OPTIMIZATION OF DISULFIDE MAPPING USING MASS SPECTROMETRY

by

NOZOMI MATSUMIYA

B.S., Kansas State University, 2007

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Biochemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Approved by:

Major Professor
Dr. John Tomich

Abstract

One of the important keys to characterize the biological function of a protein is the study of post-translational modification (PTM). Formation of disulfide bond linkages between cysteine residues within a protein is a common PTM which not only contributes to folding and stabilizing the protein structure, but also to accomplishing its native function. Therefore, the study and discovery of structural-functional relationships of expressed proteins using an isolated proteomics approach has been one of the biggest advances within the field of structural biology in recent years. In this study, rapid disulfide bond mapping of freshly obtained equine serum albumin (ESA) was performed using matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS). Highly sensitive MALDI-TOF MS is commonly used for the investigation of disulfide bond linkages in the proteomics field. However, it has also been known that the presence of disulfide bond linkages absorbs the energy which is created by the cysteine-cysteine kinetic vibration, resulting in a decrease of the instrumental sensitivity. To overcome this problem, the disulfide bond mapping method was optimized by applying a combination of chemical labeling, proteolytic enzymes, and matrices. With the optimized method, we were also able to achieve high protein sequence coverage. Obtaining higher sequence coverage of a protein provides more information about a protein which helps to identify the protein by peptide mass fingerprint (PMF) technique. These analyses eventually contribute to the estimation of the possible PTM sites.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest appreciation to my supervisors, Dr. Takeo Iwamoto and Dr. John Tomich, for their inexhaustible guidance, support and considerate advice which has always led me to learn new things. I have learned so much from both of them from the time I joined this lab (4 years ago) as an undergraduate research assistant. I cannot thank them enough for supporting me for such a long period.

I extend my gratitude to Dr. Lawrence Davis and Dr. Om Prakash for agreeing to be my committee members. Their candid and sincere views and suggestions have been instrumental to improving the quality of my work. I also would like to thank Dr. James Lillich in the K-State Veterinary Medicine for providing the fresh sample for this study.

My special thanks go to Dr. Yasuaki Hiromasa, our research associate, for extensive technical assistance. I acquired innumerable useful research techniques working with him. Sushanth Gudlur, my respected colleague and great friend since we simultaneously began working in the same lab, also deserves merit for helping me with the spectrophotometer. I also thank Satyabrata Das, my colleague from Dr. Takemoto's lab, for assistance with the densitometer.

It has been my pleasure to work with wonderful colleagues: Pinakin Sukthankar, Urska Bukovnik, Xiao Yao, Matt Warner and all the other present and past members of "the Tomich Lab." I will never forget the time we spent and the memories we have made together.

I am also thankful to have such great family and friends, especially my brother Hiroki Matsumiya, my sister Midori Matsumiya, my fiancé Nick Caton, and my best friends, Maki Kashiwaya and Ethan Primm for their continuous love and support, without which all of my work would not be possible.

Dedication

This thesis is dedicated to my parents, Masahiro Matsumiya and Kayoko Matsumiya who gave me the opportunity to study in this university. Their endless support and love kept me encouraged and inspired all the time. None of this could have happened without having such a great parents who raised me to be the person I am today.

CHAPTER 1 - Introduction

Post-Translational Modification

To better understand structural-function relationships in proteins, studying their post-translational modifications (PTM) is essential. A variety of PTMs have been discovered and studied, including attachment of phosphates, lipids and sugars. Due to the presence of these additional appended groups, proteins can be altered with respect to stability or activity. As of 2006, at least 350 different types of PTMs had been reported.[1] Due to the extensive progress in PTM research in recent years, the number of distinct PTMs discovered has been increasing. Even though many PTMs have been discovered, characterization of many of them has proven to be a difficult task. Identification and characterization of a PTM can be difficult because such identification is labor-intensive, the database for the genome can be incomplete and there are no standardized methodologies.[1] Various clinical research groups produced protein-based biopharmaceuticals involving a broad range of PTMs, with many of these having at least one, or more likely a combination of two or more PTMs.

Glycosylation is one of the most common PTMs along with carboxylation and hydroxylation. It has been reported that nearly 50 % of human proteins are glycosylated.[2-3] Glycosylation includes the attachment of mono, oligo, and extended polysaccharide chains. Glycosylation is reported to help proteins fold correctly and therefore to be stabilized better with often enhanced solubility.[2] Besides glycosylation, one of the other significant PTMs is formation of disulfide bonds.

Formation of Disulfide Bonds

The formation of disulfide bond linkages contributes to the stabilization of the protein's tertiary structure. This makes proteins which have larger numbers of amino acid residues between two bond forming cysteines more stable. The location of cysteines within a protein greatly affects the conformation of natively folded protein. Besides stabilizing the structure of protein, disulfide bonds also influence activation or regulation

of potential function of protein by forming correct or incorrect bondings.[3-5] In order to form a disulfide bond, two cysteine residues in a protein must be co-oxidized. Cysteine is the only amino acid among 20 major encoded amino acids that contains a reactive sulfur atom. This sulfhydryl group can function in many ways. It generally exists by forming native disulfide bond with other present cysteines in the lumen of the endoplasmic reticulum to maintain its native bioactivity or stay as free (reduced) cysteine which also functions as a regulator or a substance carrier in many given biological systems.[4-8] In a review article by Mamathambika,[9] four major factors are emphasized; concentration of thiolate anion ($-S^-$), accessibility, proximity, and reactivity of the thiolate anion or bond linkage can all affect formation of disulfide bonds. Also, when the native disulfide bond and a sulfhydryl group of free cysteine come closer, disulfide bond exchange can occur between already formed disulfide bond and a free cysteine as it is shown below.

$$R_1\text{-SH} + R_2\text{S-S}R_2 \leftrightharpoons R_1\text{S-S}R_2 + R_2\text{-SH}$$ [10]

The equilibrium constant for the disulfide bond exchange can be varied by the stability of each bond. Because sulfhydryl groups usually stay in the reduced form within a cell, new disulfide bond formation is more likely to occur extracellularly where the pH is higher by breaking original bond with possible presence of reducing agent such as glutathione. The pKa of cysteine sulfhydryl group normally is 8.37, therefore reoxidation to form new bond most likely happens at higher pH values.[9]

$$\text{R-SH} \leftrightharpoons \text{R-}S^- \cdots H^+ \leftrightharpoons \text{R-}S^-$$

Under certain condition, pKa value of cysteine can be shifted depending on the surrounding environment within a protein. For example, Keap1 protein, the repressor of Nrf2 which acts in the role of transcriptional activation of phase 2 genes, could have a lower pKa values than usual. Nrf2 is normally bound to the cytoplasm through Keap1 which is anchored to the cytoskeleton. These two form a complex and regulate phase 2 gene transcription. Murine Keap1 is 624 amino acid long protein and contains 25 cysteines in which 4 cysteines are identified to be most reactive. When these 4 cysteines which are located on the binding domain are modified, Nrf2/Keap1 complex is disrupted. This disruption causes Nrf2 to be dissociated from Keap1 and moved to nucleus where it accelerates the transcription. Because these 4 cysteines are located adjacent to the basic

residues which have higher side chain pKa values, pKa of cysteines would be lower than usual. It is not uncommon to see the case where cysteine pKa values are shifted due to the change in its surrounding conditions.[11-12]

Formation of disulfide bonds is certainly necessary for plants as well. Disulfide bond formation can reversibly be regulated by signaling element which controls gene expression level of chloroplast in many plant's metabolic pathways.[13] It is also known that some protein folding mechanisms first require disulfide isomerase and a formation of non-native disulfide bonding in order to form correct disulfide bonds later.[9, 14] If a protein was not folded correctly, the protein can lose its original function and become aggregated or degraded by proteases. Proteins containing multiple or odd numbers of cysteines are still capable of forming correct disulfide bonds and this fact indicates high specificity of disulfide bond formation. The bovine cathepsin C (CC), a lysosomal cysteine protease, is a tetramer, made of four identical monomers, and contains 13 cysteines per each monomer. Out of 13 cysteines, 10 cysteines were confirmed to be participating in the formation of disulfide bonds while the remaining 3 cysteines were identified to be free. Among those three cysteines, one is classified to be an inaccessible buried residue while another cysteine is a catalytic residue located in the active site. One last free cysteine becomes exposed during the activation process and can cause dissociation of the tetramer into dimers which indicates that this free cysteine significantly contributes to the maintenance of the original structure.[15] Identifying existing disulfide bonds in the protein leads to understanding of functional mechanism of the protein.[16] Disulfide bond formation is necessary for a protein to perform its function. However, determination of the disulfide bond can be challenging due to the presence of unstable intermediates and other technical concerns which discussed later.[9, 16-17]

Disulfide Bond Assignment

Determination of the structure of a protein in its native state gives us many clues to understand the functional properties of the protein. It is often difficult to perform a full structural analysis of a protein at its native state due to larger molecular weights and numerous PTMs.[18] NMR (nuclear magnetic resonance spectroscopy), X-ray Crystallography, and Edman Degradation have essentially been used to study structural

aspects of protein. These methods are used to analyze and identify disulfide bonds. However, these methods can be limited by several factors. Sample concentration is one aspect that limits an analytical method. It will also depend on protein's purity and sensitivity of the instrument. NMR usually requires highly purified sample, greater than 90 % purity, in a milligram scale. Likewise, x-ray crystallography will most likely require equal or higher amounts of purified samples for crystallizations. Purifying samples can be challenging if the sample is not abundant or is difficult to extract. Furthermore, x-ray crystallography uses proximity of neighboring cysteines to estimate disulfide bond linkages rather than actually measuring real disulfide bonds. Also, there is always extra time needed for the data processing steps. After raw experimental data are collected, both NMR and x-ray crystallography require data conversion or processing by computer to actually obtain analytically comparable data. Edman Degradation can be useful if sample is a peptide, generally less than 5000 in molecular weight. The reaction efficiency of each degradation step is relatively high which leads to prevention of sample loss. However, this method still requires a large quantity of chemical reagents and time throughout the experiment.[8,19] Additionally, it can only analyze one pure peptide at a time and sample must have a free N-terminal to initiate degradation reaction.

Mass Spectrometry Analysis

Mass Spectrometry (MS) has been contributing to the wide range of research fields especially in the proteomics research in the past few decades. Recently, it has often been used to obtain a variety of information such as sequencing of protein and investigation of protein's original structural characteristics including the assignment of disulfide bonding.[3-4,8,14,16,20-26] MS is highly sensitive, regularly using only pico or femto mol amounts which is much less than that required by the other analytical instruments discussed above. Because amount of sample required for MS analysis is relatively small, it can be processed from either a solution or a gel. MS can analyze sample at the native state since it does not require any crystallization or treatment with organic or inorganic solutions.

Electrospray ionization mass spectrometry (ESI-MS) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) are the two

most commonly used mass spectrometry instruments for the determination of disulfide bonds and identification of related PTMs of a protein.[1-4,23-24] Another significant point to mention is that by using MS, we can expect relatively high coverage of the protein sequence. High yield protein sequence coverage is an essential key to the characterization of a protein because it provides information across the full length of the sequence. This additional information will lead to the complete assignment of the PTM sites leading to understanding of protein regulation mechanisms. Now, researching biomarkers of different diseases is very much demanded from the society, detection of any kind of sequence modification such as deletion or truncation can give clues to the research goal.[20] Sequence coverage sometimes helps ascertaining the amount of expressed protein as well. Obtaining high yield sequence coverage depends on many small factors for instance, amount of protein, sample preparation and matrix adjustments.[23] Chemical modification followed by enzymatic digestion of a protein is the basic proteomics approach for detecting disulfide bond linkages using MS.[4,14,26] MALDI-TOF MS particularly is useful to detect disulfide bonds because of its high sensitivity and relatively simple methodology. However, it is well known that the sample containing multiple cysteines or disulfide bonds is hard to be detected causing the energy absorbance or kinetic vibration caused by the multiple cysteines. Also any salts or contaminants including non-cysteine-containing digested fragments lead to poor ionization efficiency of the sample. Due to these factors, only 40-50 % is the range of protein sequence coverage by MALDI-TOF MS. When it is combined with CE (capillary electrophoresis)/ESI-TOF MS, sequence coverage can be up to 90 % as it is reported.[20,23] Overall, adjustment of MALDI matrices as well as careful sample preparation would greatly contribute to the collection of reliable data and high yield sequence coverage.

Serum Albumin

Serum albumin is the most abundant protein in the blood plasma of vertebrates, up to 50 % of all protein.[27] It is generally rich in cysteines and the molecular weight is also relatively large, 67 kDa for human serum albumin. See Figures 1-1and 1-2 for 3D structure and primary sequence.[29-30] Serum albumin is quite hydrophilic and this hydrophilicity contributes to keeping up the oncotic (colloid osmotic) pressure. For

example, when a large amount of protein was suddenly lost or body was lacking in enough nutrition, the protein level can suddenly be dropped which causes dramatic decrease in blood pressure levels. This condition accumulates fluids in surrounding tissues which eventually causes our body to get swollen, known as edema. Besides maintaining oncotic pressure, the free cysteine can specifically transport molecules or substances such as ions or variety of drugs to the necessary area. Because of these function, serum albumin is often called "multifunctional plasma transport protein." [27] Equine Serum Albumin (ESA) used in this study, contains 607 residues and 35 cysteines. See Figures 1-3 and 1-4 for primary and secondary structure of ESA. [28, 31] Furthermore, the ESA sequence is 76.1 % and 73.9 % identical to the human serum albumin[30] and the bovine serum albumin[32] respectively (see Figure 1-5 for comparison).[28] This high sequence similarity also applies to the position and the number of disulfide bonds reported. Both human and bovine serum albumins contain 17 disulfide bonds in similar sequence positions. The three dimensional structure of ESA determined using x-ray crystallography with a resolution of 0.27 nm was studied and positions of 17 disulfide bonds and 1 free cysteine site were predicted.[28] It was also known that ESA contains 3 phosphorylation sites. Being very abundant and a cysteine rich protein, ESA was an excellent candidate for this study.

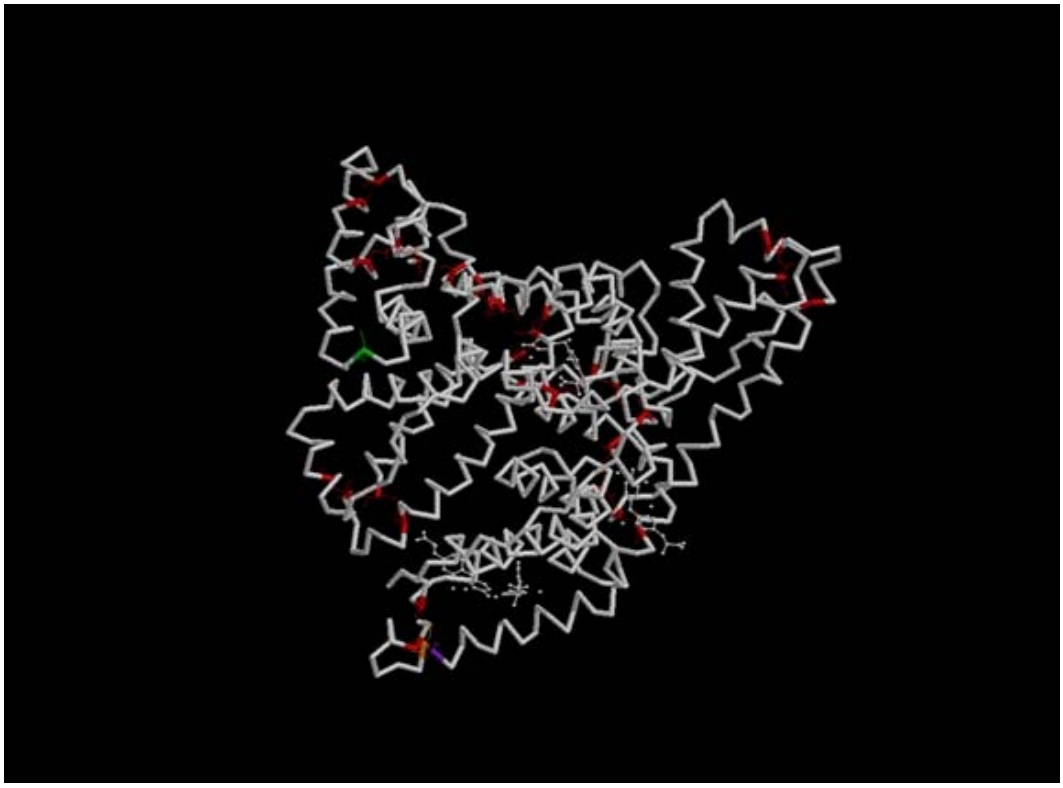

Figure 1-1 3D structure of Human Serum Albumin by X-ray Crystallography Residue 25-609(complexed with thyroxine)
Red=Cysteines, Green=Cys58 (reported as a free cysteine), Orange=Cys582, Purple=Cys583

```
         10         20         30         40         50         60
MKWVTFISLL FLFSSAYSRG VFRRDAHKSE VAHRFKDLGE ENFKALVLIA FAQYLQQCPF

         70         80         90        100        110        120
EDHVKLVNEV TEFAKTCVAD ESAENCDKSL HTLFGDKLCT VATLRETYGE MADCCAKQEP

        130        140        150        160        170        180
ERNECFLQHK DDNPNLPRLV RPEVDVMCTA FHDNEETFLK KYLYEIARRH PYFYAPELLF

        190        200        210        220        230        240
FAKRYKAAFT ECCQAADKAA CLLPKLDELR DEGKASSAKQ RLKCASLQKF GERAFKAWAV

        250        260        270        280        290        300
ARLSQRFPKA EFAEVSKLVT DLTKVHTECC HGDLLECADD RADLAKYICE NQDSISSKLK

        310        320        330        340        350        360
ECCEKPLLEK SHCIAEVEND EMPADLPSLA ADFVGSKDVC KNYAEAKDVF LGMFLYEYAR

        370        380        390        400        410        420
RHPDYSVVLL LRLAKTYETT LEKCCAAADP HECYAKVFDE FKPLVEEPQN LIKQNCELFE

        430        440        450        460        470        480
QLGEYKFQNA LLVRYTKKVP QVSTPTLVEV SRNLGKVGSK CCKHPEAKRM PCAEDCLSVF

        490        500        510        520        530        540
LNQLCVLHEK TPVSDRVTKC CTESLVNGRP CFSALEVDET YVPKEFNAET FTFHADICTL

        550        560        570        580        590        600
SEKERQIKKQ TALVELVKHK PKATKEQLKA VMDDFAAFVE KCCKADDKET CFAEEGKKLV


AASQAALGL
```

