3) D'Amico, S.; Gerday, C.; Feller, G. *J. Biol. Chem.* **2002**, *277*, 46110-46115.

(4) Mills, J. E.; Whitford, P. C.; Shaffer, J.; Onuchic, J. N.; Adams, J. A.; Jennings, P. A. *J. Mol. Biol.* **2007**, *365*, 1460-1468.

(5) Reading, N. S.; Aust, S. D. *Biotechnol. Prog.* **2000**, *16*, 326-333.

(6) Trivedi, M. V.; Laurence, J. S.; Siahaan, T. J. *Curr. Protein Pept. Sci.* **2009**, *10*, 614-625.

(7) Hapuarachchi, S.; Fodor, S.; Apostol, I.; Huang, G. *Anal. Biochem.* **2011**, *414*, 187-195.

(8) Gupta, K.; Kumar, M.; Balaram, P. *Anal. Chem.* **2010**, *82*, 8313-8319.

(9) Brgles, M.; Kurtović, T.; Halassy, B.; Allmaier, G.; Marchetti-Deschmann, M. *J. Mass Spectrom.* **2011**, *46*, 153-162.

(10) Go, E. P.; Zhang, Y.; Menon, S.; Desaire, H. *J. Proteome Res.* **2011**, *10*, 578-591.

(11) Clark, D. F.; Go, E. P.; Toumi, M. L.; Desaire, H. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 492-498.

(12) Janecki, D. J.; Nemeth, J. F. *J. Mass Spectrom.* **2011**, *46*, 677-688.

(13) Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 2857-2862.

(14) Chrisman, P. A.; Pitteri, S. J.; Hogan, J. M.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1020-1030.

(15) Xia, Y.; Cooks, R. G. *Anal. Chem.* **2010**, *82*, 2856-2864.

(16) Zhang, Y.; Dewald, H. D.; Chen, H. *J. Proteome Res.* **2011**, *10*, 1293-1304.

(17) Lee, M.; Lee, Y.; Kang, M.; Park, H.; Seong, Y.; Sung, B. J.; Moon, B.; Oh, H. B. *J. Mass Spectrom.* **2011**, *46*, 830-839.

(18) Agarwal, A.; Diedrich, J. K.; Julian, R. R. *Anal. Chem.* **2011**, *83*, 6455-6458.

(19) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.

(20) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. USA.* **2004**, *101*, 9528-9533.

(21) Wu, S. L.; Jiang, H.; Lu, Q.; Dai, S.; Hancock, W. S.; Karger, B. L. *Anal. Chem.* **2009**, *81*, 112-122.

(22) Wu, S. L.; Jiang, H.; Hancock, W. S.; Karger, B. L. *Anal. Chem.* **2010**, *82*, 5296-5303.

(23) Wang, Y.; Lu, Q.; Wu, S. L.; Karger, B. L.; Hancock, W. S. *Anal. Chem.* **2011**, *83*, 3133-3140.

(24) Pike, G. M.; Madden, B. J.; Melder, D. C.; Charlesworth, M. C.; Federspiel, M. J. *J. Biol. Chem.* **2011**, *286*, 17954-17967.

(25) Liao, H. X.; Sutherland, L. L.; Xia, S. M.; Brock, M. E.; Scearce, R. M.; Vanleeuwen, S.; Alam, S. M.; McAdams, M.; Weaver, E. A.; Camacho, Z. T.; Ma, B. J.; Li, Y. Y.; Decker, J. M.; Nabel, G. J.; Montefiori, D. C.; Hahn, B. H.; Korber, B. T.; Gao, F.; Haynes, B. F. *Virology* **2006**, *353*, 268-282.

(26) Ihling, C.; Berger, K.; Höfliger, M. M.; Führer, D.; Beck-Sickinger, A. G.; Sinz, A. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1240-1246.

(27) Leonard, C. K.; Spellman, M. W.; Riddle, L.; Harris, R. J.; Thomas, J. N.; Gregory, T. J. *J. Biol. Chem.* **1990**, *265*, 10373-10382.

Chapter 4: Reporter Peptides: How to Guarantee Disulfide Assignments are not Artifacts of Sample Preparation

The use of electron transfer dissociation has significantly increased recently and is partially due to its unique ability to leave labile post-translational modifications intact yet cause homolytic disulfide scission. Disulfide bonds are of particular importance because protein pharmaceuticals are becoming more prevalent in medicine and ensuring the correct tertiary structure is imperative for proper efficacy and safety. Extreme care must be exercised because disulfides can easily be changed from native to a shuffled form through improper handling during expression, purification, or formulation. Furthermore, disulfide bonds that are engineered and the possibility of multiple disulfide maps necessitates a method to ensure disulfide assignments are not simply artifacts due to the sample preparation required. We present a method for checking disulfide shuffling during enzymatic deglycosylation and proteolytic digestion utilizing known reporter peptides that are mixed equimolar with the target protein during sample preparation. When subjected to ETD, these reporter peptides will produce a characteristic *m/z* which enables an expedient data analysis search to determine optimal digestion conditions for disulfide mapping. These experiments revealed with model proteins that when optimal digestion conditions were used, no detectable shuffling of disulfides within the target protein or between target protein and reporter peptide occurred. However, use of non-optimal conditions produced shuffling both within the target protein and with the reporter peptide.

4.1 Introduction:

Sample preparation techniques have improved substantially over the past few decades but can be a major source of error and trouble with analytical systems.[1-3] Solid phase extraction, off-line fractionation, proteolytic digestion, and enzymatic degradation all have drawbacks including inefficiency in sample recovery, time consumption, and incomplete proteolytic or enzymatic digestion.[1,3] When such sample preparation techniques are required for the experiment each procedure must be optimized to provide the best possible results. To this extent, significant time is often invested ensuring the sample preparation techniques used do not have substantial deleterious effects on the analysis. Advancements in ionization, instrumentation, and experimental design have significantly decreased the requirement of sample preparation techniques and has allowed increasingly complex proteomic experiments to be performed.[4] At the same time sample preparation for any type of mass spectrometry experiment is more complicated than most other techniques because not only must the aforementioned procedures be optimized, but special requirements to make the sample amenable to mass spectrometry must be used. Specifically, the solvent must easily evaporate and the buffer and mobile phases must use pH modifiers that easily volatilize. The requirements for a mass spectrometer "friendly" sample are often not the same as a molecular biologist's and this may lead to detrimental effects on analytes.

The disulfide bond is a post-translational modification (PTM) that imposes a strict distance constraint via a covalent bond and it is this nature that makes the disulfide bond unique among most other PTMs. Analysis of disulfide bonds is necessary because incorrect or shuffled disulfides could mean reduced efficacy or elicit an immune

response. Recombinant and natural immunoglobulins have previously been shown to take on scrambled disulfide structures due to sample handling[5,6] and molecular dynamics simulations have been employed to understand the shuffling.[7] Engineered disulfide bonds have been popular as researchers look for ways to covalently hold distant parts of the protein close together to either stabilize or destabilize the structure.[8-13] These engineered disulfides must be checked to be sure an alternative disulfide structure is not responsible for biological effects experienced. Analysis of disulfides is often done with mass spectrometry when samples are impure or in small quantities because electrospray ionization can couple a liquid-based separation technique to mass spectrometry. Liquid chromatography has become the separation technique of choice and can be used to separate peptides so the mass spectrometer's mass measurement, low limit of detection, and tandem MS abilities can be utilized to get better coverage of the protein. However, using procedures amenable to mass spectrometry requires taking the sample out of its biologically-acceptable buffer, solvent, and salt concentration. Other groups have previously demonstrated a top-down approach to map disulfide bonds.[14-16] However, with the recent observation of disulfide shuffling during ECD and ETD fragmentation techniques, a bottom-up approach would be beneficial to the analysis so the number of disulfides possibly shuffled during dissociation can be minimized. Employment of a bottom-up approach then requires exposure to various digestion procedures, which could have detrimental effects on the disulfide bonds of the protein and must be analyzed for proper quality control.

The use of reporter ions that form during fragmentation has been present in research for more than a decade. Isotope-coded affinity tags (ICATs)[17,18], tandem mass

tags (TMTs)[19,20], and isobaric tags for relative and absolute quantification (iTRAQ)[21-27] have been applied as a means of quantification of proteins previously. Experiments involving isotope-coded affinity tags (ICATs) originally served to compare changes in protein expression but were expanded by Sethuraman et Al. to include the investigation of oxidant-sensitive cysteine thiols.[28] However, Sethuraman used the cysteine-specific tag for the relative quantification of oxidant-sensitive cysteine residues and used CID to produce the reporter ion.[28] ETD is often used for disulfide analysis as it also allows the researcher to both assign the disulfide bonding structure and monitor for aberrant structures or shuffling as well.[5,6,29] To make a comparable analysis using ETD, we need a reporter ion that produces a characteristic peak when it under goes ETD and there is disulfide scrambling present in the sample.

Here, we describe a qualitative method for checking disulfide-bonded peptides that may have shuffled due to improper sample handling or digestion conditions. The reporter ion approach was expanded to employ the use of a small reporter peptide which if detected bonded to any peptide from the target protein indicates disulfide shuffling. We evaluated the method with lysozyme in both optimal and non-optimal digestion conditions. The reporter ion was not detected as disulfide bonded to any peptides from either of the target proteins when optimal digestion conditions were used. Incorrect or non-optimal conditions produced detectable shuffling between reporter and target peptides within a standard proteolytic timeframe.

4.2 Experimental:

4.2.1 Materials and Reagents

Chicken lysozyme, iodoacetamide (IAM), acetonitrile, formic acid, acetic acid, Tris-HCl, and Tris base were purchased in high purity from Sigma-Aldrich (St. Louis, MO). Somatostatin-14 was procured from AnaSpec (San Jose, CA) at >95% purity. Peptide N-glycosidase F (PNGase F) purified from *Flavobacterium meningosepticum* was obtained from New England BioLabs (Ipswich, MA). Sequencing-grade trypsin was purchased from Promega (Madison, WI). Ultrapure water was obtained from a Millipore Direct-Q3 filtration system (Billerica, MA).

4.2.2 Protein Digestion

About 300 μg of lysozyme was alkylated with 5 mM IAM in the dark at room temperature to cap free cysteine residues prior to enzymatic digestion. Following alkylation with IAM, lysozyme was digested with trypsin. In-solution trypsin digestion was carried out at 37 °C for 18 h at a 1:30 enzyme-to-protein ratio (w/w). The digestion was stopped by the addition of 1 μL concentrated formic acid per 50 μL of digestion volume and the resulting protein digest mixture was analyzed by reversed phase HPLC/ESI-LIT-ETD MS/MS.

4.2.3 LC/ESI-LIT-ETD MS/MS Analysis

The tryptic digests were subjected to reversed phase-high performance liquid chromatography (RP-HPLC, Waters Acquity, Milford, MA) coupled with a dual linear ion trap (LTQ Velos) mass spectrometer equipped with electron transfer dissociation (Thermo Scientific Corp, San Jose, CA). The mobile phase A was 99.9% water with 0.1% formic acid. Mobile phase B was 99.9% acetonitrile with 0.1% formic acid. Approximately 5 μL of the sample was injected and separated on an Aquasil C18 column (150 mm × 300 μm i.d., 5 μM, 300 Å; Thermo Scientific, San Jose, CA) at a flow rate of 5 μL/min. The tryptic peptides were eluted using the following gradient. The mobile phase initially contained 3% B and the level of B increased linearly to 50% over 40 min, then ramped to 80% B over 20 min. The column was re-equilibrated after holding at 80% B for 10 min.

The samples were infused into the electrospray ion source, and the LTQ Velos data acquisition was performed in a data-dependent scanning mode. Briefly, the MS[1] spectra were recorded and the three most abundant ions in the MS scan were sequentially selected for ETD performed in the LTQ Velos mass analyzer with a two minute dynamic exclusion window. A reaction time of 100 ms was used and the fluoranthene signal was optimized to approximately 1-6x10[7] counts prior to instrument use.

Electrospray ionization was achieved with a spray voltage of ~3.0 kV. Nitrogen was used as a nebulizing gas and set at a pressure of 10 psi. The capillary temperature

was maintained at 250 °C. All data were acquired in the positive-ion mode and analyzed using Xcalibur 2.1 software (ThermoElectron Corp, San Jose, CA).

4.3 Results and Discussion:

The disulfide reporter peptide method is outlined in **Figure 1**. Briefly, prior to sample preparation a reporter peptide or protein is introduced to the sample with approximately an equimolar concentration compared to the target protein. The reporter is present during all stages of sample preparation so that if any disulfide shuffling occurs there should also be evidence of shuffling either within the reporter or between target and reporter. The sample is subjected to LC-MS/MS with ETD as a fragmentation technique. The well-characterized homolytic cleavage of the disulfide bond is then used to quickly check the data for shuffling between reporter and target analytes. Extracted ion chromatograms (XICs) of the unique mass or masses of reporter peptides provides an expedient check to see if any disulfide shuffling was detected as reporter peptides being bonded to any peptides from the target protein. The major benefit of this method is in case aberrant disulfide bonds are detected, they can be proven not to be a consequence of the sample preparation, but instead are indicative of a sample that contained alternative disulfide structures.
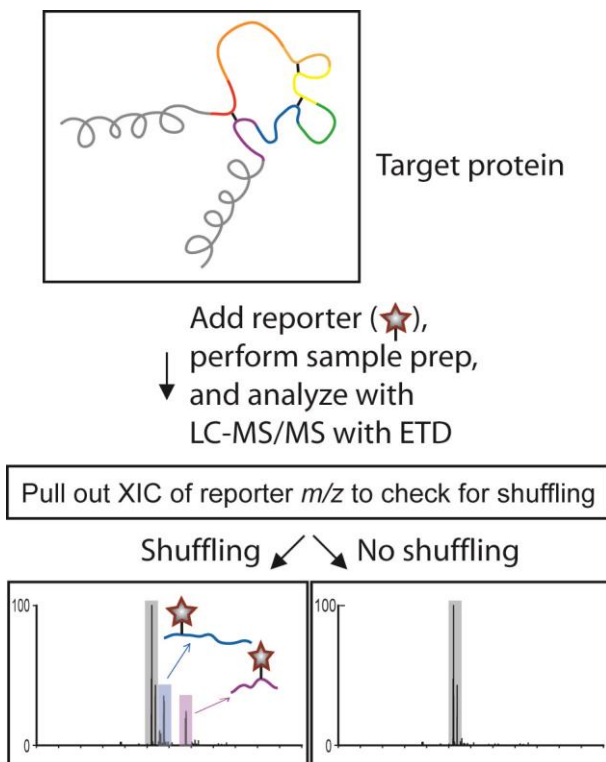
Figure 1. Reporter peptide approach to disulfide analysis. The reporter is introduced before any sample preparation takes place. If shuffling occurs, shuffling will also occur with the reporter which can be easily detected.