Figure 1-2 Primary sequence of Human Serum Albumin

Figure 1-3 Primary structure of ESA

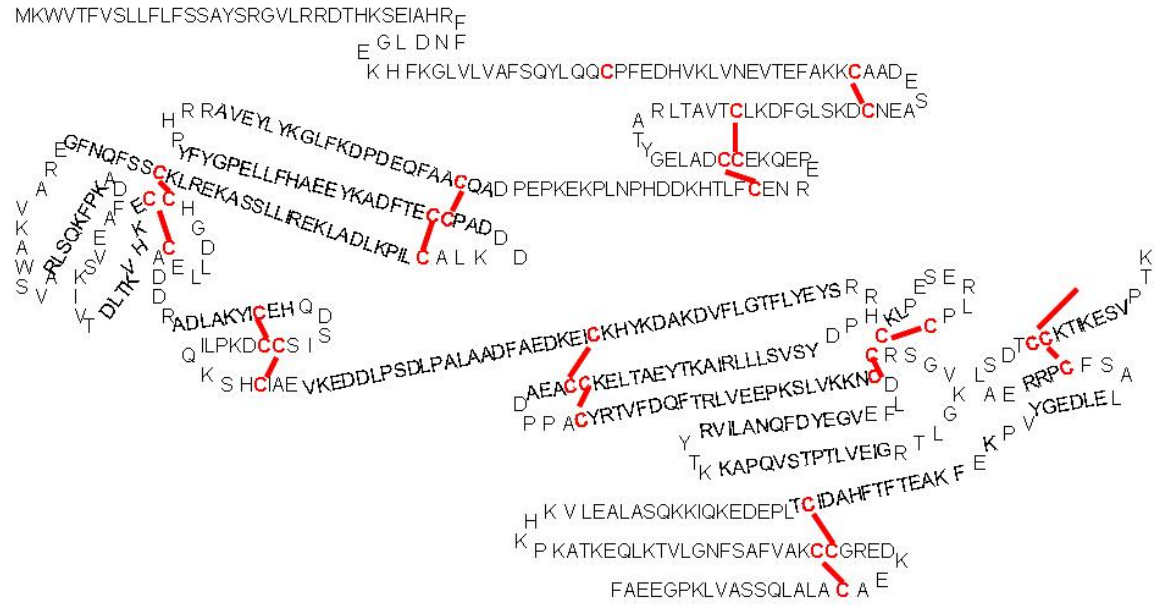

Figure 1-4 Secondary structure of ESA

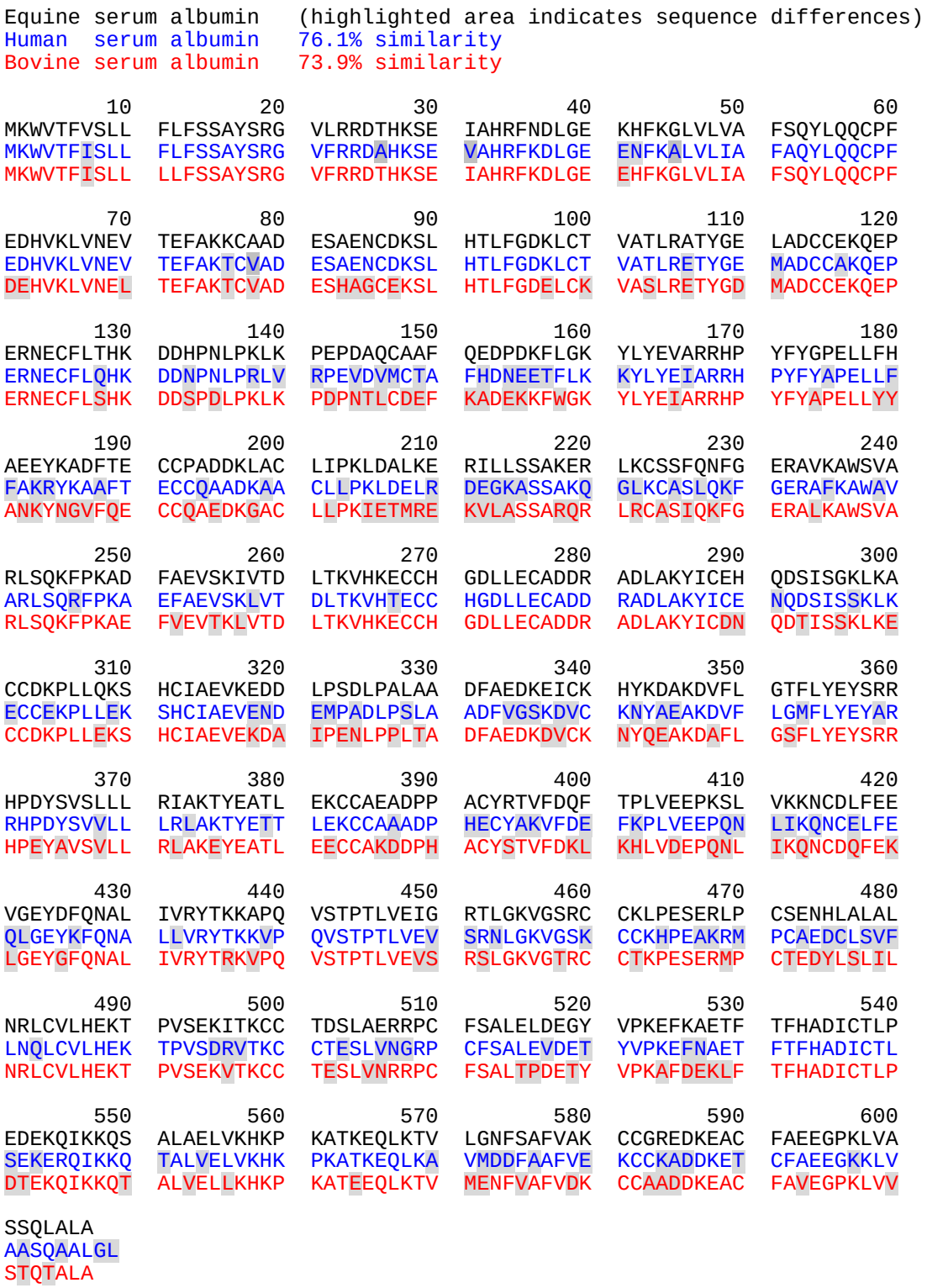

Figure 1-5 Sequence comparison between Equine (black), Human (blue), and Bovine (red) serum albumin

Differences between three sequences are highlighted in grey

Research Outline

In this study, several kinds of in-gel chemical modification and enzyme digestion were performed, followed by MALDI-TOF MS to analyze the disulfide bond combinations present in fresh ESA. In order to improve quick and easy disulfide bond mapping methodologies and also to achieve high yield protein sequence coverage using MS, we optimized the protocol by chemically labeling the cysteines and also used a new type of matrix. Nearly 90 % of sequence coverage was obtained by single enzyme digestion after reduction and carboxymethylation of sample while two enzyme combined data showed up to 92 % sequence coverage. The possibility of other PTMs such as glycosylation and phosphorylation, in the ESA sample, were also investigated using a proteomics approach.

CHAPTER 2 - Experimental Procedure

Sample Separation and Confirmation

Fresh equine serum (KSU ES) was kindly provided from Dr. James Lillich, Kansas State University College of Veterinary Medicine, collected during a routine check-up. The sample was obtained from a healthy, 14 years old male *Equus caballus* (horse), weigh 450 kg. SDS-PAGE and HPLC were used to compare the fresh sample against purchased equine serum (Sigma ES) (Sigma-Aldrich, St. Louis, MO).

SDS-PAGE

Electrophoresis was carried out using a SDS-PAGE apparatus and reagents (Invitrogen, Carlsbad, CA) and by following the protocol as described below. KSU and Sigma ES 1.0 µL each was diluted with 9.0 µL of Novex Tricine SDS sample buffer, followed by the heat denaturation at 85 °C for 2 minutes. Prepared samples 10.0 µL along with 10.0 µL of Mark 12 unstained molecular weight standard were applied to the each well in the 10-20 % Tricine gel 1.0 mm, 10 well which was pre assembled with 500.0 mL of Novex Tricine SDS running buffer. The prepared gel was run for 90 minutes at a constant 125 V followed by staining with Coomassie blue R-250 (Bio-Rad, Hercules, CA) and then destained with deionized distilled water (D.D.water).

HPLC

An HPLC column, Hydro-RP (reverse phase) 80 A, 150 X 4.60 mm 4 micron column (Phenomenex, Torrance, CA) was equilibrated with 0.1 % trifluoroacetic acid (TFA) for 3 hours prior to the experiment. Both KSU and Sigma ES were diluted with D.D.water to 1/10 of the original concentration. The diluted sample 20.0 µL was injected into the System Gold HPLC (Beckman, Fullerton, CA) with pre equilibrated Hydro-RP column and run with a mobile phase, D.D water containing 0.1 % TFA, for 30 minutes while the absorbance was monitored at 215 nm.

Protein Concentration Determination

KSU ES was diluted 1/5, 1/10, and 1/100 of the original concentration using D.D.water. To the 2.0 µL of diluted sample, 13.0 µL of D.D.water and 5.0 µL of the SDS sample buffer were added to make a total volume of 20.0 µL in each sample. Bovine serum albumin (BSA) (Pierce, Rockford, IL) standard, 2.0 mg/mL, was diluted to a concentration of 0.2 mg/mL with D.D.water. Sample buffer 5.0 µL was added to 0.5, 1.0, 5.0, and 10.0 µL of diluted BSA standard. D.D.water was added into the BSA mixture to make a total volume of 20.0 µL in each sample mixture. A total 20.0 µL of prepared serum and BSA standards were loaded into the prepared tricine gel. The electrophoresis was performed by following the standard protocol using Invitrogen SDS-PAGE apparatus as described in the previous section. UN-SCAN-IT gel software version 6.1 (Silk Scientific, Orem, UT) was used to determine the density of each band after scanning the completed gel using a ScanJet ADF (Hewlett Packard, Palo Alto, CA). The prescribed dilution of the BSA concentration standards were plotted in the x-axis and average pixels of the corresponding bands were plotted in the y-axis to make a standard calibration graph. This standard calibration graph was later used to determine the concentration of equine serum albumin.

Chemical Modification

6-Iodoacetamido Fluorescein (6-IAF)

The dye 6-Iodoacetamido fluorescein (6-IAF, molecular weight = 515.26) was purchased from Molecular Probes (Invitrogen, Carlsbad, CA). Freshly obtained KSU ES (native), 50.0 µL, was mixed with equal volume of 5 mM of 6-IAF stock solution (2.58 mg/mL 50 % DMF in D.D.water.) and placed in 4 °C for an overnight reaction in the dark (more than 10 fold excess 6-IAF to a free cysteine). Unreacted 6-IAF was removed by using a Sephadex G-25 column (Pharmacia Fine Chemicals, Piscataway, NJ) using D.D.water. SDS-PAGE and UV-visible spectrophotometer were used to confirm the attachment of 6-IAF to the free cysteine site in the serum. Purified sample was covered and kept in 4 °C.

6-IAF Label Confirmation and Separation of ESA Using SDS-PAGE

SDS sample buffer, 10.0 µL, was added to 10.0 µL of the collected 6-IAF labeled serum fractions and then the mixture was applied to the SDS-PAGE using the standard protocol. After running, the gel was first rinsed with D.D.water to wash out the SDS and 6-IAF. It was then visualized using a Typhoon 9410 Imager (Amersham Biosciences, Piscataway, NJ), which was set at 488 nm and 520 nm for the excitation and emission wavelengths, respectively. Image Quant TL 7.0 software (GE Healthcare bioscience AB, Uppsala, Sweden) was used for the image analysis. The sample was prepared for the in-gel digestion by following the standard protocol; 6-IAF attached serum albumin band from completed gel was cut into pieces and washed with 50 % acetonitrile twice for 15 minutes each. It was then frozen and dried using FreeZone 2.5Plus lypholizer (LABCONCO, Kansas City, MO).

6-IAF Label Confirmation Using UV-visible Spectrophotometer

Purified 6-IAF labeled equine serum fractions were placed into a quartz cuvette with 10 mm pathlength (Uvonic Instruments, Inc., Plainview, NY) and scanned over the range of 200-700 nm, using a Cary 50 Bio UV-visible spectrophotometer (Varian, Palo Alto, CA) and Cary Win UV software (Varian, Palo Alto, CA). The blank used in this experiment was prepared by taking a 1/100 diluted equine serum in D.D.water.

Carboxymethylation

The freshly obtained KSU ES was treated with iodoacetamide (IAA, molecular weight = 184.96). To 10.0 µL of the KSU ES (native) sample, 90.0 µL of 50 mM IAA (Fluka, Sigma-Aldrich, St. Louis, MO) in 4 M urea (Invitrogen, Carlsbad, CA) was added. It was then incubated at the room temperature for 1 hour under dark (more than 500 fold excess IAA to a free cysteine). The incubated sample, 1.0 µL, was mixed with 9.0 µL of SDS sample buffer and applied to the tricine gel followed by the application of standard protocol. The serum albumin band was cut from the stained gel and prepared for in-gel enzyme digestion. For reducing the sample, the unmodified serum albumin band from the completed SDS-gel was excised and cut into smaller pieces. Pierce In-Gel

Tryptic Digestion Kit (Thermo scientific, Rockford, IL) was used for in-gel enzymatic digestion by following the experimental procedure described here. The gel pieces were first destained with 200.0 µL of destaining solution, 50 % acetonitrile, and then incubated in the room temperature for 30 minutes. This step was repeated several times to completely remove the staining dye. After the destaining solution was removed, 30.0 µL of 50 mM Tris (2-carboxyethyl) phosphine Hydrochloride [33] (TCEP·HCl, structure and chemical equation are shown in Figure 2-1 [34]) was added and incubated for 10 minutes at 60°C for in-gel reduction (more than 100 fold excess to 17 disulfide bonds). The reducing solution was removed and 30.0 µL of 100 mM IAA in digestion buffer (more than 500 fold excess to 35 cysteines) was added to the sample followed by 1 hour of incubation at room temperature in dark for in-gel alkylation (carboxymethylation). The alkylation solution was removed and the sample was washed twice, 30 minutes each, with 200.0 µL destaining buffer to remove excess IAA. It was then prepared for the in-gel digestion by lypholyzation.

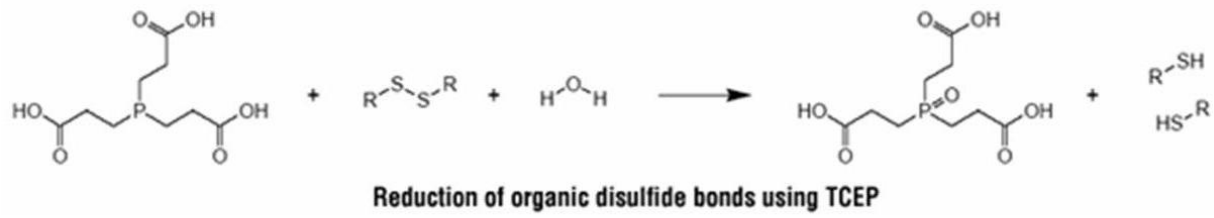

Figure 2-1 Tris(2-carboxyethyl)phosphine (TCEP) structure and chemical equation

Proteolytic Digestion

Trypsin Gold, mass spectrometry grade (Promega, Madison, WI) 10 ng/µL in 50 mM ammonium bicarbonate (Fluka/Sigma-Aldrich, St. Louis, MO), 30.0 µL, was added to the prepared gel pieces followed by the addition of 30 µL of digestion buffer (50 mM ammonium bicarbonate). The sample was then incubated for overnight at 30 °C. The same procedure was repeated for the digestion using the endoproteinase Lys-C (Roche, Mannheim, Germany) 10 ng/µL in D.D.water. The digested sample was extracted from

the gel pieces by first transferring the digestion mixture to a new tube, then adding 100.0 μL of 0.1 % TFA in 50 % acetonitrile extraction solution. The sample was again incubated for 30 minutes at the same temperature for further extraction. The extraction solution was removed and combined with previously extracted digest mixture. This step was repeated one more time. The combined solution was freeze dried for 3 hours or until the sample was completely dried.