ETD often provides a way to calculate the charge state of the precursor based on the un-dissociated charge-reduced precursor species that are present in the product spectrum. The charge state can be used to calculate the experimental mass and the theoretical precursor mass can be checked against the assignment. Typically there are also lower intensity peaks than the disulfide scission products in ETD spectra as a result of the standard c and z ion formation mechanism of cleavage of the N-Cα bond which can be employed to further verify the assignment of each disulfide-shuffled ion. Detection and identification of shuffled disulfides represents a situation where sample preparation needs to be performed in a different manner in order to minimize artifact

disulfide assignments. **Figure 2**(a) shows an XIC of *m/z* 249.1 that corresponds to a tryptic peptide from lysozyme that contains a cysteine. This sample was purposely mistreated during sample preparation to illustrate the shuffling that can occur and subsequently be detected. There were two assignments of shuffled disulfides with this peptide (Figures 2(b) and (c)) that formed due to the improper handling and incorrect conditions. **Figure 2**(d) represents the native disulfide and composes the most intense of all the species. Two of the XIC peaks at retention times of 14 and 25 minutes could not be assigned as tryptic peptides.
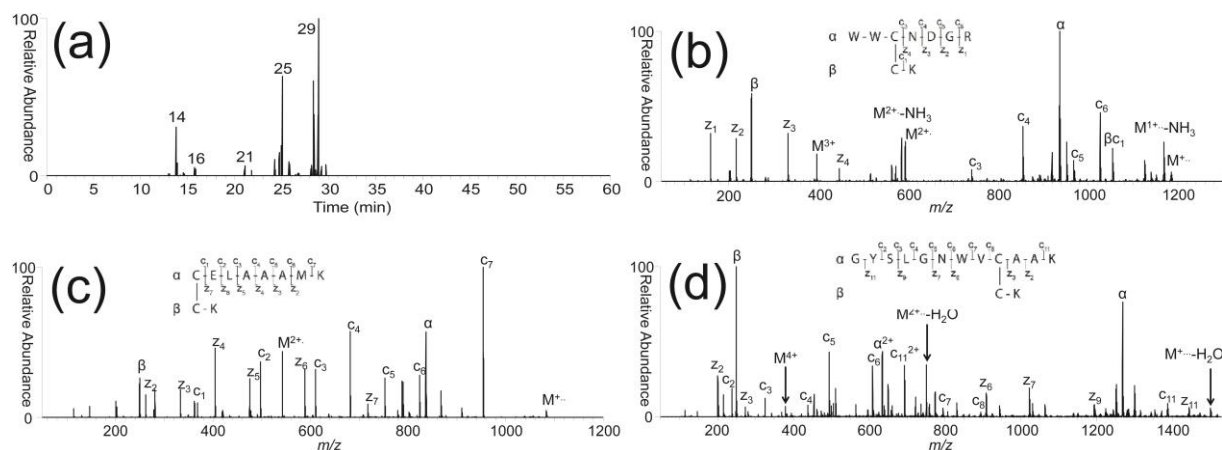
Figure 2. Results of non-optimal lysozyme proteolysis represented with (a) an XIC from ETD data of *m/z* 249 corresponding to peptide CK, (b) a shuffled disulfide at 16 minutes, (c) a shuffled disulfide at 21 minutes, (d) and the native disulfide at 29 minutes. Peaks at 14 and 25 minutes were not assigned.

4.3.1 Reporter protein/peptide selection

Proper selection of a reporter protein or peptide is a task that needs to be undertaken for every target protein of interest as one single suitable reporter is not feasible while utilizing this method of data analysis. Above all else, the reporter peptide

needs to be a different mass than any other cysteine-containing peptide from the target protein. This allows the researcher to use the XIC method of assigning disulfide connectivity because the possibility of isobaric interferences between reporter and target proteins has been eliminated. The reporter should also be small enough that the disulfide is not sterically-hindered and reactive enough that using non-optimal protein digestion conditions will result in shuffling. For example, insulin represents a poor choice of reporter due to the well-characterized cysteine-knot structure and difficulty of reduction with any standard reducing agent. Somatostatin-14 was used here to illustrate a good reporter choice with lysozyme. **Figure 3** represents an ETD spectrum of shuffling between peptides from lysozyme and somatostatin-14 under marginally non-optimal conditions. This experiment represents a positive control to show that the reporter is working properly and can be used to illustrate the prospect of shuffling because a reporter peptide that will not shuffle is useless for the purpose of this experiment.
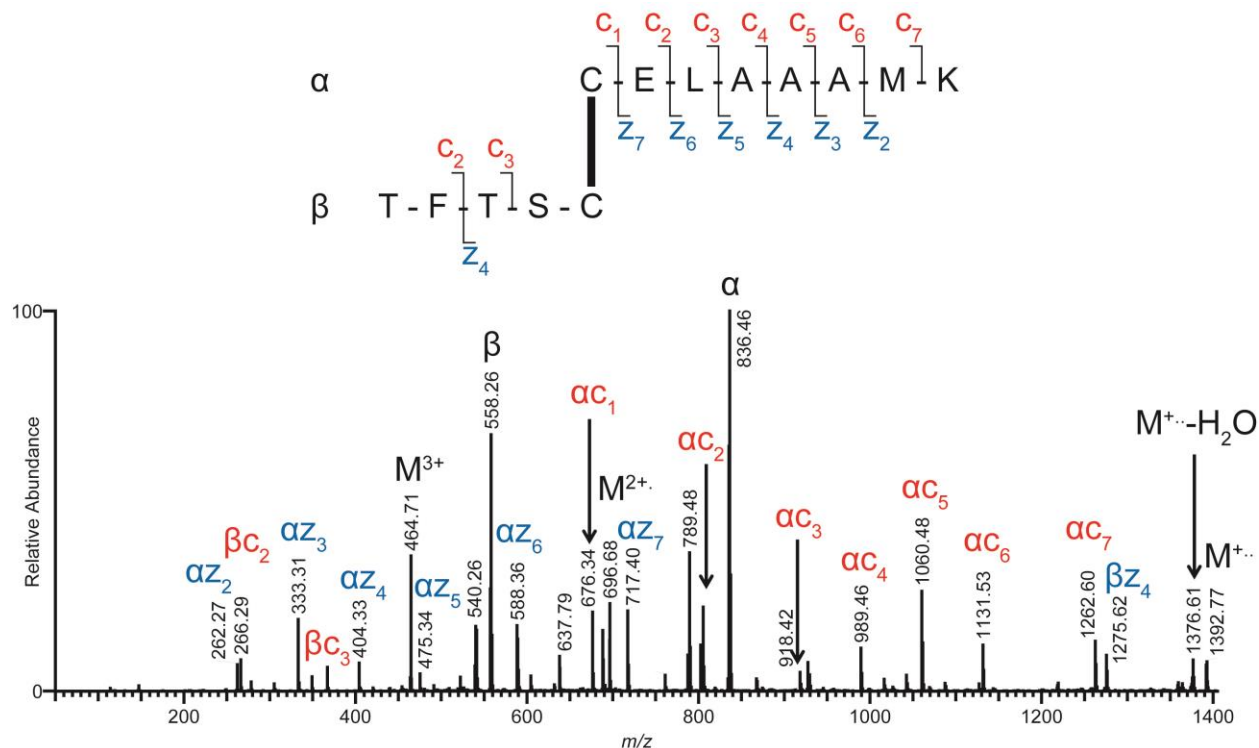
Figure 3. ETD spectrum of the detectable shuffling between a target protein (lysozyme) and the reporter (somatostatin-14) disulfides that occurred when non-optimal digestion conditions were used.