MS and MS/MS Measurements

Matrices; 2,5-Dihydroxybenzoic acid (DHB) (Sigma-Aldrich, St. Louis, MO) 30 mg and α-Cyano-4-hydroxycinnamic acid (CCA) 97 % (Sigma-Aldrich, St. Louis, MO) 10 mg were each dissolved in 1.0 mL of 33 % acetonitrile/ 0.1 % TFA. Another matrix; 1,5-Diaminonaphthalene (DAN) (Fluka/Sigma-Aldricha, St. Louis, MO) 10 mg was dissolved in 1.0 mL 50 % acetonitrile. DAN was prepared just before mixing with the sample and must be shielded against light to avoid auto oxidation. The chemical structures of all three matrices are shown in Figures 2-2 though 2-4. DAN is a new type of matrix known to have a reducing capability with addition of laser energy from the MS instrument.[25] The reducing capability of DAN was tested using a synthesized vasopressin by mixing sample and prepared matrix in 1:1 ratio. Vasopressin is a peptide hormone containing 2 cysteines, 1 disulfide bond, and synthesized in the magonocellular cells of brain to be released into the blood stream. Its main function is to regulate the water retention in the body leading to the control of urine levels.[35] For the MS measurements of ESA, the dried extracted sample was first resuspended in 10.0 μL of 0.1 % TFA, then 1.0 μL of resuspended sample mixture was mixed into 5.0 μL of prepared DHB or 1.0 μL of DAN. 2.0 μL of sample/matrix mixture was applied to a MTP 384 massive target plate (Bruker Daltonics GmbH, Billerica, MA) and measured using positive ion reflector mode. The Ultraflex II TOF/TOF mass spectrometer, Flex control-compass 1.1 data acquation 3.0, Flex analysis 3.0, Bio Tools 3.0 software were purchased from Bruker Daltonics GmbH (Billerica, MA) and used for mass spectral measurement, data acquisition and analyses. With the use of peak detection, algorithm centroid monoisotopic, masses in the range of 0-5000 Da were assigned with the signal to noise threshold fixed to 5.

The MS/MS analyses were performed by the MALDI LIFT-TOF/TOF MS technique which increases the ion potential energy. Using laser-induced-dissociation (LID), the sequence of produced peptide fragments within the precursor ion can be easily assigned. The MS measurements were externally calibrated using Peptide Calibration Standard I (Bruker Daltonics GmbH, Billerica, MA). Mixture of monoisotopic ion mass $[M+H]^+$ in the standard includes 757.40 Da (Bradykinin 1-7), 1046.54 Da (Angiotensin II), 1296.68 Da (Angiotensin I), 1347.47 Da (Substance P), 1619.82 Da (Bombesin), 2093.09 Da (ACTH 1-17), 2465.20 Da (ACTH 18-39), and 3147.47 Da (Somatostatin 28). Additional internal calibration was also achieved in the flex analysis 3.0 by taking pre-calculated monoisotopic ion mass $[M+H]^+$ of the digested ESA fragments. Background subtraction or smoothing of acquired data was not performed in these data sets.

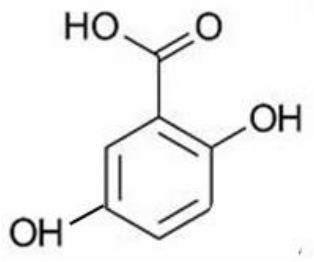

Figure 2-2 Chemical structure of 2,5-Dihydroxybenzoic acid (DHB)

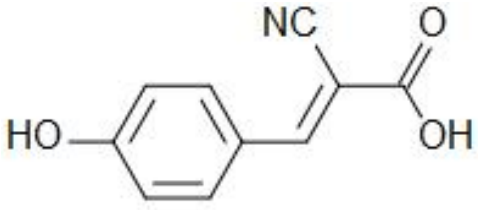

Figure 2-3 Chemical structure of α-Cyano-4-hydroxycinnamic acid (CCA)

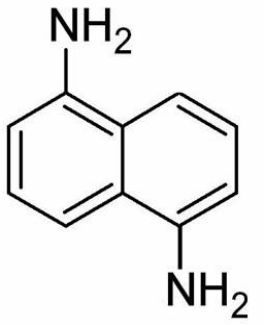

Figure 2-4 Chemical structure of 1,5-Diaminonaphthalene (DAN)

PTM Investigation

Glycosylation Detection using Glycoprotein Staining Kit

GelCode Glycoprotein Staining Kit was purchased from Pierce (Rockford, IL). Horseradish Peroxidase (HP) and soybean trypsin inhibitor (TI) (included in the kit) were used for positive and negative controls respectively. Both dried controls were resuspended by adding 0.5 mL of D.D.water to make 2 mg/mL solution. It was then diluted to make a concentration of 1 mg/mL using the SDS sample buffer. 1.0 µL of the equine serum was diluted with 10.0 µL of D.D.water and 22.0 µL of SDS sample buffer to make 1 ug/µL solution. BSA (used in the protein concentration determination experiment) was also diluted with SDS sample buffer to make 1 mg/mL solution. Both 5.0 and 10.0 µL of each control and prepared sample were applied to the tricine gel. The SDS-PAGE standard protocol previously described was used to carry out electrophoresis. The completed gel was stained by following the direction briefly described here. The gel was fixed in 50 % methanol, and then washed with 3 % acetic acid. After removing the wash solution, the oxidizing solution (include in the kit) was added to the gel and gently agitated for 15 minutes. To the gel, 3 % acetic acid was added again to gently wash the gel for a few times. The glycoprotein staining reagent (included in the kit) was added which react only with the glycoproteins. After discarding the staining solution, a reducing solution (include in the kit) was added. Finally, the gel was washed with 3 % acetic acid and D.D.water then scanned using Epson TWAIN5 USB Scanner.

Phosphorylation Detection Using Phosphoprotein Staining Kit

The Pro-Q Diamond Phosphoprotein Gel Stain Kit was purchased from Molecular Probes (Invitrogen, Carlsbad, CA). The same tricine gel which was used for the glycoprotein investigation was again used for this experiment. The completed gel was stained by following the direction briefly described here. The gel was fixed with the fixing solution; 50 % methanol/ 3 % acetic acid followed by rinsing with D.D.water. Then Pro-Q Diamond stain was added to the gel followed by 1 hour incubation in dark. Destaining solution was prepared by mixing 20 % acetonitrile and 50 mM sodium acetate at pH 4. The stained gel was transferred into the destaining solution for three times to remove the background, and then the gel was once again washed with D.D.water for twice. The stained gel was visualized using Typhoon 9410 with 532 nm excitation wavelength and 560 nm long pass emission wavelength. Image Quant TL 7.0 software was used for further image analysis.

CHAPTER 3 - Results

Sample Confirmation

The KSU and Sigma ES showed identical bands based on the SDS-PAGE analysis (Figure 3-1). Serum albumin showed a distinctively broad and diffused band at approximately 50 kDa based on the comparison with a GDH (glutamic dehydrogenase = 55.4 kDa) which is contained in the molecular marker. The HPLC data of the diluted samples showed almost identical peaks as that shown in the Figure 3-2. The main peak for both samples appeared approximately 14 minutes after the sample was injected. KSU serum showed more small peaks than the Sigma sample in the period between 7 and 10 minutes and again after 17 minutes.

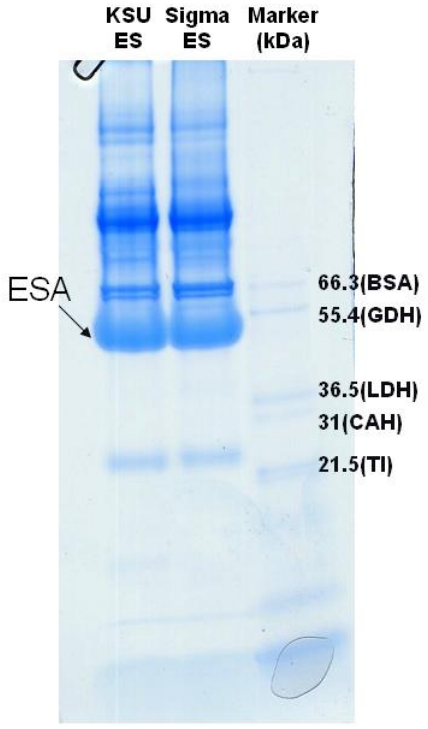

Figure 3-1 Sample confirmation by SDS-PAGE
BSA= Bovine Serum Albumin, GDH= Glutamic Dehydrogenase, LDH= Lactate Dehydrogenase, CAH= Carbonic Anhydrase, TI= Trypsin Inhibitor

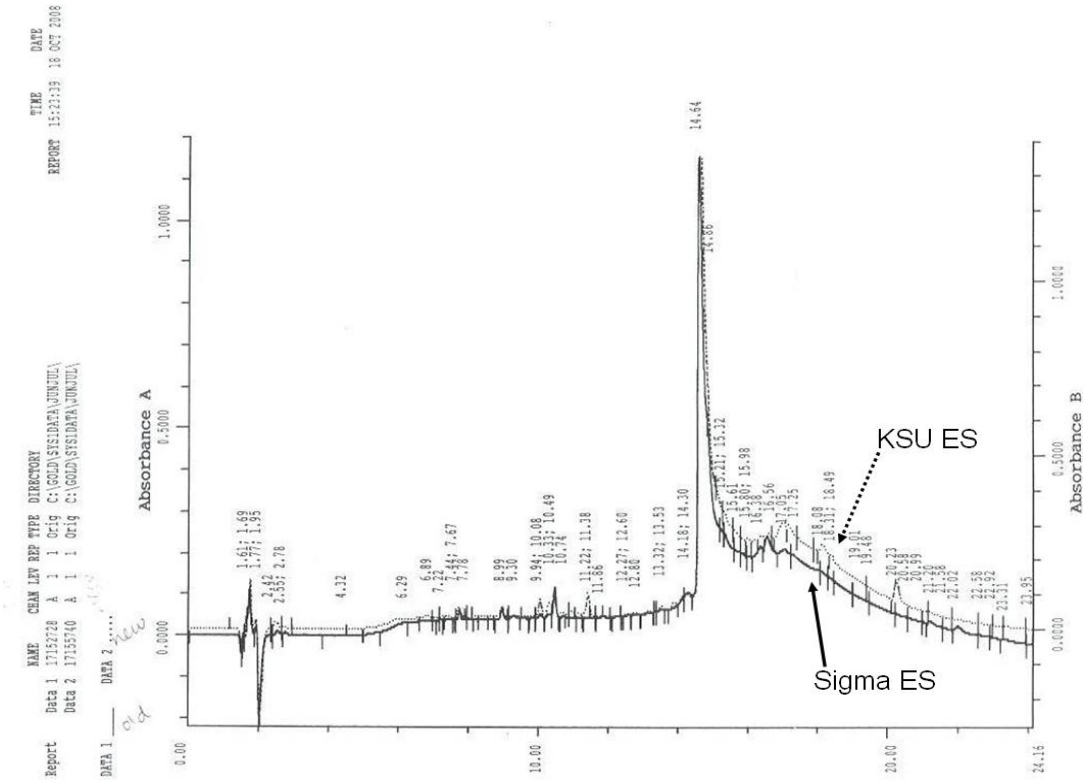

Figure 3-2 Sample confirmation by HPLC
Dotted line = KSU ES
Solid line = Sigma ES

Protein Concentration Determination

The protein concentration determination standard calibration graph (Figures 3-3 for the gel image, 3-4 for the data table, and 3-5 for the graph) was made by taking μg of standard BSA along the x-axis and average number of pixel on y-axis. A best fit line was calculated as $y = 13.016x + 1.0665$ with the R^2 value of 0.9982 as shown in the Figure 3-5. A 1/100 diluted equine serum albumin band, far right on the Figure 3-3, was used to find the concentration of the sample. The protein concentration of the ESA sample was determined to be 33 mg/mL. Since the molecular weight of ESA was known as 68,554 g/mol, molar concentration of the sample was calculated to be 480 μM. Based on these calculations, the concentrations of disulfide bond and cysteine were determined to be 8.16 nmol and 16.8 nmol in 1μL of serum, respectively.

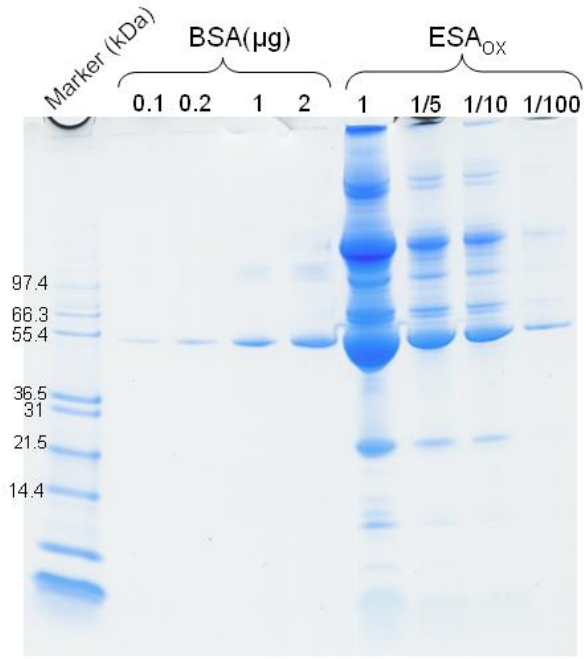

Figure 3-3 Protein concentration determination gel image

BSA (µg)	Avg.pixel
0.1	1.76
0.2	4.23
1	14.23
2	27

ESA concentration: 33 mg/mL
17 disulfide bonds/mol→ 8.16 nmol disulfide bonds
35 cys/mol → 16.8 nmol cys

Figure 3-4 Protein concentration determination data table

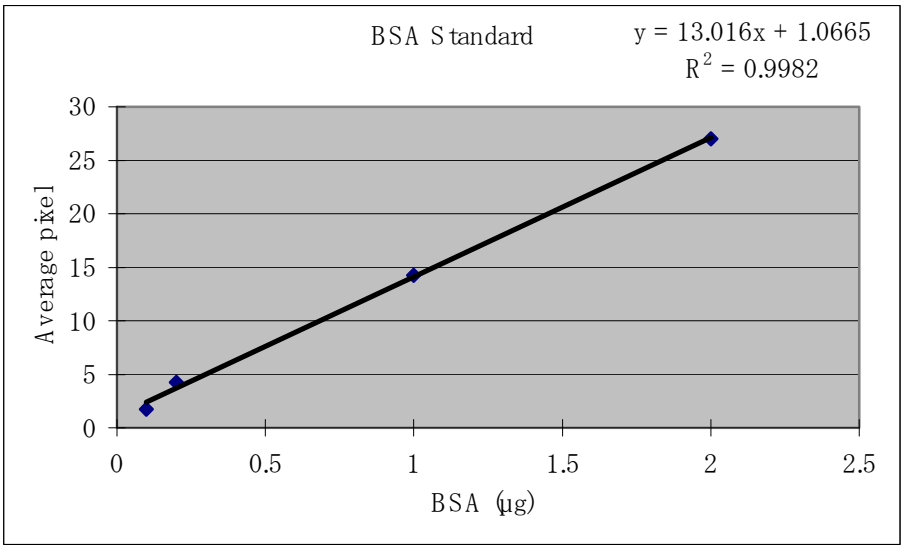

Figure 3-5 Protein concentration standard graph
BSA in µg (x-axis), Average number of pixels from the densitometer (y-axis)

Chemical Modification

6-Iodoacetamido Fluorescein

Figure 3-6 represents an image of a tricine gel containing 6-iodoacetamido fluorescein (6-IAF) labeled equine serum after it was purified from the excess 6-IAF. The chemical structure of the 6-IAF attachment to the free cysteine is shown in the Figure 3-7. The 6-IAF labeled serum albumin band is in yellow. Weak yellow bands are also visible in lower (< 5 kDa) and higher (~97 kDa) molecular weight region. 6-IAF modification of the serum was also confirmed by scanning sample from 200 to 700 nm using spectrophotometer. Figure 3-8 shows the spectrum of scanned data. The blue line represents the blank, 1/100 diluted unmodified equine serum in D.D.water, and the yellow line represents 6-IAF labeled sample. One clear broad peak was observed near 500 nm region. The λmax of the sample was determined to be 0.347 at 484.1 nm.