4.3.2 Non-optimal conditions

Some disulfide bonds are held tightly in place by intermolecular forces of the peptides of the two cysteine residues and are difficult to chemically reduce even with a mild denaturant present. For a proof of principle experiment non-optimal protein digestion conditions were used with lysozyme and somatostatin-14 and the digestion was monitored for shuffling. To create the non-optimal digestion conditions the pH was increased by approximately half a unit and alkylating reagent was not introduced for the duration of the experiment. **Figure 3** represents shuffling between tryptic peptides from lysozyme and somatostatin-14 with experimental conditions (time, temperature, trypsin ratio, buffer, concentration) representative of a typical trypsin digestion. Both correct

and incorrect disulfide bonds from somatostatin-14 and lysozyme peptides were detected. The shuffling between lysozyme and somatostatin-14 was detectable after only 18 hours and continues to increase both in intensity and the number of shuffling products over the timespan monitored.

4.3.3 Disulfide shuffling analysis

Before the invention of ETD, the only way to detect disulfide shuffling was to search for every possible *m/z* of every possible tryptic disulfide-bonded peptide which increases exponentially with more cysteine amino acid residues. Computer programs such as MassMatrix made this feat easier with an "exploratory" mode that would automate the search.[30] However, the approach taken here requires only an XIC of the reporter peptide *m/z* to be produced which allows the researcher to quickly investigate the possibility of shuffling before substantial time and effort is invested to assign the disulfide map. If shuffling is present, the XIC will show numerous peaks where the reporter peptide has shuffled with peptides from the target protein and these can be assigned as such. When detectable shuffling with the reporter peptide is present it is also probable that the there is also shuffling of the disulfide bonds in the target protein. This allows the researcher to test the sample preparation conditions to minimize or eliminate entirely any disulfide shuffling occurring.

4.4 **Conclusion:**

Previously we have reported aberrant disulfide connectivity in a recombinant monomer of an HIV envelope protein. The current study was performed to verify that the assignments made were representative of the protein and not merely artifacts of sample preparation. In order to achieve this goal, a known reporter disulfide bonded peptide was incorporated during the proteolytic digestion of a model target protein, lysozyme.

We have demonstrated an expedient and simple method for determining the suitability of sample preparation conditions for the analysis of disulfide bonds. The addition of a reporter disulfide bonded peptide and utilization of ETD provides an ideal combination to allow fast searching of data for possible shuffling between the target and reporter disulfides. Using optimal digestion conditions no shuffling between reporter and the model target peptides was detected whereas non-optimal conditions produced several shuffled products. Subsequent investigation into gp120 disulfides supported our previous findings and no detectable shuffling occurred with the reporter peptide. This method allows the researcher to be more certain of cases where novel or aberrant disulfides are detected and sample preparation must be eliminated as a source of artifactual disulfide bond assignments.

4.5 References:

(1) Poole, C. F. *TrAC Trend. Anal. Chem.* **2003**, *22*, 362-373.

(2) Goodwin, R. J. A. *J. Proteomics* **2012**, *75*, 4893-4911.

(3) Ramos, L. *J. Chromatogr. A* **2012**, *1221*, 84-98.

(4) Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49-79.

(5) Glocker, M. O.; Arbogast, B.; Deinzer, M. L. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 638-643.

(6) Wang, Y.; Lu, Q.; Wu, S.-L.; Karger, B. L.; Hancock, W. S. *Anal. Chem.* **2011**, *83*, 3133-3140.

(7) Wang, X.; Kumar, S.; Singh, S. K. *Pharmaceutical Research* **2011**, *28*, 3128-3144.

(8) Matsumura, M.; Becktel, W. J.; Levitt, M.; Matthews, B. W. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 6562-6566.

(9) Clarke, J.; Fersht, A. R. *Biochemistry* **1993**, *32*, 4322-4329.

(10) Pellequer, J. L.; Chen, S. W. W. *Proteins* **2006**, *65*, 192-202.

(11) Siadat, O. R.; Lougarre, A.; Lamouroux, L.; Ladurantie, C.; Fournier, D. *BMC Biochem.* **2006**, *7*, 12.

(12) Wozniak-Knopp, G.; Stadlmann, J.; Ruker, F. *PLoS One* **2012**, *7*.

(13) Wurzburg, B. A.; Kim, B.; Tarchevskaya, S. S.; Eggel, A.; Vogel, M.; Jardetzky, T. S. *J. Biol. Chem.* **2012**, *287*, 36251-36257.

(14) Fagerquist, C. K.; Sultan, O. *Analyst* **2011**, *136*, 1739-1746.

(15) Zhang, Y.; Cui, W.; Zhang, H.; Dewald, H. D.; Chen, H. *Anal. Chem.* **2012**, *84*, 3838-3842.

(16) Peng, Y.; Chen, X.; Sato, T.; Rankin, S. A.; Tsuji, R. F.; Ge, Y. *Anal. Chem.* **2012**, *84*, 3339-3346.

(17) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994-999.

(18) Han, D. K.; Eng, J.; Zhou, H. L.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 946-951.

(19) Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. *Anal. Chem.* **2003**, *75*, 1895-1904.

(20) Viner, R. I.; Zhang, T.; Second, T.; Zabrouskov, V. *J. Proteomics* **2009**, *72*, 874-885.

(21) 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. *Mol. Cell. Proteomics* **2004**, *3*, 1154-1169.

(22) Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C. *Anal. Chem.* **2008**, *80*, 2921-2931.

(23) Han, H. L.; Pappin, D. J.; Ross, P. L.; McLuckey, S. A. *J. Proteome Res.* **2008**, *7*, 3643-3648.

(24) Phanstiel, D.; Zhang, Y.; Marto, J. A.; Coon, J. J. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1255-1262.

(25) Phanstiel, D.; Unwin, R.; McAlister, G. C.; Coon, J. J. *Anal. Chem.* **2009**, *81*, 1693-1698.

(26) Dayon, L.; Pasquarello, C.; Hoogland, C.; Sanchez, J. C.; Scherl, A. *J. Proteomics* **2010**, *73*, 769-777.

(27) Chen, Z.; Wang, Q. H.; Lin, L.; Tang, Q.; Edwards, J. L.; Li, S. W.; Liu, S. Q. *Anal. Chem.* **2012**, *84*, 2908-2915.

(28) Sethuraman, M.; McComb, M. E.; Huang, H.; Huang, S. Q.; Heibeck, T.; Costello, C. E.; Cohen, R. A. *J. Proteome Res.* **2004**, *3*, 1228-1233.

(29) Clark, D. F.; Go, E. P.; Desaire, H. *Anal. Chem.* **2012**, *85*, 1192-1199.

(30) Xu, H.; Zhang, L. W.; Freitas, M. A. *J. Proteome Res.* **2008**, *7*, 138-144.

Chapter Five: Future Directions Toward a Disulfide Hunter Program

5.1 Dissertation Summary:

This dissertation describes analysis of disulfide maps in both simple model proteins and application to complex recombinant biotherapeutic proteins. Chapter two described the analysis of fragmentation product ions produced when disulfide bonded peptides were subjected to low energy collision induced dissociation. The results supported the notion that some disulfide bonded peptides produced multiple cleavage product ions and these could be used to either assign more peaks in spectra or confirm the location of disulfide bonds when more than one cysteine residue is present on a single tryptic peptide.

Chapter three detailed a method with which to assign disulfide partners using electron transfer dissociation. The method makes use of the knowledge that disulfide bonded peptides will preferentially homolytically break the disulfide bond when subjected to ETD. When the extracted ion chromatograms of each cysteine-containing peptide mass are aligned, the peaks that co-elute are assigned as disulfide partners until the entire disulfide map is assigned. Chapter four represents efforts to prove the results obtained were not artifactual from sample handling or preparation but instead indicative of the protein as it was expressed in the cell. Reporter proteins and peptides with very specific tryptic peptide masses were incorporated in the sample preparation.

5.2 Future Directions:

5.2.1 *Origins of the Disulfide Hunter*

One of the most significant limitations to performing disulfide analysis by hand is the time it takes to perform the analysis. This is particularly true when analyzing the recombinant protein gp120 as it had considerable heterogeneity when it came to the disulfide network. While I could after many years of practice often perform the disulfide analysis on model proteins within one day, a faster method was desired. The most obvious way to increase the speed of analysis is to have a computer perform the analysis. However, to fully automate the analysis a method must be reduced to simple math calculations the computer can perform quickly. In other words, the human aspect must be effectively removed from the equation. At this point two of the most popular mass spectral searching programs, Mascot and Sequest, do not offer disulfide analysis and there are a limited number of computer programs that perform disulfide analysis (refer to Chapter 2). The method set up in chapter three was purposely designed with the idea of a computer program to be later written, affectionately named *Disulfide Hunter* for my love of hunting, to expeditiously assign disulfide bonds. However, due to time constraints on the in-house programmer, this was not completed prior to graduation. *Disulfide Hunter* could allow automated disulfide assignments in minutes instead of hours or days. The general workflow required is represented by **Figure 1**. Simply, the user supplies information regarding how the sample was treated and the sequence of the protein. The data file is queried to populate a data matrix which will provide preliminary assignments and spectra will be analyzed to investigate the

possibility of extra multi-cysteine peptides. Finally, a data file is written for the user to view.
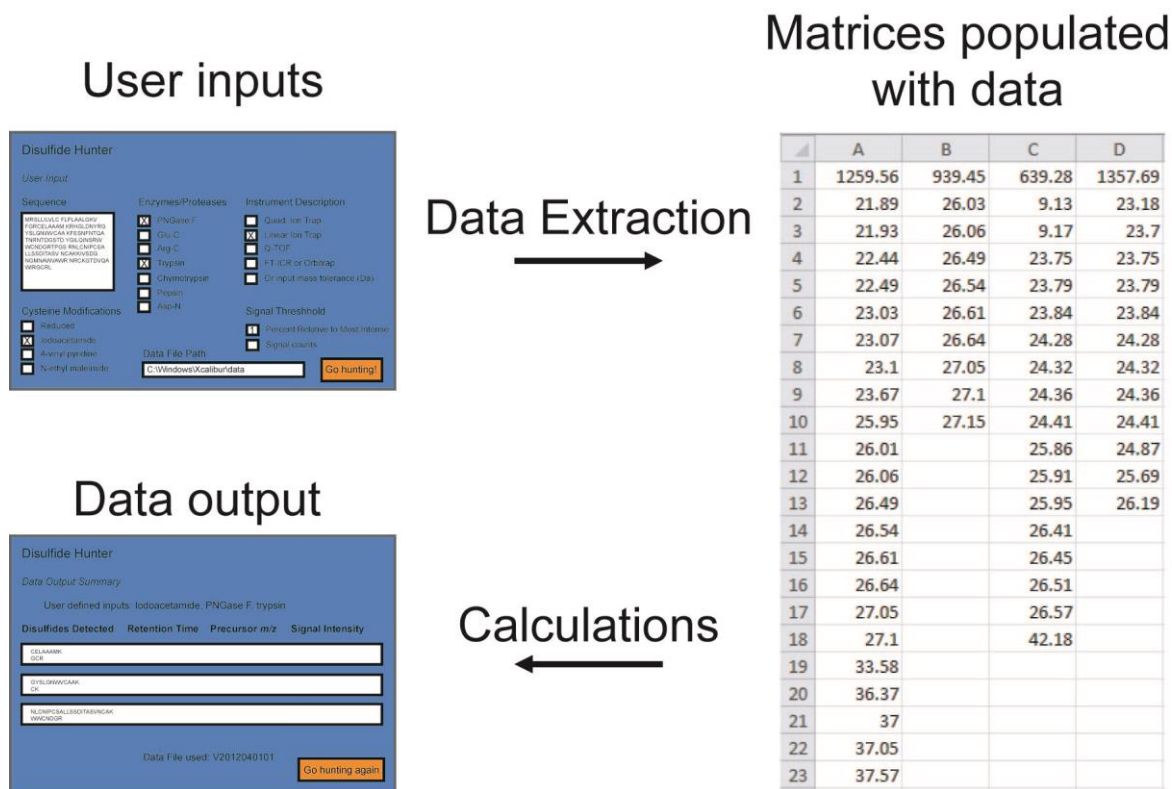
Figure 1: The workflow of *Disulfide Hunter* entails extracting data from the data file while using information provided by the user to perform calculations for the assignment of disulfide bonded peptides.