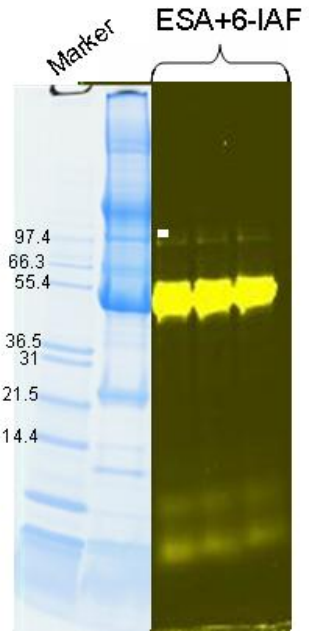

Figure 3-6 6-IAF labeled ESA_{ox} gel image

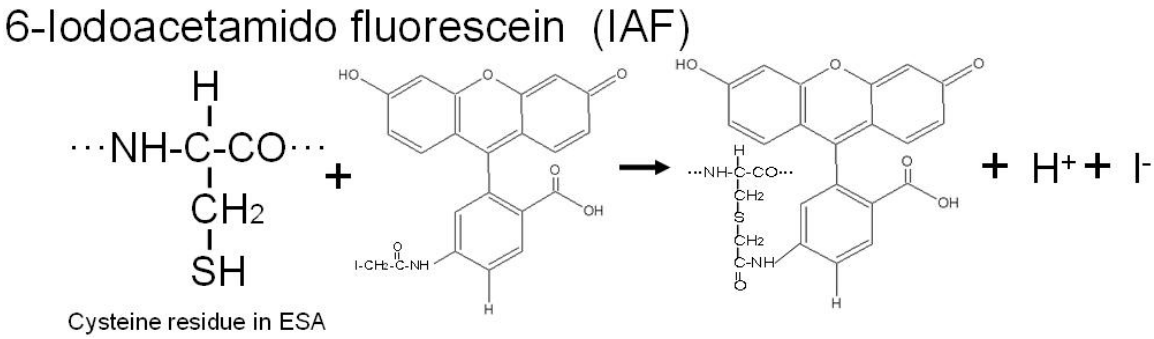

Figure 3-7 Chemical equation for the 6-IAF labeling of cysteine residue

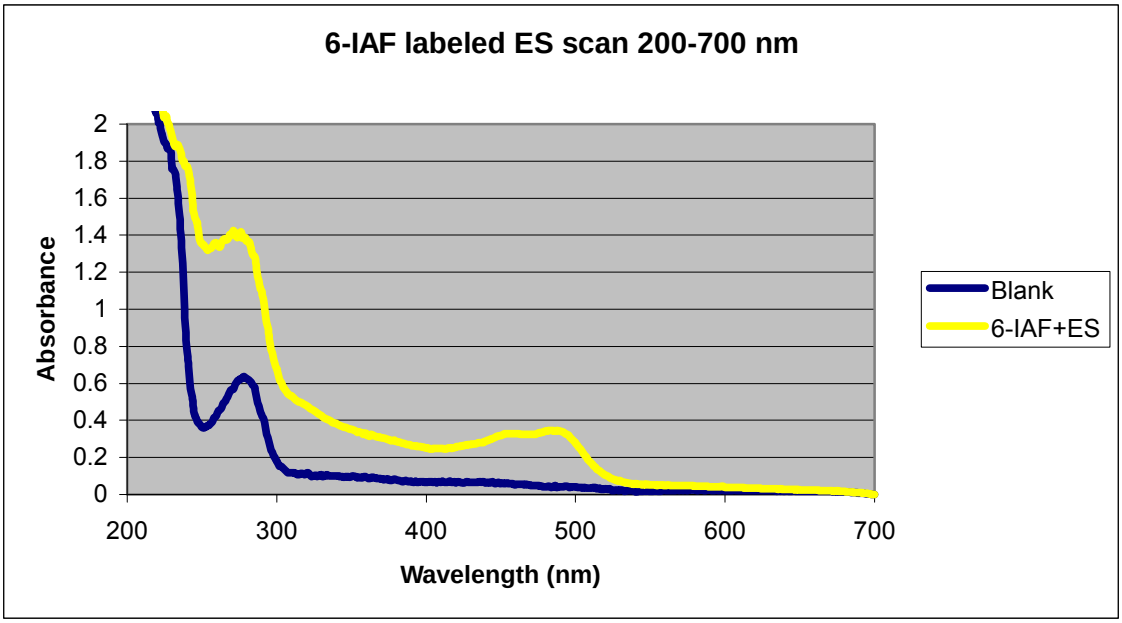

Figure 3-8 UV-visible wavelength scan spectrum of 6-IAF labeled ES

Carboxymethylation

Both IAA labeled and unmodified equine serum sample were run in the SDS gel as shown in the Figure 3-9. The large diffused ESA bands are visible in the first and third lanes. The molecular weight of the sample is estimated around 50 kDa based on the molecular marker applied in the middle lane (GDH = 55.4 kDa). Figure 3-10 shows the chemical equation of carboxymethylation of the cysteine. Visible changes due to this chemical modification were not observed from the gel image. The ESA band (indicated

by an arrow) from the unmodified equine serum (ES only) was later cut out, destained, reduced and carboxymethylated within the gel.

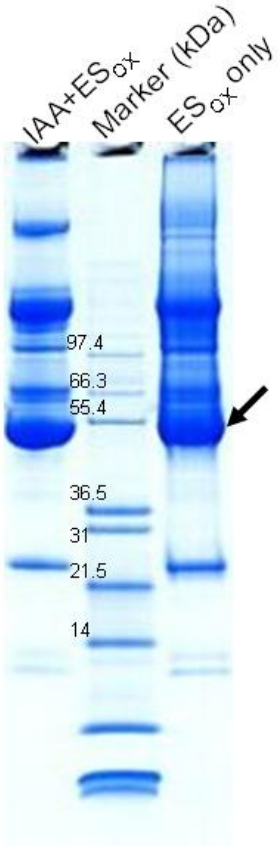

Figure 3-9 IAA labeled ES gel image

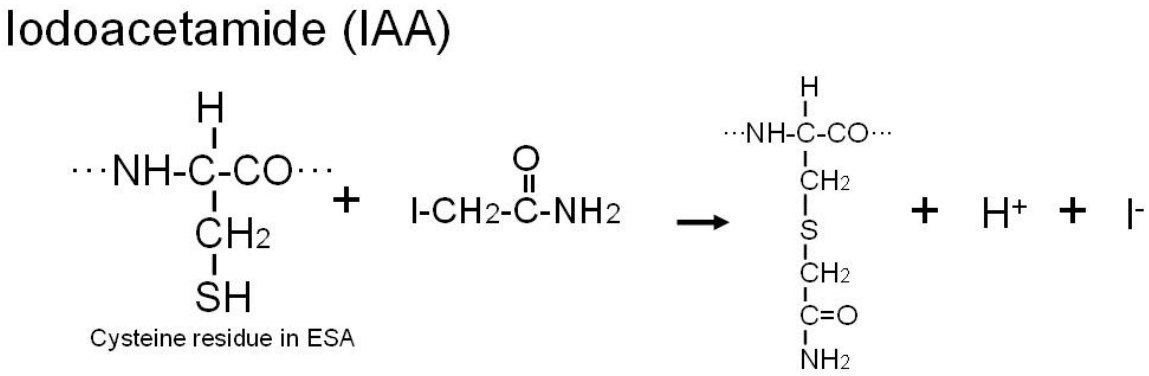

Figure 3-10 Chemical equation for the carboxymethylation with cysteine residue

Proteolytic Digestion

The expected disulfide bonded peptides produced by the enzymatic digestion are summarized in the Table 3-1 and 2. The ESA_{ox} (oxidized = native equine serum albumin) produces 9 disulfide peptide clusters using trypsin and 8 with the protease Lys-C. The disulfide bonded peptides included intrapeptide bond (blue), interpeptide bond (red), and a combination of both. Digestion of the ESA_{red} (reduce equine serum albumin) using trypsin or Lys-C produced 81 and 63 fragments, respectively.

Tryptic fragment	Fragment/Expected mass
FREE 58 GLVLVAFSQYLQQCPFEDHVK	Fragment | (red) 45-65 | 2421.227
Tryp1 77 86 CAADESAENCDK	Fragment | (red) | (ox) 77-86 | 1255.457 | 1253.457
Tryp2 99 LCTVATLR 114 ATYGELADCCEK 115 NECFLTHK 125	Fragments | (red) 98–105 | 876.497 106–117 | 1302.534 123–130 | 991.467 2177.031 2292.001 3166.498
Tryp3 147 LKPEPDAQCAAFQEDPDK 191 ADFTECCPADDK 192 LACLIPK 200	Fragments | (red) 139–156 | 2001.922 186–197 | 1314.498 198–204 | 757.464 3314.42 2069.962 4069.884
Tryp4 223 CSSFQNFGER 268 276 ECCHGDLLECADDR 269	Fragments | (red) | (ox) 223–232 | 1174.495 | 267–280 | 1578.598 | 1576.596 2749.093
Tryp5 288 YICEHQDSISGK 301 ACCDKPLLQK 302 SHCIAEVK 312	Fragments | (red) 286–297 | 1379.626 300–309 | 1118.57 310–317 | 886.445 2496.196 2003.015 3380.641
Tryp6 339 EICK 383 392 CCAEADPPACYR 384	Fragments | (red) | (ox) 337–340 | 492.249 | 383–394 | 1298.496 | 1296.496 1786.745
Tryp7 415 NCDLFEEVGEYDFQNALIVR 460 CCK 461 LPCSENHLALALNR 471	Fragments | (red) 414–433 | 2374.102 460–462 | 353.131 469–482 | 1550.811 2725.233 1901.942 4274.044
Tryp8 484 LCVLHEK 499 CCTDSLAER 500 RPCFSALELDEGYVPK 510	Fragments | (red) 483–489 | 841.46 499–507 | 997.408 508–523 | 1823.9 1836.868 2819.308 3658.768
Tryp9 537 AETFTFHADICTLPEDEK 581 CCGR 582 EACFAEEGPK 590	Fragments | (red) 527–544 | 2066.938 581–584 | 438.159 588–597 | 1080.467 2503.097 1516.626 3581.564

Table 3-1 Expected disulfide bonded peptide clusters by trypsin digestion

Lys-C fragment	Fragment/Expected mass
FREE GLVLVAFSQYLQQ⁵⁸CPFEDHVK	Fragment | (red) 45-65 | 2421.227
Lysc1 ⁷⁷CAADESAENCDK⁸⁶	Fragment | (red) | (ox) 77-86 | 1255.457 | 1253.457
Lysc2 ⁹⁹LCTVATLRATYGELADCCEK^{114 115} QEPERNECFLTHK¹²⁵	Fragments | (red) | (ox) 98–117 | 2160.013 | 2158.013 118–130 | 1630.764 | } 3786.777
Lysc3 PEPDAQC¹⁴⁷AAFQEDPDK ADFTECC¹⁹¹PADDK¹⁹² LACLIPK²⁰⁰	Fragments | (red) 141–156 | 1760.743 186–197 | 1314.498 198–204 | 757.464 } 3073.241 + 2069.962 } 3828.705
Lysc4 ²²³CSSFQNFGERAVK ²⁶⁸ECCHGDLLECADDRADLAK²⁷⁶ ²⁶⁹	Fragments | (red) | (ox) 223–235 | 1472.695 | 267–285 | 2076.878 | 2074.878 } 3545.573
Lysc5 ²⁸⁸YICEHQDSISGK ³⁰¹ACCDK³⁰² SHCIAEVK³¹²	Fragments | (red) 286–297 | 1379.626 300–304 | 539.195 310–317 | 886.445 } 1916.821 + 1423.64 } 2801.266
Lysc6 ³³⁹EICK ³⁸³CCAEADPPAC³⁹²YRTVFDQFTPLVEEPK ³⁸⁴	Fragments | (red) | (ox) 337–340 | 492.249 | 383–408 | 2929.321 | 2927.321 } 3417.57
Lysc7 ⁴¹⁵NCDLFEE VGEYDFQNAL IVRYTK ⁴⁶⁰VGSRCCK⁴⁶¹ ⁴⁸⁴LPESERLPCSENHLALALNRLCVLHEK⁴⁷¹ ⁴⁹⁹CCTDSLAERRPCFSALELDEGYVPK⁵¹⁰ ⁵⁰⁰	Fragments | (red) | (ox) 414–436 | 2766.308 | 456–462 | 752.354 | 463–489 | 3084.608 | 499–523 | 2802.29 | 2800.29 } 3516.662 + 3834.962 + 5882.898 } 9397.56
Lysc8 ⁵³⁷AETFTFHADICTLPEDEK ⁵⁸¹CCGREDK⁵⁸² EACFAEEGPK⁵⁹⁰	Fragments | (red) 527–544 | 2066.938 581–587 | 810.323 588–597 | 1080.467 } 2875.261 + 1888.79 } 3951.728

Table 3-2 Expected disulfide bonded peptide clusters by Lys-C digestion

MS and MS/MS Analysis

DAN (1,5-Diaminonaphthalene) Reduction Potential Confirmation

All of the following MS analyses were achieved by taking an average of 1000 sets of calibrated data. The reducing capability of DAN was initially tested with a synthesized vasopressin as shown in the Figure 3-11. The oxidized vasopressin was first measured using the matrix CCA with the detected peaks indicated in blue on the Figure 3-12. The molecular ion $(M+H)^+$ of the oxidized vasopressin is 1085.4 m/z and was observed as the highest peak in the full spectrum (inset in, top right corner). Vasopressin with the deletion of C-terminus glycine had a molecular ion of 1028.4 m/z which was also detected (inset in, top left corner) in the same spectrum. The sample reduced by DAN should gain 2 protons which shifts the molecular ions to be 1087.4 m/z and 1030.4 m/z for glycine deleted vasopressin. Both of those two peaks (indicated in red) were observed with high intensities, up to 10^5 scale. A peak of non-reduced vasopressin was observed however, which overlapped with the oxidized vasopressin peak.

Figure 3-11 Vasopressin sequence and molecular weight

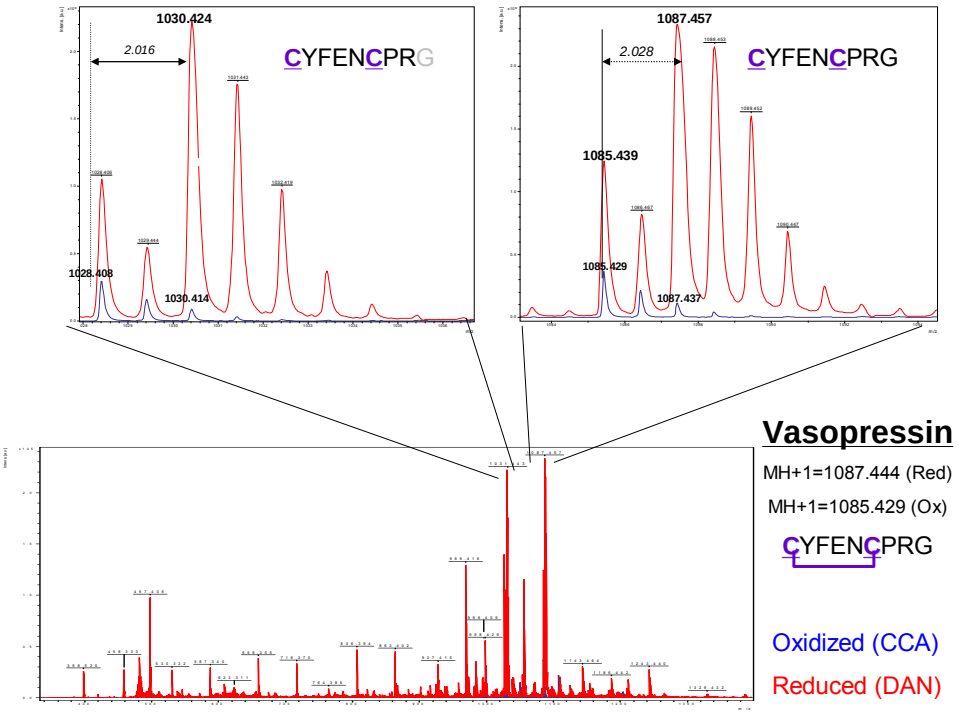

Figure 3-12 MS spectra of Vasopressin CCA/DAN

Trypsin Digested Fragments

DHB (2,5-Dihydroxybenzoic acid)

Figure 3-13 shows the trypsin digested fragments of IAA labeled ESA_{ox} using DHB as a matrix. The IAA-labeled free cysteine fragment (residues 45-65), $[M+H]^+$ = 2478.2 m/z, was detected. Peptide clusters Tryp 1 (1 miscleavage), 4, 5, 7, and 8 from the Table 3-1 were also detected with smaller intensities. The total sequence coverage of this spectrum was calculated to be 47.0 % as indicated in blue in the Figure 3-14. The 6-IAF labeled ESA_{ox}, $[M+H]^+$ = 2809.2 m/z , was also detected using DHB as shown in the Figure 3-15. The 6-IAF labeled ESA spectrum had a reduced intensity (10^3) compared to the IAA labeled ESA spectrum (10^4). Also, 6-IAF labeled ESA had lesser number of peaks, particularly larger than 2000 m/z molecular ion peaks were rarely observed. IAA labeled (carboxymethylated) ESA_{red} was measured using DHB (Figure 3-16) and showed peaks of higher intensities throughout the entire spectrum. Total sequence coverage of the reduced sample was determined to be 78.3 % (Figure 3-18) after MS/MS

measurement was applied. The intensity of the reduced sample was generally higher (10^5) compared to the oxidized sample.

Additionally, produced fragment from the ESA_{red} peak 1231 m/z was analyzed using MS/MS method. Precursor ion 1231.5 m/z had a possibility of being one of the following fragments; IAA labeled CSSFQNFGER (residues 223-232) or ACCDKPLLQK (residues 300-309). After the analysis of precursor ion 1231 m/z spectrum, SFQNF peptide sequence was successfully identified. This data verified the peak 1231 m/z to be IAA labeled CSSFQNFGER (residues 223-232), as shown in the Figure3-17.