5.2.2 *User Interface and Inputs*

Disulfide Hunter would have similar inputs to most other data analysis programs (**Figure 2**). First, the sequence of the protein is required and could be provided by the user or in the form of a FASTA file or Uniprot Accession number. Separate fields including modifications such as alkylation, methionine oxidation, termini modifications, etc. are needed to specify as many experimental variables as possible. The user must also enter either a type of instrument used or specify a mass tolerance so that the data

can be analyzed appropriately. Finally, all enzymes experimentally used will also be required to perform an accurate *in silico* digest. This opens up the possibility of using multi-enzyme approaches to obtain data, such as that used in chapter 3 (PNGase F and trypsin). Similarly, other researchers have utilized multi-enzyme approaches to solve cysteine knot structures because there were not arginine and lysine residues at places that could decrease the ambiguity of disulfide mapping. The program should be arranged to collect as much information about the experiment as possible such that the data can be "intelligently" analyzed.

Figure 2: *Disulfide Hunter* inputs would include sequence, cysteine modifications, enzymes, instrument information, data source, and the user-defined signal threshhold.

An *in silico* digest is performed with the specified proteases and masses of cysteine-containing peptides would be stored as elements of the matrix. The size of the matrix would be determined from the inputs provided by the user. Each matrix element mass would be converted to theoretical *m/z* values with all expected modifications present and represents the *m/z* where the peptide would be expected when ETD is performed. Some limits on these values must also be in place. For instance on our LTQ Velos linear ion trap instrument, when peptides are between 100 and 1000 Daltons, only a 1+ charge will be represented. Peptides between 1000 and 2000 Daltons should have values for both 1+ and 2+ charge states. Once the peptides are above 2000 Daltons the 1+ charge state can be eliminated from consideration as it is not in the scan range and instead 2+ and 3+ charge states should be searched. Peptides with more than one cysteine need to be given special designation. This special designation is in place because multiple disulfides have not been shown to cleave during ETD. Instead *Disulfide Hunter* will treat each multi-cysteine peptide in two ways. First it will calculate *m/z* values in conjunction to the rules stated above as possibly connected to each of the single cysteine peptides. This will allow detection of tripeptides, tetrapeptides, and so on. Second, *Disulfide Hunter* will calculate each multi-cysteine peptide *m/z* as if they are bonded to themselves as well. This aids the detection of intrapeptide disulfide bond as they are quite common in the analyses that have been performed here.

Another important input needed from the user is the kind of mass tolerance that is to be used. This has a two-fold purpose: to account for possible minor mass modifications from sample preparation or the ETD process and to incorporate the mass

error of the instrument. For instance, when PNGase F is used, the mass tolerance needs to be automatically increased to account for the mass difference between non-glycosylated peptides (with asparagine still intact) and the deglycosylated peptides (deamidated to form aspartate). Based on glycosylation site coverage, both forms may or may not be detected simultaneously. Isotope abundance is a fundamental mass spectrometry concept but performing ETD creates complications when processes such as neutral hydrogen ejection and multiple electrons neutralize multiple protons. The former is often noted in ETD of peptides. **Figure 3**A represents an experimentally-obtained MS[1] spectrum of peptide HSTIFENLANK from human transferrin. The inset shows the expected isotopic profile, generated by IsoPro 3.0, and the two closely resemble each other as expected. **Figure 3**B is an experimentally-obtained MS[2] spectrum when the 2+ charge state is subjected to ETD and the charge reduced precursor is viewed. The inset is again the expected isotopic profile based solely on natural abundance. The most obvious difference between the two is the relative abundances of the isotopes. The ion subjected to ETD is represented by $(M+2H)^{2+}$ where M is the mass of the peptide and two protons are responsible for the 2+ charge state of the peptide. The *m/z* that was chosen for ETD was 637.44 and when one electron is transferred to it, the absolute minimum *m/z* should be $(M+2H)^{+\cdot}$ or 1274.88. While 1274.72 is the most intense peak, there is also a "forbidden" peak at 1273.72 which corresponds to a neutral hydrogen ejection. This process complicates the mass tolerance value of *Disulfide Hunter* because now what appears to be the monoisotopic peak is actually a neutral hydrogen atom ejection and what appears to be M+1 peak is a

combination of peptides that did not eject a neutral hydrogen and isotopes from peptides that did.
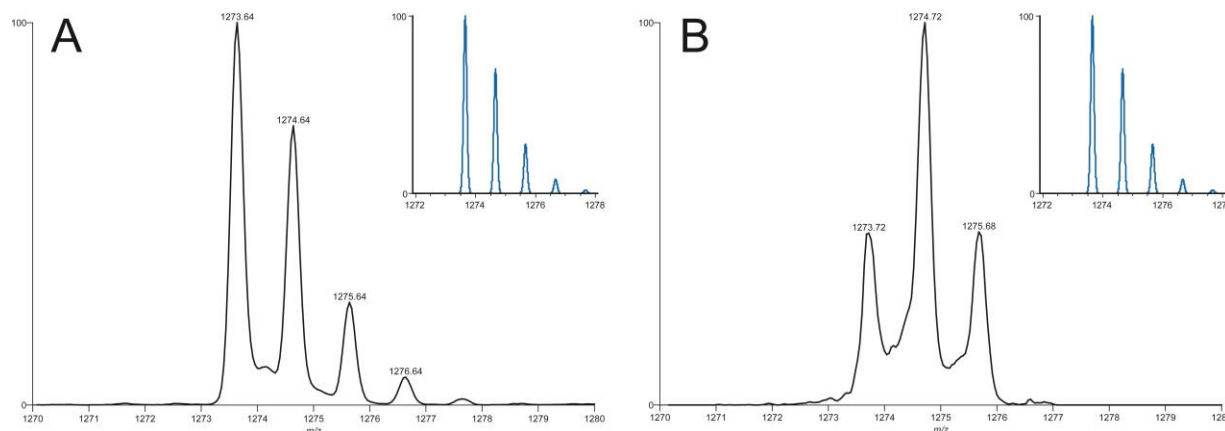

Figure 3: Spectra of peptide HSTIFENLANK from human transferrin. 3A is the MS1 spectrum and 3B is the resulting charge reduced precursor from performing ETD on the 2+ charge state. The differing relative abundances detected are due to ejection of a neutral hydrogen atom.