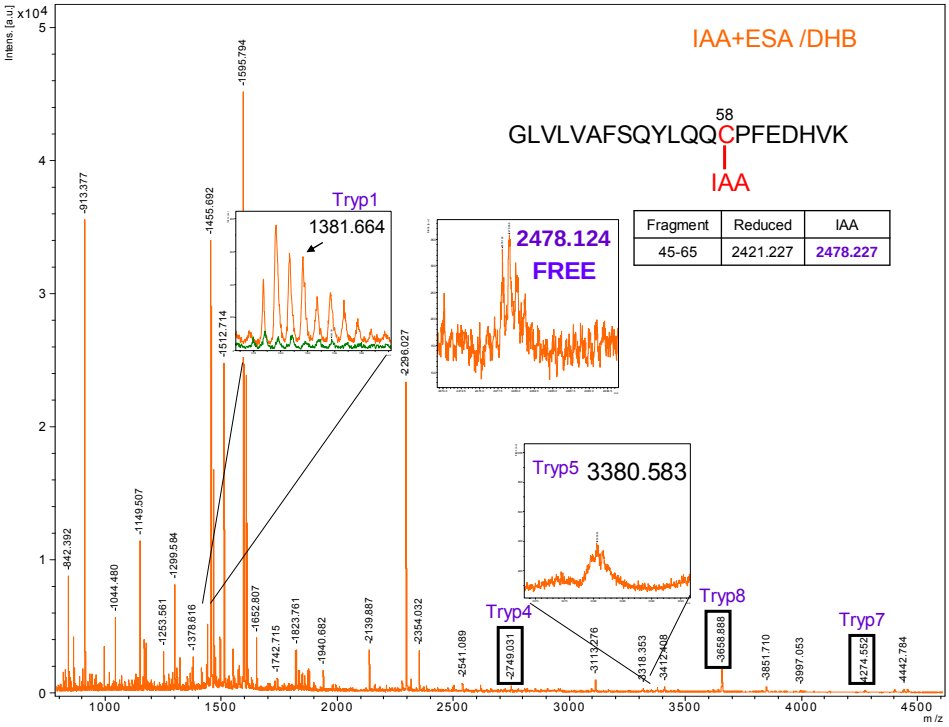

Figure 3-13 MS spectra of IAA labeled ESA_{ox}/DHB

```
  1 MKWVTFVSLL FLFSSAYSRG VLRRDTHKSE IAHRFNDLGE KHFKGLVLVA
 51 FSQYLQQCPF EDHVKLVNEV TEFAKKCAAD ESAENCDKSL HTLFGDKLCT
101 VATLRATYGE LADCCEKQEP ERNECFLTHK DDHPNLPKLK PEPDAQCAAF
151 QEDPDKFLGK YLYEVARRHP YFYGPELLFH AEEYKADFTE CCPADDKLAC
201 LIPKLDALKE RILLSSAKER LKCSSFQNFG ERAVKAWSVA RLSQKFPKAD
251 FAEVSKIVTD LTKVHKECCH GDLLECADDR ADLAKYICEH QDSISGKLKA
301 CCDKPLLQKS HCIAEVKEDD LPSDLPALAA DFAEDKEICK HYKDAKDVFL
351 GTFLYEYSRR HPDYSVSLLL RIAKTYEATL EKCCAEADPP ACYRTVFDQF
401 TPLVEEPKSL VKKNCDLFEE VGEYDFQNAL IVRYTKKAPQ VSTPTLVEIG
451 RTLGKVGSRC CKLPESERLP CSENHLALAL NRLCVLHEKT PVSEKITKCC
501 TDSLAERRPC FSALELDEGY VPKEFKAETF TFHADICTLP EDEKQIKKQS
551 ALAELVKHKP KATKEQLKTV LGNFSAFVAK CCGREDKEAC FAEEGPKLVA
601 SSQLALA
```

→Total sequence coverage = 47.0%

Figure 3-14 Sequence coverage of IAA labeled ESA_{ox} by trypsin
Blue = Detected sequence, Highlighted C = Cys 58 (free cys)

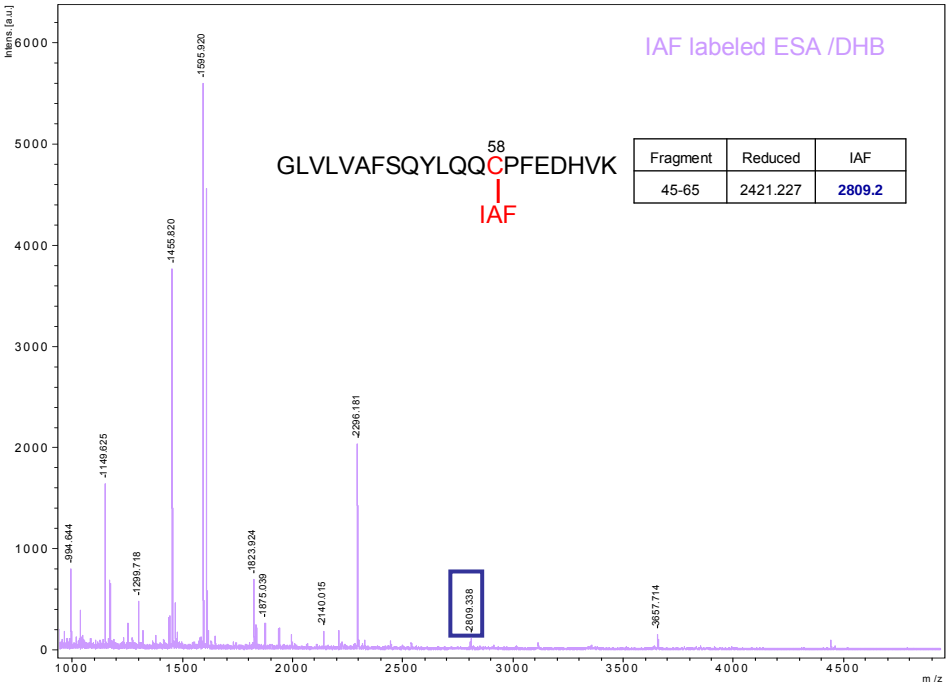

Figure 3-15 MS spectra 6-IAF labeled ESA_{ox}/DHB MS spectra

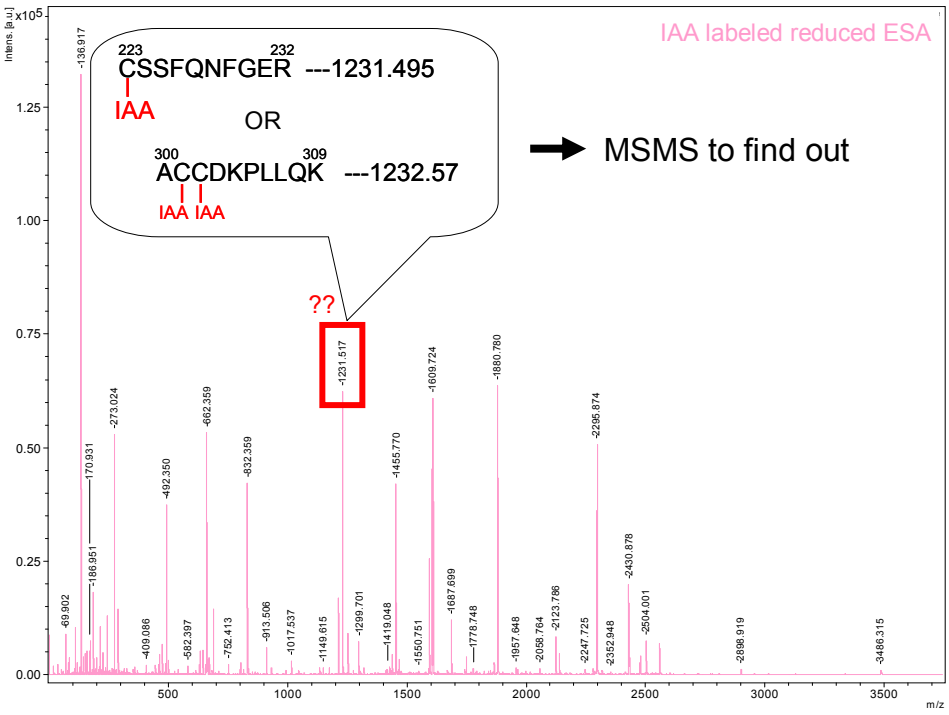

Figure 3-16 MS spectra of carboxymethylated ESA_{red}/DHB
Peak 1231 is indicated in red square, 2 IAA labeled fragments are possible

MSMS on peak 1231

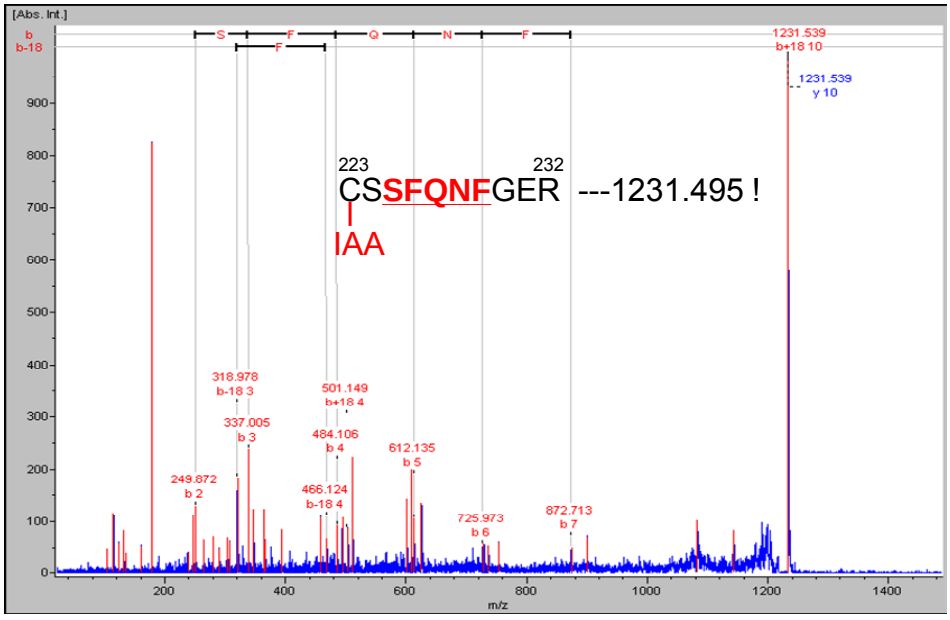

Figure 3-17 MS/MS spectra of peak 1231

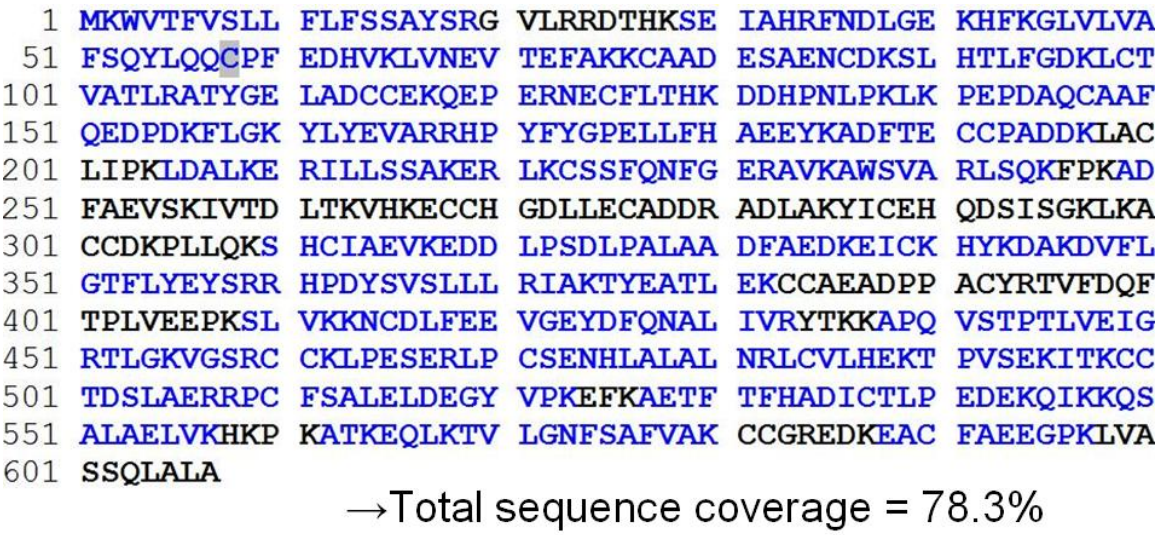

Figure 3-18 Sequence coverage carboxymethylated ESA_{red} by trypsin
Blue = Detected sequence, Highlighted C = Cys 58 (free cys)

DAN

Trypsin digested ESA_{ox} fragments were analyzed by the MS using DAN as a matrix which has a reducing capability upon increased laser energy from the MS instrument. Figure 3-19 shows MS spectra of IAA labeled ESA cluster Tryp8 with DHB (orange) and DAN (green). Tryp8 is composed of 3 digestion fragments which are linked together with 2 disulfide bonds. The total molecular ion of this cluster was 3658.8 m/z and it was measured using DHB as shown in the right bottom corner of the Figure. Reduced fragments ions appeared using DAN and each are indicated in the corresponding colors (841.4 in purple, 997.4 in lime green, and 1823.9 in light blue). Similar analysis was performed to the other tryptic peptide clusters as shown in the Appendix A, Figures A-1 through A-7.

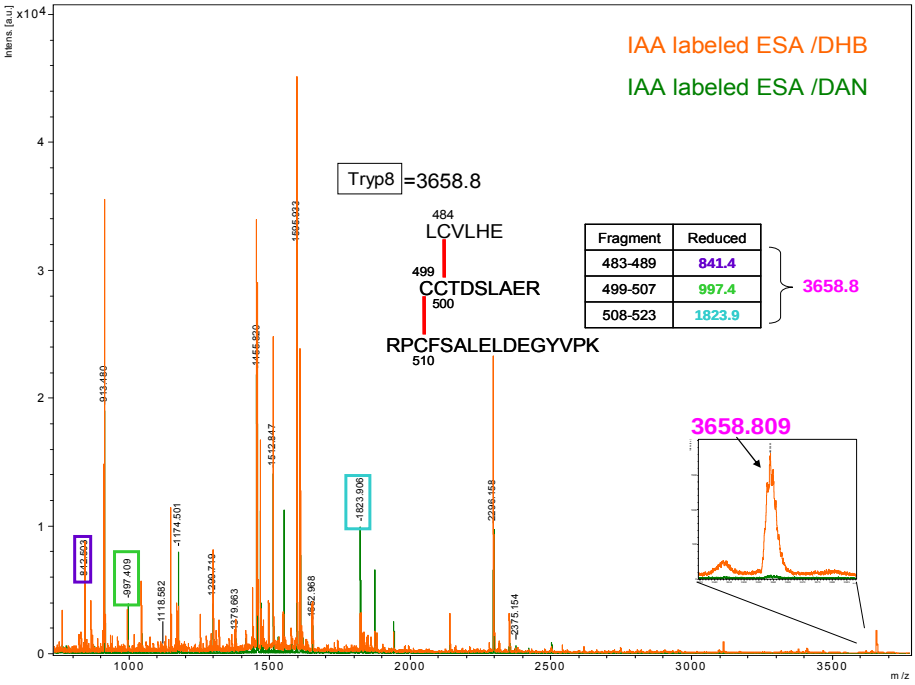

Figure 3-19 MS spectra of Tryp8
Orange = MS spectra of the ESA_{ox}
Green = MS spectra of the ESA_{red} (with DAN).
Pink = Tryp8 peptide cluster (3658.8 m/z)
3 reduced fragments are shown with corresponding colors

Lys-C Digested Fragments

DHB

The IAA labeled ESA_{ox} was first analyzed with DHB as shown in the Figure 3-20. An IAA labeled free cysteine fragment as well as the disulfide bonded peptide clusters Lysc2, 4, and 5 (from the Table 3-2) were detected. Total sequence coverage including free cysteine fragments was 54.0 % (Figure 3-21). Carboxymethylated ESA_{red} was also measured using DHB and produced 89.1 % sequence coverage with higher MS signal intensities (10^5) as shown in the Figures 3-22 and 23. The MS spectrum of ESA_{red} peptides had a higher intensity throughout the entire spectrum. The spectrum of IAA labeled ESA_{ox} (Figure 3-20) was observed with lesser intensity (10^4) and also noticeable intensity gaps were seen between peaks of disulfide bonded peptide clusters and non-disulfide bonded fragments. For example, disulfide bonded clusters Lysc4 and Lysc5 had

intensities of approximately 0.2 x 10^4 a.u. (arbitrary unit) while non-disulfide bonded fragments such as 1145.365 m/z and 1994.701 m/z had intensities of 2.0 X 10^4 a.u.

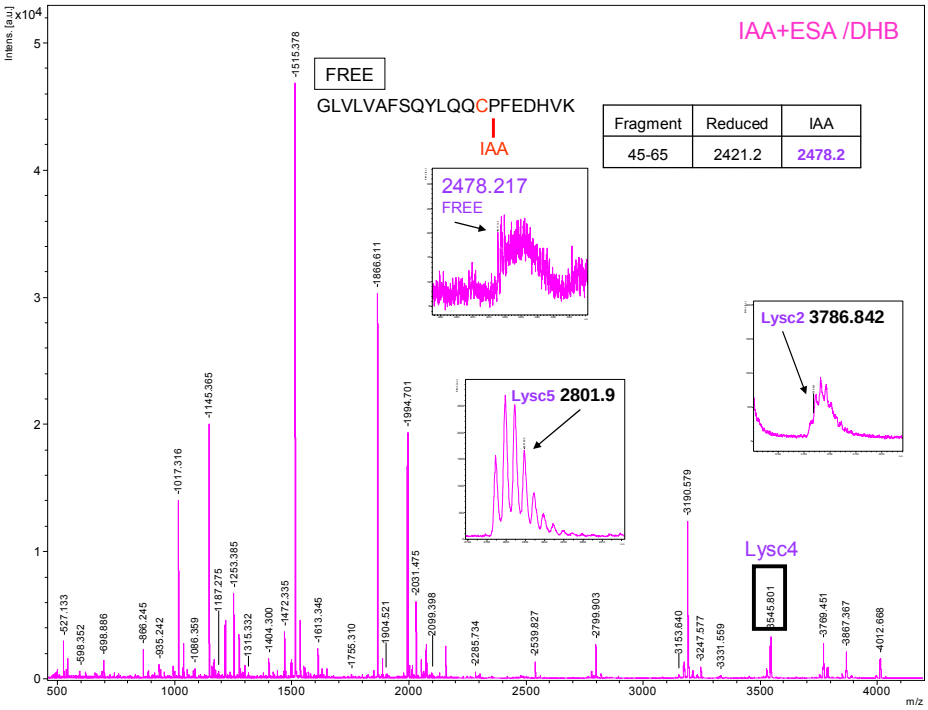

Figure 3-20 MS spectra of IAA labeled ESA_{ox} /DHB

```
  1 MKWVTFVSLL FLFSSAYSRG VLRRDTHKSE IAHRFNDLGE KHFKGLVLVA
 51 FSQYLQQCPF EDHVKLVNEV TEFAKKCAAD ESAENCDKSL HTLFGDKLCT
101 VATLRATYGE LADCCEKQEP ERNECFLTHK DDHPNLPKLK PEPDAQCAAF
151 QEDPDKFLGK YLYEVARRHP YFYGPELLFH AEEYKADFTE CCPADDKLAC
201 LIPKLDALKE RILLSSAKER LKCSSFQNFG ERAVKAWSVA RLSQKFPKAD
251 FAEVSKIVTD LTKVHKECCH GDLLECADDR ADLAKYICEH QDSISGKLKA
301 CCDKPLLQKS HCIAEVKEDD LPSDLPALAA DFAEDKEICK HYKDAKDVFL
351 GTFLYEYSRR HPDYSVSLLL RIAKTYEATL EKCCAEADPP ACYRTVFDQF
401 TPLVEEPKSL VKKNCDLFEE VGEYDFQNAL IVRYTKKAPQ VSTPTLVEIG
451 RTLGKVGSRC CKLPESERLP CSENHLALAL NRLCVLHEKT PVSEKITKCC
501 TDSLAERRPC FSALELDEGY VPKEFKAETF TFHADICTLP EDEKQIKKQS
551 ALAELVKHKP KATKEQLKTV LGNFSAFVAK CCGREDKEAC FAEEGPKLVA
601 SSQLALA
```

→Total sequence coverage = 54.0%

Figure 3-21 Sequence coverage of IAA labeled ESA_{ox} by Lys-C digestion
Red = Detected sequence, Highlighted C = Cys 58 (free cys)

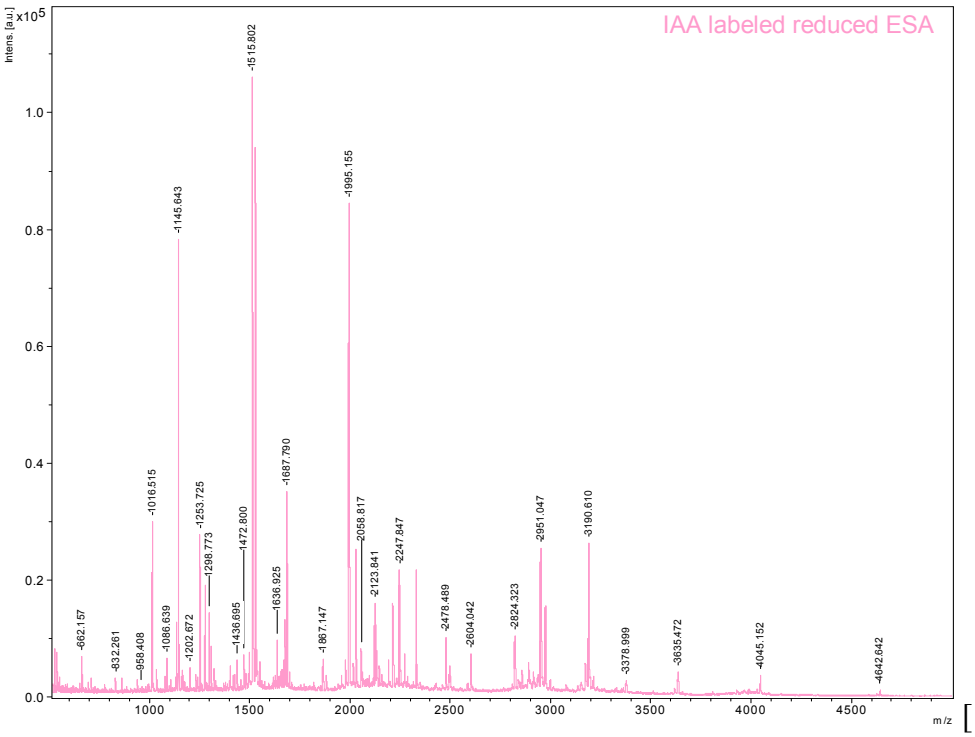

Figure 3-22 MS spectra of carboxymethylated Lys-C digested ESA_{red} /DHB

```
  1 MKWVTFVSLL FLFSSAYSRG VLRRDTHKSE IAHRFNDLGE KHFKGLVLVA
 51 FSQYLQQCPF EDHVKLVNEV TEFAKKCAAD ESAENCDKSL HTLFGDKLCT
101 VATLRATYGE LADCCEKQEP ERNECFLTHK DDHPNLPKLK PEPDAQCAAF
151 QEDPDKFLGK YLYEVARRHP YFYGPELLFH AEEYKADFTE CCPADDKLAC
201 LIPKLDALKE RILLSSAKER LKCSSFQNFG ERAVKAWSVA RLSQKFPKAD
251 FAEVSKIVTD LTKVHKECCH GDLLECADDR ADLAKYICEH QDSISGKLKA
301 CCDKPLLQKS HCIAEVKEDD LPSDLPALAA DFAEDKEICK HYKDAKDVFL
351 GTFLYEYSRR HPDYSVSLLL RIAKTYEATL EKCCAEADPP ACYRTVFDQF
401 TPLVEEPKSL VKKNCDLFEE VGEYDFQNAL IVRYTKKAPQ VSTPTLVEIG
451 RTLGKVGSRC CKLPESERLP CSENHLALAL NRLCVLHEKT PVSEKITKCC
501 TDSLAERRPC FSALELDEGY VPKEFKAETF TFHADICTLP EDEKQIKKQS
551 ALAELVKHKP KATKEQLKTV LGNFSAFVAK CCGREDKEAC FAEEGPKLVA
601 SSQLALA
```

→Total sequence coverage = 89.1%

Figure 3-23 Sequence coverage of IAA labeled ESA_{red} by Lys-C
Red = Detected sequence, Highlighted C = Cys 58 (free cys)

DAN

The Lys-C digested disulfide bonded clusters were measured using the MS with DAN as a matrix. Figure 3-24 shows the MS spectra of IAA labeled peptide cluster Lysc7. The Lysc7 is the largest disulfide bonded peptide cluster with a $[M+H]^+$ = 9397.56 m/z in the native state. It contains 3 interpeptide and 1 intrapeptide disulfide bonds. The oxidized sample peak with DHB (pink) was not detected, however, all 4 fragments (indicated in corresponding color) were observed. Peak 2802.068 m/z indicates intrapeptide bond and after reduction of the bond with DAN, a 2 mass unit shifted peak was seen. Analysis was similarly done to the rest of peaks as shown in the Appendix B, Figures B-1 though B-5.

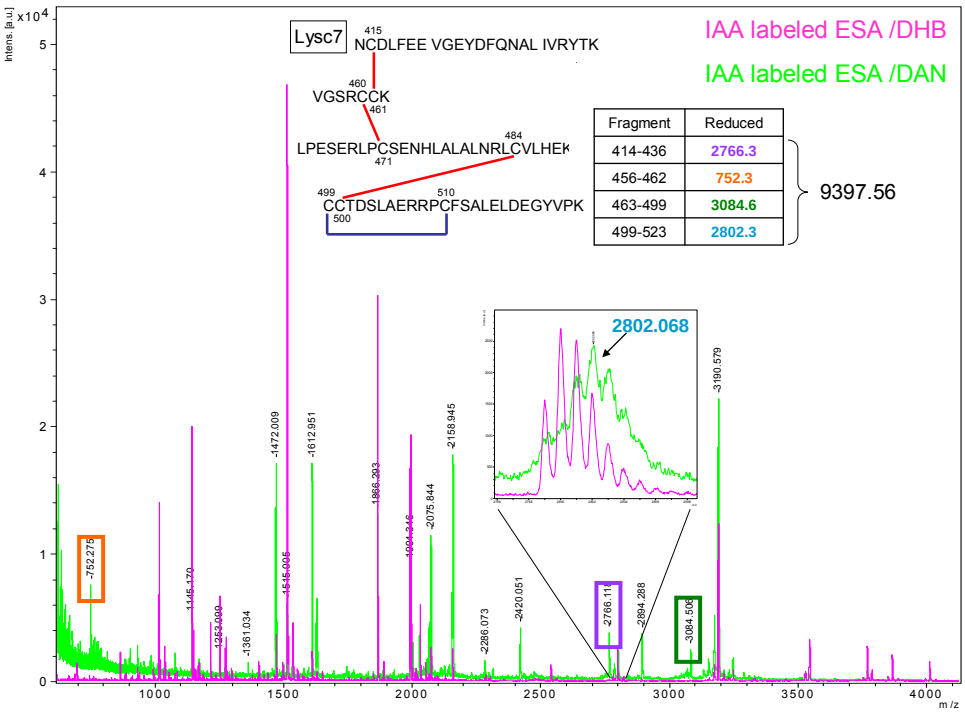

Figure 3-24 MS spectra of IAA labeled ESA_{ox} Lysc7
9397 m/z peak = not detected (not shown in the Figure above)
4 separated fragments = purple, orange, green, and light blue

PTM Investigation

Glycosylation Detection

The glycosylation state of the sample was investigated against both positive and negative controls, as well as BSA using a glycoprotein staining kit. Presence of glycoprotein in gel was indicated with pink color stain. Figure 3-25 shows a comparison between samples loaded in different amounts and stained with Coomassie or the glycoprotein staining kit. Image on the left shows the gel stained with R-250 Coomassie blue while the image on the right shows the same gel which is stained with the glycoprotein staining kit. As a positive control, the HP (horse radish peroxidase) band was observed around 48 kDa in both gels while the negative control TI (soybean trypsin inhibitor) band was seen only in the Coomassie stained gel with molecular weight of nearly 16 kDa. The KSU ES was also loaded into the gel and the regular bands were seen with Coomassie staining, though no ESA band was observed in the glycoprotein stained gel. Some of the higher molecular weight bands (66, 97, and 120 kDa) within the serum sample were observed in the glycoprotein stained gel. The BSA band was also not visible in the glycoprotein stained gel.

Phosphorylation Detection

Phosphoprotein detection was performed using the phosphoprotein gel staining kit on the same gel used for the glycoprotein detection. Figure 3-26 shows Coomassie stained gel on the left and phosphoprotein stained gel on the right. ESA bands were observed in the phosphoprotein stained gel as indicated by the arrows. Compare to the Coomassie stained gel, the size of the ESA band is much narrower in the phosphoprotein stained gel. The HP and TI, positive and negative control, bands were also visible in this experiment. The HP bands were significantly stained compared to the TI bands. BSA bands were slightly visible in the phosphoprotein stained gel.

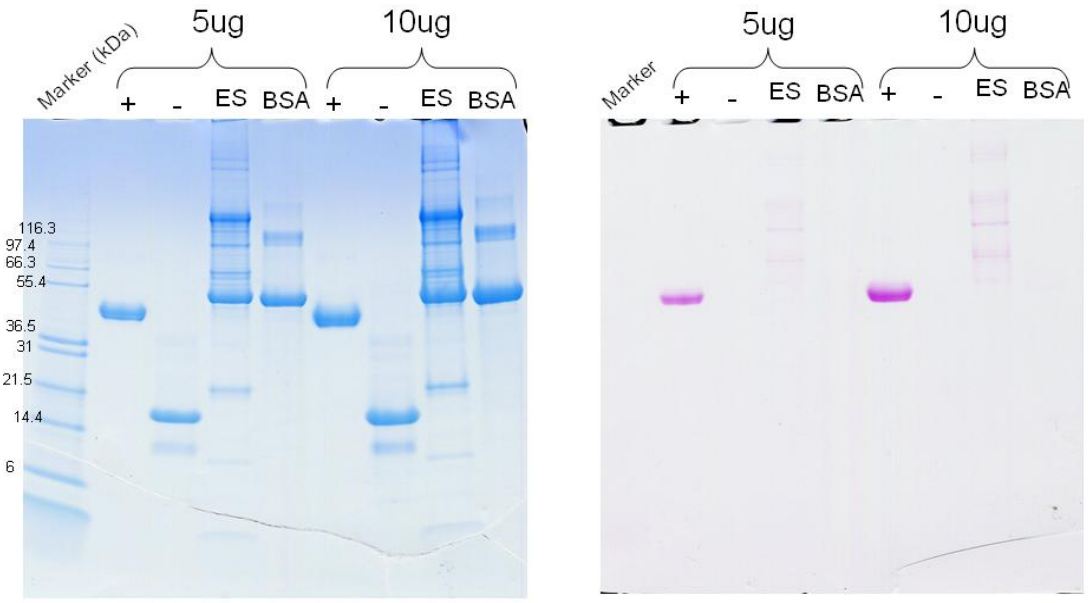

Figure 3-25 Glycoprotein stained gel image
Glycoprotein gel image (right), Coomassie stained gel image (left)
+ Positive control = HP (horseradish peroxidase)
– Negative control = TI (soybean trypsin inhibitor)
Different amount of controls, equine serum (ES), and bovine serum albumin (BSA) were loaded along with the molecular marker

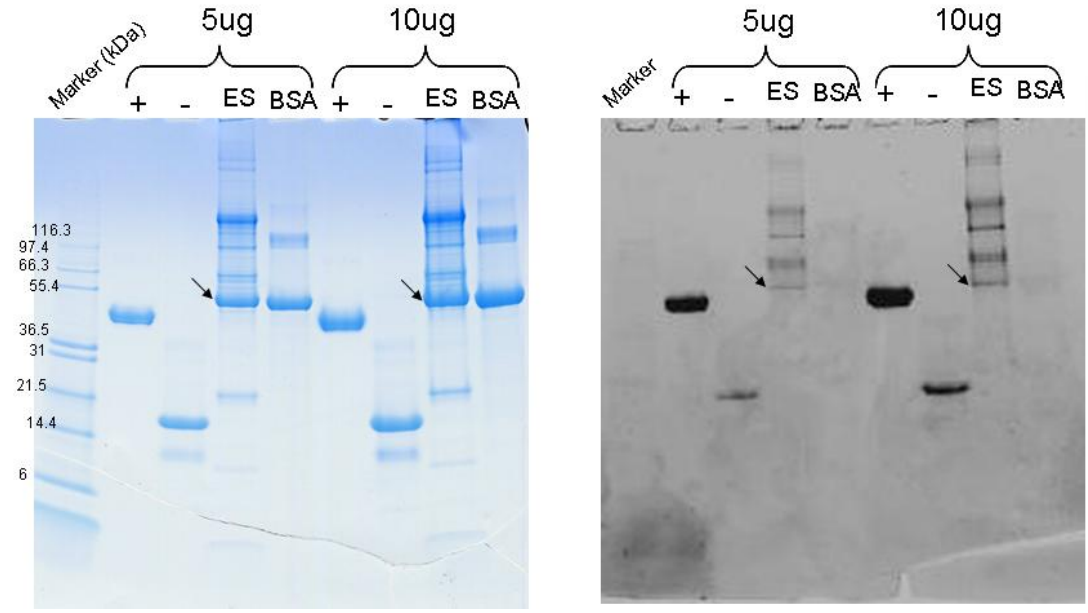

Figure 3-26 Phosphoprotein stained gel image
Phosphoprotein gel image (right), Coomassie stained gel image (left)
Same gel was used as described in the Figure 3-25. Arrow indicates ESA band.

CHAPTER 4 - Discussion

Sample Confirmation

The distribution of the bands from the SDS-PAGE (Figure 3-1) and peak pattern from the HPLC (Figure 3-2) of KSU ES were virtually identical to the ones seen with the Sigma ES. These results confirmed that the band identified as ESA from the K-State veterinary medicine was the protein of interest. The Sigma ES was essentially used to standardize the experimental procedures for KSU ES analyses.

Protein Concentration Determination

Based on the graph shown in the Figure 3-5 along with some additional calculations, the concentration of the ESA in the KSU ES sample was determined to be 33 mg/mL. The distinctive wide band observed in the SDS-PAGE (Figure 3-3) confirmed a high abundance level of the albumin within the serum albumin sample. The sequence based molecular weight of the ESA_{ox} is 68,554.3 Da while observed molecular weight which is based on the molecular weight standard was approximately 50 kDa. This molecular weight shift can be explained by the presence of disulfide bonds. The sample applied to the SDS-PAGE was heat denatured but not reduced in order to maintain the native disulfide pattern of the ESA. Due to the presence of 17 disulfide bonds in the ESA, the protein could only be denatured up to certain level, possibly preventing the SDS from fully interacting with the protein and increasing its net charge. In other words, the presence of this particularly large number of the disulfide bonds preserves a tightly folded structure which affects its mobility in the SDS gels. This change in the size of protein could perhaps lead to the observed molecular shift in the gel. To confirm this hypothesis, a reduced ES sample was applied to the SDS gel. The ESA_{red} band was shifted up to that of the BSA standard 66.3 kDa (gel image not shown). Based on the observed concentration of ESA, the calculated amount for the cysteine is 16.8 nmol, and 8.16 nmol for the disulfide bonds.