Disulfide ETD also presents another layer of complexity and can obscure an expected value for the cysteine-containing peptides. Disulfide bonded peptides often occur in higher charge states, which is advantageous for ETD because at least one charge will be neutralized upon the addition of an electron and ETD has a "sweet spot" from 500 to 1000 *m/z* where it is more efficient. Multiple charging of very large disulfides will bring them into this range and efficient ETD can take place. However, disulfides are also unique because they preferentially undergo homolytic disulfide scission. When this is applied to a population of ions as we see in a mass spectrometer, charge locations will not be identical on all species. **Figure 4** is an example of a disulfide-bonded peptide from transferrin that was subjected to ETD in the 4+ charge state and the expected isotope pattern, calculated with IsoPro 3.0, is shown in each inset.
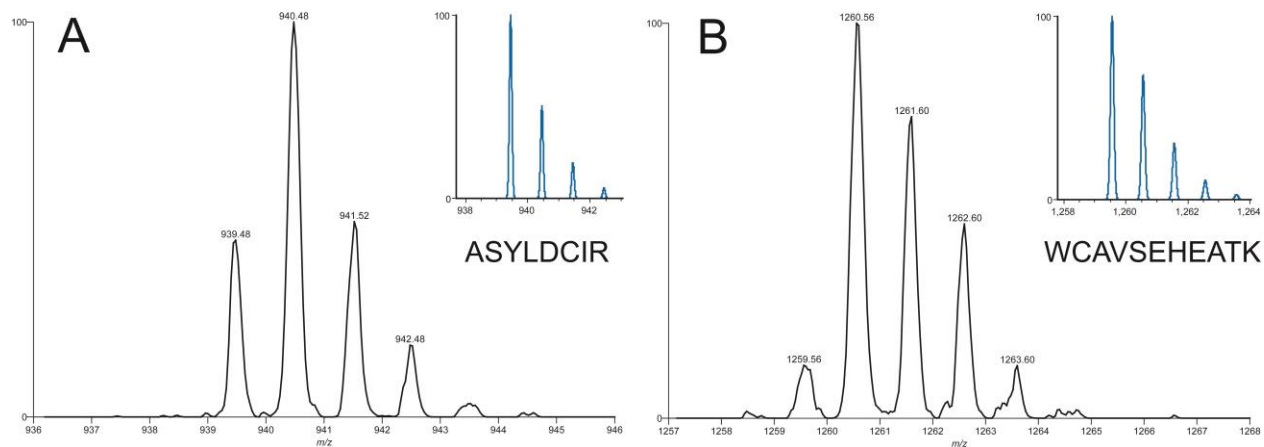
Figure 4: A disulfide-bonded peptide from transferrin was subjected to ETD and the peaks resulting from the two peptides, ASYLDCIR and WCAVSEHEATK, A and B respectively, exhibit a complex spectral pattern.

The *m/z* selected for ETD that resulted in these was 550.51 which represents $(M+4H)^{4+}$. The breaking of the disulfide leads to a vital question of where the protons are located. In both of these cases, the first peak is a combination of peptides that had only one proton located on it during fragmentation and peptides that had two protons but one was neutralized via addition of an electron and ejected (more complex cases are possible). The second peak in both spectra represents the combination of the M+1 isotope from the first peak, peptides that had only two protons during fragmentation and one was neutralized but not ejected, and peptides that had three protons and neutralized two protons while ejecting a neutral hydrogen atom after an electron was transferred. The trend continues with the other peaks. The complexity is problematic for mass tolerance application because if the 1+ charge state *m/z* is searched for a peptide but the most intense peak is the result of two protons with one neutralized as was the case in both **Figure 4**A and 4B, the assignment may not be made. Add these to the fact that

PNGase F introduces a 0.984 Da mass shift and there is a conundrum such that mass tolerance could either make or break the analysis. *Disulfide Hunter* should attempt to take as much of this as possible into account and the mass tolerance of data from low resolution instruments may need to be set to a higher value than one would expect to account for these problems.

The final input needed by *Disulfide Hunter* from the user is the data file. Portions of the data will be extracted and placed in a multi-dimensional matrix as shown in **Figure 5**. The algorithm to extract the data from each data file will require the most development because each manufacturer uses a different file type. We have also found that manufacturers are less than accommodating when attempting to mine the data from data files produced by their software.
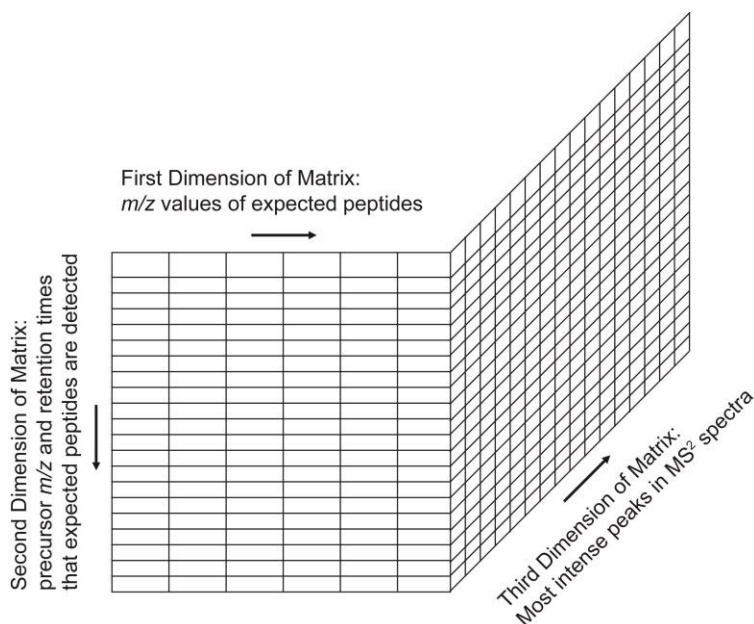
Figure 5: Representation of the data matrix in *Disulfide Hunter*.

However, there is the possibility of using the manufacturer-specific software and converting each data file to a CSV type file, which is a more standard spreadsheet-type file, and using that as an input. This constitutes an extra step and we would like to not only eliminate it to expedite the analysis but also set up *Disulfide Hunter* to perform analysis on a batch of files instead of one file at a time. In either case the overall data extraction is very similar. A matrix must be made in *Disulfide Hunter* that will contain the data. The matrix elements are then populated when the MS^2 ETD spectra are queried for the *m/z* values specified in the first dimension of the matrix and information about that spectrum is recorded. Specifically, the retention times and precursor *m/z* values are recorded in the second matrix dimension and spectral information is recorded in the third matrix dimension. In order to reduce computing requirements, the data recorded from each spectrum will consist of the retention time, precursor *m/z*, and the 25 most intense peaks. If the *m/z* that resulted in the spectrum being labeled as a "hit" is not among the top 25 peaks, the spectrum is discarded as it was probably due to low level background. *Disulfide Hunter* will proceed with data collection until all *m/z* values from the matrix are exhausted.

Searched *m/z* values

Retention times of spectra where that *m/z* was detected

Two pairs of peptides are co-eluting and have been given a preliminary disulfide pair assignment. The spectra will be checked by calculations for more partner peptides before the disulfide assignment can be written to the output file.

Figure 6: Values written to the *Disulfide Hunter* data matrix are extracted from the data file. These include the m/z searched, retention times, and spectral m/z values (not shown for simplicity).