Chemical Modification

6-Iodoacetamido Fluorescein

The broad yellow bands present in Figure 3-6 confirm the attachment of the 6-IAF to the free cysteine in the ESA. A few other proteins within the serum also appear to contain free cysteines as some fainter yellow bands were observed at a higher (97 kDa) and lower (< 5 kDa) molecular weights based on the molecular weight standard. A clear absorbance peak around 500 nm in the visible/UV spectrum (Figure 3-8) also indicated 6-IAF was attached to the free cysteine in the serum. This covalent bond between free sulfur of the cysteine side chain and carbon of the 6-IAF is highly stable. Therefore, once the 6-IAF is attached to the free cysteine, manipulation of the sample will not alter this modification even with a change of its surrounding conditions (Figure 3-7 for chemical equation). After the electrophoresis of 6-IAF labeled serum was finished, the SDS gel needed to be rinsed several times with D.D.water to all remove excess dye present in the lower portion of the gel. Labeling the protein using fluorescein is a practical method for protein analysis not only due to its high stability, but also completion of the modification can visually be assessed.

Iodoacetamide

The IAA labeled and non-modified equine serum had the identical bands in SDS PAGE as shown in the Figure 3-9. For the native (oxidized) sample, IAA should only modify the single free cysteine. When the sample was reduced using more than 100 fold of TCEP and quickly treated with more than 500 fold of IAA, all of the cysteine residues should be modified and therefore unable to reform disulfide bonds. Due to this chemical modification of the cysteines, the protease digestions were carried out much more effectively by increasing the enzyme's accessibility to the preferred cleavage sites. By generating small peptide fragments and eliminating the disulfide linkage, performance in MS is improved due to the greater ionization of digested fragments which in turn leads to the higher sequence coverage in the MS spectrum.

IAA labeling of the protein is a very important tool for peptide mapping along with the labeling with fluorescein. It covalently alkylates free sulfurs and therefore blocks any unwanted oxidations that would interfere which interpreting the data. As shown in the Figure 3-10, when IAA attaches to the free sulfur, both the I^- and H^+ are products of the reaction. This modification causes the cysteine containing fragments to gain 57 molecular mass per cysteine and this should be small enough compound to go and interact with a tightly folded protein to label any free sulfhydryl group in the native protein. Labeling the fully reduced protein helps to make sure that the protein will fully unfold during the proteolytic digestion process.

Proteolytic Digestion

The chemically modified samples were digested with two different enzymes; trypsin and Lys-C. The units of enzymes as well as the wash and extraction procedures were identical for both enzymes. Comparing these two enzymes, they both have similar activities and cleavage preferences. Trypsin cleaves protein at a C-terminus of both arginines and lysines while Lys-C specifically cleaves only at a C-teminus of lysines. These enzymes both have pH optima of 7 to 9 and are mostly resistant to 0.1 % SDS, 1M urea, or up to 10 % acetonitrile (Product information from Promega and Roche). However a notable difference is the cost for conducting a single experiment. In our protocol, it only costs $ 0.19 for trypsin while costs $5.52 for Lys-C to digest the same sample. This significant difference in cost between two enzymes seems to reflect the percent sequence coverage obtained for ESA sample. Besides trypsin and Lys-C, chymotrypsin was used either independently or in combination with the other enzymes. However, due to lack in the specificity of chymotrypsin, satisfactory data were not obtained.

MS and MS/MS Analysis

DAN Reducing Potential Confirmation

Using the synthesized vasopressin, the reducing capability of DAN was tested. Synthesized vasopressin was first analyzed with the MS using CCA as a matrix. The oxidized vasopressin spectrum (blue line) in the Figure 3-12 indicated a peak at 1085.4 m/z with the highest intensity. Vasopressin was then analyzed using DAN to confirm its reducing capability. The MS spectrum obtained using the DAN (red line) showed a peak shift (to the higher mass) between the oxidized and reduced vasopressin. The reduced vasopressin had a molecular ion of $[M+H]^+ = 1087.5$ which also produced the highest intensity peak in the spectrum. Although the majority of the peak indicated a reduced vasopressin, a minor oxidized vasopressin peak was still detected. Based on this experiment, reducing capability of DAN was verified but not at 100 % efficiency for this sample. MS measurements using DAN as the matrix can be challenging mainly due to its preparation. DAN must be prepared immediately before applying to the sample plate because it quickly self-oxidizes within the solution. Also after the DAN solution was prepared, it needed to be kept in dark to avoid chemical reaction caused by exposure to light. The crystal of the peptide-DAN mixture formation on the MALDI target plate also contributes to improved sample ionization.

Analysis of Digested Fragments

MS Analysis Using DHB

As shown in the Figure 3-13, 5 out of 9 trypsin digested disulfide bonded oxidized peptide clusters (from Table 3-1) were found using DHB. This result generates only a total ESA_{ox} sequence coverage of 47.0 % including IAA/6-IAF labeled free cysteine fragment. IAA and 6-IAF labeled cysteine (Cys58) peaks, $[M+H]^+ = 2478.2$ m/z and $[M+H]^+ = 2809.2$ m/z, were detected thus confirming each of the two different chemical modifications. The 6-IAF labeled cysteine fragment (Figure 3-15) had a peak

with reduced intensity compared to that observed for the IAA labeled fragment. The ESA_{ox} modification with 6-IAF possibly reduced the ability of the digested fragments to become ionized in the MS. The identified peptide clusters contained mostly interpeptide disulfide bonds fragment except for the cluster Tryp4 which contained a single intrapeptide segment bond. Fragments containing intrapeptide bonds such as Tryp1 and 6 can be difficult to get ionized due to its compact structure.

Lys-C digested fragments showed similar digestion patterns compared to the trypsin digested fragments. Lys-C produced identical IAA/6-IAF labeled ESA_{ox} fragments as that observed with trypsin, again confirming each chemical modification. Due to the reduced number of cleavage sites compared to the trypsin, the sizes of the digested fragments were generally larger. Disulfide bonded peptide clusters Lysc2, Lysc4, and Lysc5 were detected using DHB and covered 54.0 % of the total IAA labeled ESA_{ox} sequence. With Lys-C, more non-disulfide bonded fragments were detected which contributed to this higher sequence coverage. Unlike the trypsin digested fragments, Lys-C created cluster Lysc7 which was composed of 4 large fragments linked by both intra and interpeptide disulfide bonds. This largest cluster had a calculated molecular ion $[M+H]^+ = 9397.6$ m/z which was not observed even using a different MS measurement methods, yet all the composed fragments were identified after reduction by DAN.

Comparing the two proteolytic enzymes, Lys-C produced higher sequence coverage for both native (oxidized) and reduced ESA. These results could be caused by the fact that the cleavage site for Lys-C is less in number, and therefore, less miscleavages. Because there are a fewer number of cleavage sites for Lys-C, the size of the peptide clusters and non disulfide bonded fragments are generally larger. Also, the ESA sequence contains several sites where lysine and arginine are located next to each other. These types of sites caused miscleavages during tryptic digestion and also produced some short (mono-, di- and tri-) non-detectable peptides which affected the overall sequence coverage. The observed difference in the cost of these two enzymes is also a consideration. It costs considerably more to use Lys-C compared to trypsin. Therefore, as data suggests, one should start proteolytic digestion with trypsin. Only

when higher sequence coverage is required, Lys-C can be used to enhance coverage. Using Lys-C as a primary proteolytic enzyme is not recommended due to its higher cost.

For determining the cleaved peptides, the MS method used predominantly in these experiments was MALDI-TOF positive ion reflector mode. This method only detects fragments which have a total positive ion charge. Since extracted fragments were resuspended in 0.1 % TFA and DHB was also dissolved in 33 % acetonitrile/ 0.1 % TFA, sample applied to the MS measurement was dried in an acidic environment. This condition protonates all the acidic amino acids such is glutamic acid and aspartic acid, thereby making peptides containing acidic amino acids measurable. However, the intensity of the fragment peaks containing large number of acidic amino acids were generally lower due to a decrease in sensitivity. Decreases in the sensitivity are generally caused by poor ionization and also higher mass molecular ions.

The reason that we are not able to detect all of the disulfide bonded clusters can be explained by a difference in the energy absorption among the digested fragments. Some digested fragments were more easily ionized compared to the others. Easily ionized fragments absorb more energy from the energy source leaving less energy to ionize the other fragments. In other words, MS measurements are sequence dependant. It also signifies that the hydrophobic samples are easier to ionize into the gas-phase and be detected while highly hydrophilic samples are not.

Miscleavage by the enzyme can be another reason for not recovering some fragments. Miscleavage of the fragments generates new higher molecular weight fragments. When the total molecular weight of the disulfide bonded cluster increases, the total ionization potential of the cluster can be decreased due to the presence of the additional amino acids. In this case, poor ionization of the cluster could lead to not identifying certain clusters. Overall, the sequence of ESA contains a large number of lysine and arginine. Some of those cleavage sites could be located in the buried section of the native (oxidized) sample that was inaccessible to the protease. As discussed in the introduction, the presence of disulfide bonds in the sample reduces the ionization potential of the segment. This condition also contributes greatly to the lowering of the peak intensity.

MS Analysis Using DAN

The reducing capability of DAN matrix was demonstrated in the experiment using synthesized vasopressin as discussed above. Proper preparation of the DAN solution is critical for obtaining the maximum reducing capability of DAN. The trypsin digested peptide clusters of ESA were measured using DAN in order to investigate its oxidation states of the different fragments. As one of the examples, Tryp8 (Figure 3-19) was shown to be made up of 3 tryptic fragments linked by 2 interpeptide disulfide bonds. The molecular ion mass of the Tryp8 is $[M+H]^+ = 3658.8$ m/z, which was previously detected using DHB. This peak was barely present when the sample was mixed with DAN. Rather 3 peaks, fragments 483-489, 499-507, and 508-523 were observed having molecular ions of $[M+H]^+ = 841.4$, 997.4, and 1823.9 m/z respectively. As we observed in the vasopressin experiment, reduction by the DAN was not complete, therefore, some of the reduced peaks were seen with reduced intensities. For example, the disulfide bonded cluster Tryp9 (Figure A-7) had $[M+H]^+ = 3581.5$ not measured. Tryp9 is composed of 3 fragments linked by 2 intrapeptide disulfide bonds. Two out of three fragments, residues 527-544: $[M+H]^+ = 2066.9$ m/z and 581-584: $[M+H]^+ = 438.1$ m/z, were measured using DAN, but 588-597: $[M+H]^+ = 1080.4$ m/z was not. Although the molecular ion of disulfide bonded cluster could not be measured, it was possible to identify several or all of the constituent fragments by measuring the sample using DAN.

Similar results were obtained for the Lysc7 peptide cluster (Figure 3-24). The cluster Lysc7 was composed of 4 fragments containing 4 disulfide bonds. The total molecular ion was calculated as $[M+H]^+ = 9397.5$ m/z and was not detected using DHB. However, peaks for all of the 4 fragments were detected after the sample was reduced with DAN. Fragment 499-523: $[M+H]^+ = 2802.3$ m/z was linked by both intra and interpeptide disulfide bonds. In the Figure 3-24, the partially reduced intrapeptide disulfide bonded fragment 499-523 was observed as indicated in the pink trace me green trace), the main peak was observed 2 mass unit higher at $[M+H]^+ = 2802.068$ m/z, indicating the reduction of an intrapeptide disulfide bond.

Due to the energy dissipation caused by the presence of existing disulfide bonds, the measurement of oxidized peaks can be challenging. However, with the help of

reducing ability of DAN, reduced fragments within an undetected disulfide linked peak become identified. In other words, DAN helps to distinguish between intra and interpeptide disulfide bonds within a digested peptide. If interpeptide disulfide bonds were formed, 2 additional peaks appear while the oxidized peak disappears or decreases. With the presence of an intrapeptide disulfide bond, the identified peak shifts 2 mass units higher. DAN is a new type of matrix which can reduce existing disulfide bonds without pre-treating a sample with any reducing reagent. This unique property of DAN contributes greatly to the field of protein structural analysis not only by shortening the analysis time, but also excluding the addition of any extra reducing chemicals or salts to the sample which could potentially interfere with the MS measurement.

MS/MS Analysis

The MS/MS measurements were carried out to identify the peak 1231 m/z (Figure 3-16), obtained from the IAA labeled ESA_{red} fragment. There were two fragment possibilities for peak 1231 m/z; IAA labeled ESA_{red} residues: CSSFQNFGER (223-232), or ACCDKPLLQK (300-309). The precursor ion, $[M+H]^+$ = 1231.517 m/z, was trapped and then fragmented with the application of higher laser energy which increases the potential energy of ions. As a result, the partial sequence, SFQNF, was observed (Figure 3-17). This data confirmed the assignment of peak 1231 to fragment CSSFQNFGER, 223-232. MS/MS is one of the unique features of the MALDI-TOF/TOF MS instrument and becomes practical for the assignment of ambiguous peaks. This technique is able to increase the usefulness of the instrument and often used for PMF (peptide mass fingerprint)-based protein identification[21]. Combining MS and MS/MS data, protein identification/ characterization is becoming a requirement for publishing MS data. Additional protein data bases, a variety of search engines using different algorithms and de novo sequencing programs, are also contributing to the rapid process in protein identification.

Overall sequence coverage

The sequence coverage obtained for each experiment is summarized in the Figures 4-1 and 4-2 for oxidized and reduced ESA samples, respectively. Sequences generated by only trypsin are indicated in red, only Lys-C is blue, and by both enzymes is indicated in purple. Sequences which were not identified using either enzymes are indicated in black. Lys-C digestion (Red) generated more sequences that were not covered by the trypsin digestion (Blue) within the native ESA_{ox} sample, which again confirms the higher detection rate for this enzyme. Residues 20-28, GVLRRDTHK, in the ESA_{red} were not detected by either enzymes (indicated in black). While fragments up to Arg_{19} generated by trypsin digestion were identified, the theoretical fragment generated by the Lys-C hydrolysis (residues 3-28: $[M+H]^+ = 3085.673$ m/z) was not detected due to its large ion mass. The two sequences containing adjacent cysteine residues, 383-4 and 581-2 (underlined) were not discovered either in native (oxidized) or reduced ESA samples. This could be indicating that those cys-cys residues are located in the tightly folded motif where either IAA or enzymes were not able to reach. Additionally, it is also possible that in the native ESA_{ox} sample, the presence of asparatic acids and glutamic acids near cys-cys residues are making negatively charged fragments which cannot be detected by the applied measuring method. Even though the ESA sample was prepared in the acidic environment prior to the MS measurement, these particular fragments could not be fully protonated which, as result, caused poor ionization of the fragments. The presence of an intrapeptide bond between Cys_{383}-Cys_{392} could also be one of the reasons to cause the poor ionization due to the energy absorption by the cysteines.

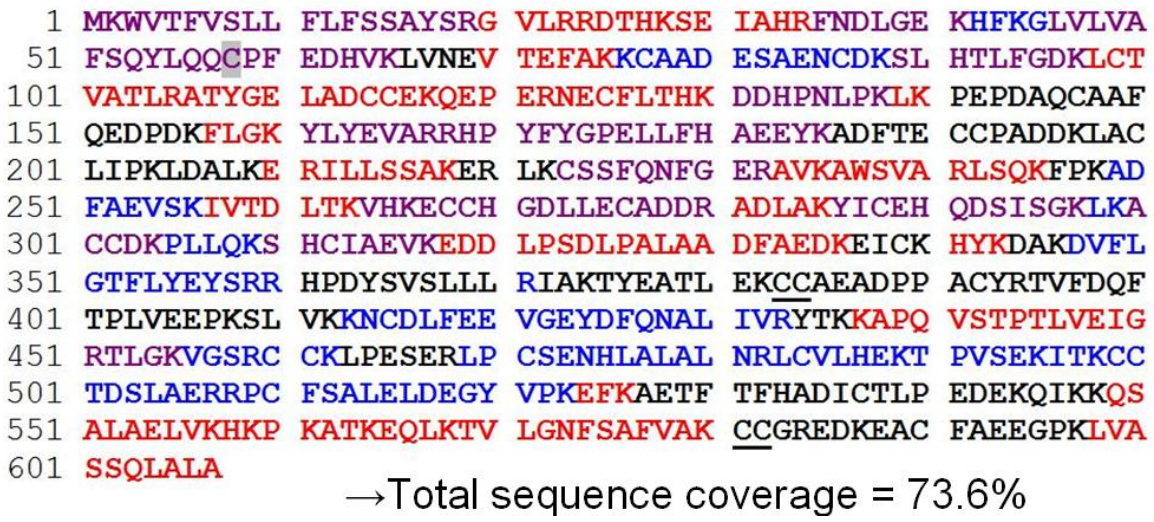

Figure 4-1 ESA_{ox} sequence coverage
Blue = by trypsin, Red = by Lys-C (red), Purple = by both enzymes

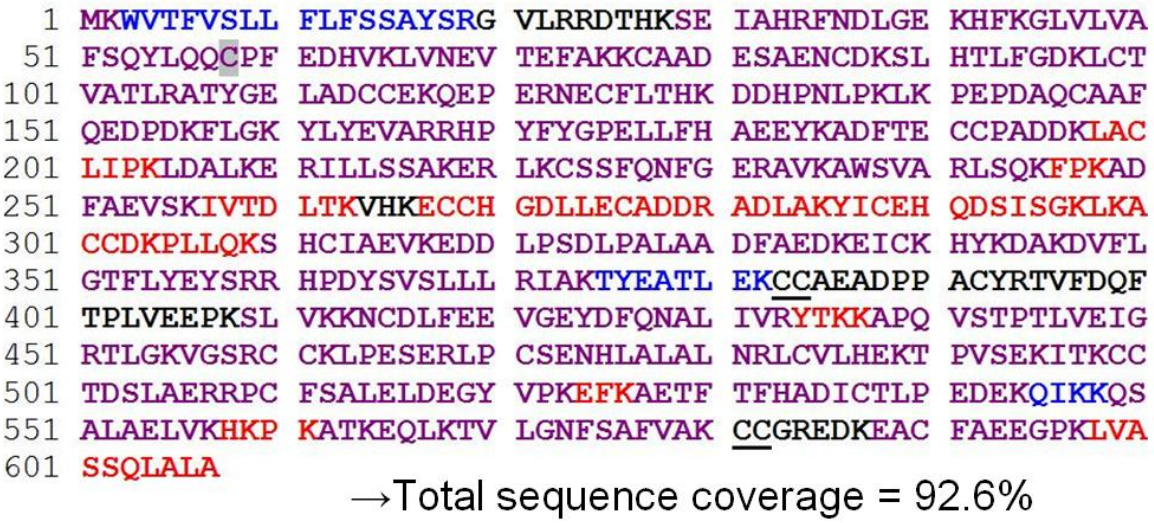

Figure 4-2 ESA_{red} sequence coverage
Blue = by trypsin, Red = by Lys-C (red), Purple = by both enzymes

Matrix Crystallization Comparison

DHB (2,5-Dihydroxybenzoic acid), CCA (α-Cyano-4-hydroxycinnamic acid), and DAN (1,5-Diaminonaphthalene) are the three matrices used throughout this particular series of experiments. Besides these three matrices, SA (Sinapic acid) and PNBA (2-amino-5-nitrobenzoic acid) are often used for MALDI-TOF MS measurement. It is

important to identify the appropriate matrix for each experiment in order to achieve the best ionization for the sample. Obtaining good crystallization of the matrix on the target plate is also a critical item along with finding an optimal sample-to-matrix ratio. Figures 4-3 though 4-6 show pictures of crystallized matrices used in the experiment.