It should be noted that the described filtering of the data is required to reach this point. Due to advances in experimentation many thousands of spectra are collected in each LC-MS experiment. Attempting to analyze every single spectrum would require an unrealistic amount of time and computational resources. For this reason the spectra were filtered from the mother liquor of data by searching for the *m/z* values that are expected and only retrieving the top 25 peaks in each spectrum. With current computing power, this represents an amount of data that could be handled by a workstation computer. However, this could be an area to improve in the future and could even be a user-defined field with a suggestion of 25 *m/z* values.

5.2.3 Hunter Analysis

Once *Disulfide Hunter* receives all the necessary input from the user and the data file, the analysis can begin. **Figure 7** provides a good visual interpretation of what the user sees versus what *Disulfide Hunter* is actually doing. Moments after the user clicks "Go hunting!" *Disulfide Hunter* has populated the matrix with values consisting of the retention time, precursor ion *m/z*, and *m/z* values of the top 25 peaks in each ETD spectrum that an expected *m/z* was detected in. Each spectrum is then analyzed for charge reduced precursors in order to verify the charge state of the precursor. The precursor *m/z* is used to calculate possible values for charge reduced precursors which are checked against the top 25 *m/z* values. For low resolution data precursor charge states from 2+ to 10+ will be investigated. Because there will be overlap between some of the charge states, all must be investigated and verified later with the disulfide assigned. There are two possible outcomes at this point: assignment of a charge state to each precursor or non-assignment. In the case of the latter, it should throw a "charge state undefined" error with the program output so the user can either look at the spectrum manually or discard it at their choosing. Spectra with undefined charge states should comprise a minimal portion of the data if prior analysis is any indication.

Figure 7: *Disulfide Hunter* is illustrated and what the user sees is compared to what is actually occurring.

With charge states of each precursor defined *Disulfide Hunter* can search for and assign free cysteine residues as well as disulfide bonding partner peptides. Spectra of iodoacetamide-alkylated free cysteine residues are detected and listed in the matrix because neutral loss of the carbamidomethylation moiety represents the most intense resulting peak from these species during ETD and represents a peak at the *m/z* expected for the cysteine-containing peptide. Comparing the precursor mass to the mass of the cysteine-containing peptide will show a 57 Dalton mass difference and this is then written to the output file as an alkylated cysteine. Disulfide bonding peptides will require more, but similar, arithmetic. *Disulfide Hunter* will assign disulfide bonding partner peptides by comparing retention time or scan numbers in the matrix: any spectra that have the same retention time represent a case that could potentially be a disulfide-bonded peptide between those two peptides. The assignment will be scrutinized via purely mathematical means. The mass of the precursor is calculated because precursor *m/z* and charge state have both been identified. Since homolytic disulfide scission has been previously characterized to produce the most intense peaks in ETD spectra of disulfide bonded peptides, each disulfide assignment is then checked for proper mass by adding the masses of the individual peptides. If the masses of the peptides sum to the correct precursor mass, a disulfide assignment is made and written to the output file. If the masses do not concur, then further analysis is required. A third peptide that contains two cysteine residues may be the culprit and in this case typically also produces peaks representative of the two dipeptides formed when only one single disulfide is cleaved. The analysis continues until an assignment can be made or all

options are exhausted, at which point the spectrum is assigned as an unknown/interfering species and written to the output file.

The output file will be a simple text file summary of all assignments applied to each cysteine-containing peptide. It should be separated by each peptide and contain information such as the scan number and the assignment, where "free cysteine", "disulfide partner with peptide(s) X", or "unknown/interfering species" are the possible assignments. When organized by peptide it will be easy for the user to view all assignments and double check any scan numbers that may have been assigned as an unknown/interfering species. One advantage to performing this type of analysis is that multiple disulfide bonding configurations can be detected and viewed in the output file.

Figure 8: A representation of what the output of *Disulfide Hunter* could potentially look like. It would include assignments and other information the user could use to find necessary data to compile figures.

5.2.4 *The Potential of Disulfide Hunter*

This is no way represents the entirety of what *Disulfide Hunter* could be and in fact includes only a minor version of capabilities that are possible. Debugging and more development with previously assigned data sets will need to be performed before any serious effort is made to apply it to unknown data sets. There exist several more areas for improvement once the basics of *Disulfide Hunter* are implemented. Most obvious is the possibility of incorporation with existing programs. *Disulfide Hunter* could provide current programs such as Mascot and Sequest a way to quickly assigning disulfide maps to samples. This would be extremely beneficial because Mascot and Sequest will identify the proteins that are present in samples and *Disulfide Hunter* can use this information to assign disulfide bonds present in those proteins.

There are also several other improvements that would include the incorporation of more data from the data file. The intensity of each peak and the relative intensities could be used to help determine precursor charge state, as has been done previously. The intensity of each ion representative of a cysteine-containing peptide could also be used to apply either a statistical scoring scheme or used to further filter out low background from the list of potential assignments. This highlights another possible area of improvement: implication of a probability-based scoring scheme. For each disulfide assignment there could be set up a series of scoring criteria such as: intensity of the peptide peaks resulting from homolytic disulfide scission, intensity of precursor ions, and other fragment ions. Intensity of the precursor ion is important because there is the possibility that a small portion of the population of proteins has slightly different disulfide bonds and is detectable. While this will still be reported, it could be given a score to

reflect the decreased signal. There also lies the possibility of using either background c and z ions that form during ETD or correlating the peak in question with CID data to assign more ions. While disulfide scission represents the most probable form of fragmentation of disulfide-bonded peptides during ETD, there are still c and z ions that form from peptide backbone fragmentation. These can be used to corroborate or disprove the assignment that was made. For an orthogonal approach at fragmentation, CID uses a different mechanism and does not break the disulfide but instead produces b and y ions from the peptide backbone. If the instrument method is set up to perform both CID and ETD on a precursor then the CID data can be treated similarly to the background c and z ions to either corroborate the assignment made or provide evidence to the contrary. CID could be included in the instrumental method as one of the scan types but the *Disulfide Hunter* would need to be able to sort MS^2 data from ETD and CID differently.

The last area for improvement would be the output. Currently the simplest output is a text file of all the assignments shown in **Figure 8**. However, this could be reworked to include visualization of either how the disulfides will influence protein tertiary structure or be given weighting based on intensity of the precursor ion. This also leads to the notion that the entire disulfide map could be represented in a figure or series of figures and provides insight into the disulfide bonding status of each cysteine residue (**Figure 9**).

Figure 9: A representation of a potential output from Disulfide Hunter. Several different disulfide bonding configurations were detected for the same portion of protein and are shown with surrounding peptides gleaned from the protein sequence input.

5.3 Conclusion:

Disulfide Hunter is the rough sketch of a software program that would help automate

assignments using the method outlined in Chapter 3. Using a program such as

Disulfide Hunter would reduce the variability due to the human component of data

analysis. After time the program could even output images of how the disulfide

assignments look according to the primary sequence of the protein.

FIN.