DHB is often the matrix of choice for proteomics analyses such as the analysis of peptides after proteolytic digestion. It forms needle-like crystals especially around the corner of the sample area on the target plate. A minimum of 2 μL matrix/sample mixture was required in order to obtain the crystals shown in the Figure 4-3. A number of conditions can interfere with crystals formation including: if samples contain high salts or contaminants, if more than 2 μL of the sample mixture is applied or if increasing the sample to matrix ratio is required to see crystals formation. Crystallized DAN also forms meshed needles similar to the DHB but shorter and thinner (Figure 4-4). Unlike DHB, matrix to sample ratio was kept 1:1 for DAN in this experiment. Increasing the concentration of DAN caused problems in ionizing the sample for the MS measurement. Also using freshly prepared DAN was the most important point for the MS measurement. DAN starts oxidizing instantly after it is prepared as observed by a color change, from clear to dark grey. On the other hand, CCA is very stable once it is prepared in a saturated solution. It crystallizes to form small spheres (Figure 4-5) and mainly used for the analysis of smaller peptides such as synthesized peptides. The baseline for CCA stays flat compared to the DHB which is useful when looking for less abundant sample. A combination of DHB and DAN matrices formed a different type of crystals (Figure 4-6). However, even though the crystal was formed, no detectable peaks were observed during the MS measurement.

It is important to adjust the type, concentration, and the ratio of matrix in order to achieve better MS measurements. Knowing the characteristics of the sample and adjusting the matrix referring to the each sample leads to improved MS measurements.

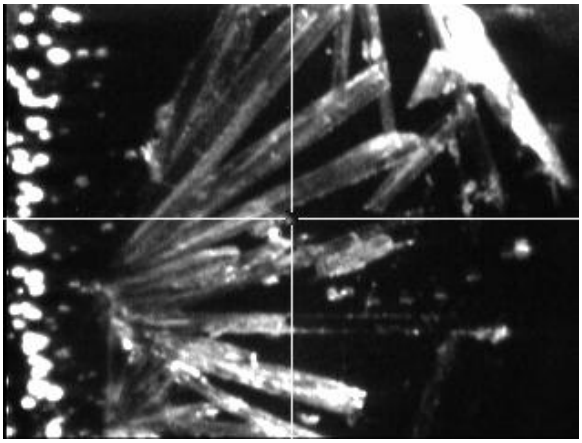

Figure 4-3 DHB crystal

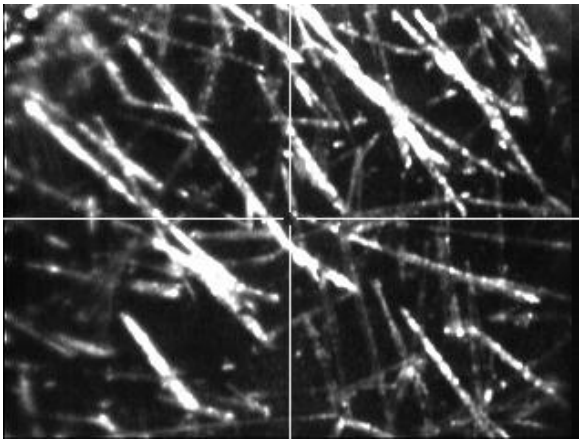

Figure 4-4 DAN crystal

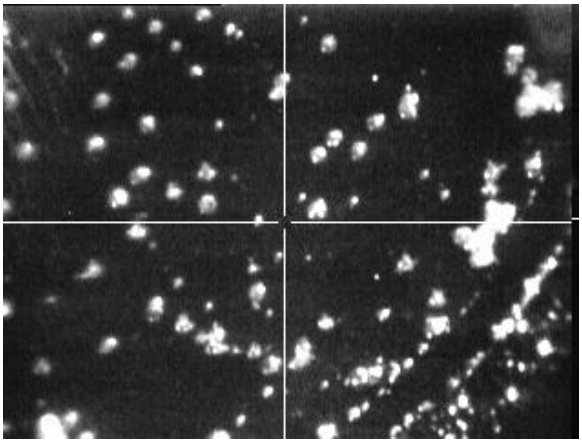

Figure 4-5 CCA crystal

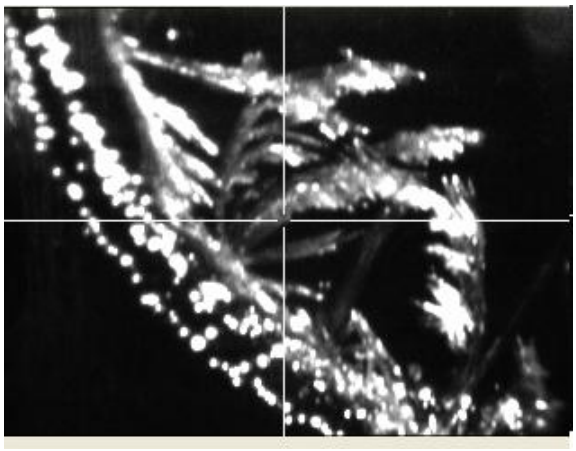

Figure 4-6 DAN:DHB crystal

PTM Investigation

Glycosylation Detection

As the glycoprotein stained gel image in the Figure 3-25 shows, no glycosylation in the native ESA_{ox} was detected. Using the glycosylation staining kit was a very quick and easy way to probe for the glycoproteins. The whole staining process took less than 2 hours and there were no additional or unusual chemicals required. However, this method can be unreliable considering the sensitivity of the staining kit. Horseradish peroxidase (HP), 5 μg was used as a positive control (included in the kit), and has numerous glycosylation sites. According to the instruction, as little as 0.16 μg of HP can be detected using this staining kit. This suggests that in cases with samples containing only a few glycosylation sites or samples at lower concentration, glycosylation might not be detected. Additionally, in order to assign the glycosylation sites, another method must be applied. MS has often been used for mapping glycosylation sites after glycosylated sample was pre-treated with a glycosidase. Mass differences observed between glycosylated and deglycosylated samples can be measured using MS to assign glycosylated sites to certain peptides. For more precise mapping, one can combine this technique with liquid chromatography (LC/MS), to more easily detect the glycosylation sites.[3]

Phosphorylation Detection

Phosphorylation of the ESA_{ox} sample was confirmed using the phosphoprotein gel staining kit. As shown in the Figure 3-26, the ESA band was stained with the kit containing the phosphate reactive fluorescent dye. The stained bands indicate phosphorylation. However, the stained ESA band was much narrower compared to the one stained by Coomassie blue. This could be indicating heterogeneous phosphorylation within the ESA sample. If the protein was phosphorylated evenly throughout each mole of protein, the thickness of the bands should be equal as observed in the Coomassie blue stained sample, but with less intensity. According to the information obtained from the Expasy Proteomics Server,[31] there are only three potential phosphorylation sites, Ser82, Tyr108, and Tyr163 within the 607 residues of ESA which explain less intensity in the staining of ESA band relative to the control HP sample. Phosphorylation of the protein is one of the most common protein modifications, particularly in the cellular signaling and regulation pathways. It is achieved by the variety of kinases or other organic/inorganic phosphate donating groups, such as ATP, present within the cell. Because it is one of the common PTMs, newer MS instruments can automatically detect the phosphorylation sites by simply recognizing the loss of 49 amu (atomic mass unit). For example, the MALDI-TOF MS instrument used in this experiment was often able to detect phosphorylation of the ESA sample during the sample analysis. This was automatically informed by the analysis program after the measured data was transferred. Because phosphate groups can be liberated from the amino acid residues by the addition of instrumental energy, the MS instrument was able to measure peaks of both phosphorylated and dephosphorylated fragments. The data analysis program could detect the 49 amu difference between those two peaks which leads to the automatic assignment of the phosphorylation.

CHAPTER 5 - Conclusion

Free cysteine and disulfide bonds present in the equine serum albumin (ESA) were characterized by using MALDI-TOF MS. After chemical modifications of the sample and matrix adjustments, majority of the expected disulfide bonded clusters were assigned which matched with the disulfide mapping reported from the x-ray crystallography experiment.[28] Total sequence coverage of native and reduced ESA reached to 73.6 % and 92.6 %, respectively when combining both the trypsin and Lys-C data. Acquired data from this experiment using MALDI-TOF MS confirmed the disulfide bonds sites reported by x-ray crystallography, but in a much more efficient way. Modifying the free cysteine in the ESA_{ox} was an important step before proceeding to the further analysis. Labeling the free cysteine with florescent dye (6-IAF) was effective because not only was the reaction irreversible, but also because it was visible to unaided eyes. Due to a large number of disulfide bonds present in the native ESA_{ox}, some of the digested fragments were still cross-linked after denaturation. However, using DAN, a new type of reducing matrix, those disulfide bonds were reduced instantly on the sample plate with just the addition of MALDI laser energy. Additionally, since the sample was separated, labeled and digested within the gel, more MS peaks with higher intensities and sensitivities were obtained. These facts make it possible to achieve higher sequence coverage of the protein which is important for the characterization of the protein sequences along with the assignment of all the PTM sites. Figure 5-1 is a summary of an optimized protocol based on our studies. For the disulfide bond mapping using MALDI-TOF MS, it is recommended to follow such protocol for obtaining better results.

MALDI-TOF MS is one of the common protein analysis tools that can provide extensive information in a short time frame while keeping the sample in its native state. A protein containing disulfide bonds is poorly ionized with MALDI which leading to a decrease in the sensitivity of the instrument. However, by using a new type of matrix and performing a sample digestion with suitable proteolytic enzymes, it is possible to overcome such problems. Nowadays, the investigation of PTMs has been a big focus in the field of proteomics. Even though different separation techniques or ionization

methods, capillary electrophoresis and liquid or gas chromatography, are combined to the MS, idealization and optimization of such will be the challenge that we all must overcome in the future.

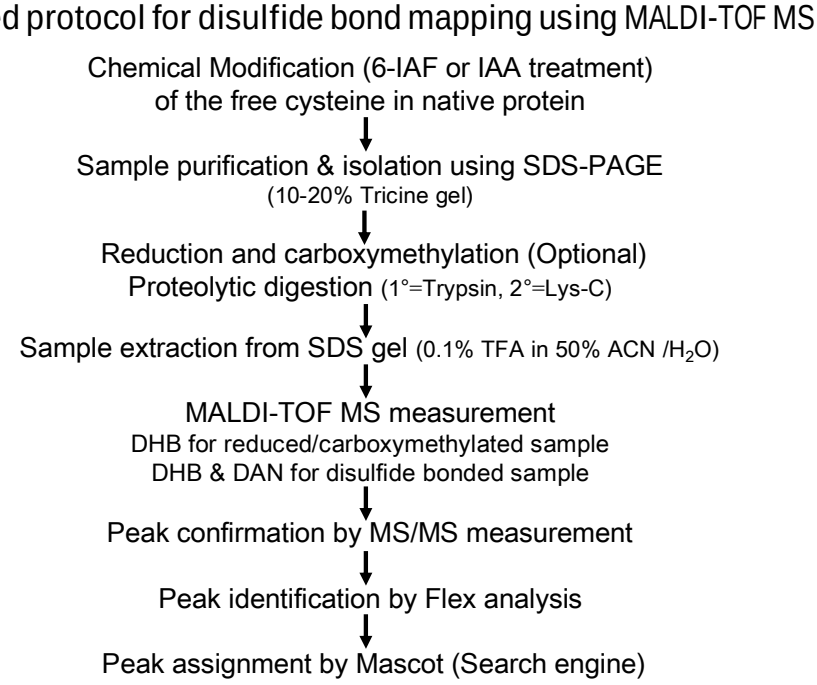

Figure 5-1 Summary of optimized protocol for disulfide bond mapping

List of Abbreviations

ES, equine serum

ESA, equine serum albumin

ESA_{ox}, native (oxidized) equine serum albumin

ESA_{red}, reduced equine serum albumin

HSA, human serum albumin

BSA, bovine serum albumin

6-IAF, 6-Iodoacetamido fluorescein

IAA, Iodoacetamide

HP, horseradish peroxidase

TI, soybean trypsin inhibitor

GDH, glutamic dehydrogenase

DHB, 2,5-Dihydroxybenzoic acid

DAN, 1,5-Diaminonaphthalene

CCA, α-Cyano-4-hydroxycinnamic acid

amu, atomic mass unit

a.u, arbitrary unit

References

1. Zhou F, Xue Y, Yao X, Xu Y. 2006. A General User Interface for Prediction Servers of Proteins' Post-Translational Modification Sites. *Nature protocols* 1(3): 1318.
2. Walsh G, Jefferis R. 2006. Post-Translational Modifications in the Context of Therapeutic Proteins. *Nature biotechnology* 24(10): 1241.
3. Yen TY, Macher BA. 2006. Determination of Glycosylation Sites and Disulfide Bond Structures using LC/ESI-MS/MS Analysis. *Methods in Enzymology* 415(7): 103.
4. Zeng R, Xu Q, Shao X, Want K, Xia Q. 2001. Determination of the Disulfide Bond Pattern of a Novel C-Type Lectin from Snake Venom by Mass Spectrometry. *Rapid communications in mass spectrometry* 15(23): 2213.
5. Gleiter S, Bardwell JC. 2008. Disulfide Bond Isomerization in Prokaryotes. *Biochimica et biophysica acta*. 1783(4): 530.
6. Kveder M, Kriško A, Pifat G, Steinhoff H. 2003. The Study of Structural Accessibility of Free Thiol Groups in Human Low-Density Lipoproteins. *Biochimica et biophysica acta*. 1631(3): 239.
7. Varughese KI, Ahmed FR, Carey PR, Hasnain S, Huber CP, Storer AC. 1989. Crystal Structure of Papin-E-64 Complex. *Biochemistry* 28(3): 1330.

8. Xu H, Zhang L, Freitas MA. 2008. Identification and Characterization of Disulfide Bonds in Proteins and Peptides from Tandem MS Data by use of the MassMatrix MS/MS Search Engine. *Journal of proteome research* 7(1): 138.
9. Mamathambika BS, Bardwell JC. 2008. Disulfide-Linked Protein Folding Pathways. *Annual Review of Cell and Developmental Biology* 24(211): 35.
10. Gilbert HF. 1995. Thiol/Disulfide Exchange Equilibria and Disulfide Bond Stability. *Methods in Enzymology* 251(8): 2.
11. Kobayashi A, Kang M, Okawa H, Ohtsuji M, Zenke Y, Chiba T. 2004. Oxidative Stress Sensor Keap1 Functions as an Adaptor for Cul3-Based E3 Ligase to Regulate for Proteasomal Degradation of Nrf2. *Molecular and cellular biology* 24(16): 7130.
12. Dinkova-Kostova AT, Holtzclaw WD, Cole RN, Itoh K, Wakabayashi N, Katoh Y. 2002. Direct Evidence that Sulfhydryl Groups of Keap1 are the Sensors Regulating Induction of Phase 2 Enzymes that Protect Against Carcinogens and Oxidants. *Proceedings of the National Academy of Sciences of USA* 99(18): 11908.
13. Wittenberg G, Danon A. 2008. Disulfide Bond Formation in Chloroplasts - Formation of Disulfide Bonds in Signaling Chloroplast Proteins. *Plant science* 175(4): 459.
14. Zhang M, Kaltashov I. 2006. Mapping of Protein Disulfide Bonds using Negative Ion Fragmentation with a Broadband Precursor Selection. *Analytical chemistry* 78(14): 4820.

15. Horn M, Baudyš M, Voburka Z, Kluh I, Vondrašek J, Mareš M. 2002. Free-Thiol Cys331 Exposed during Activation Process is Critical for Native Tetramer Structure of Cathepsin C (Dipeptidyl Peptidase I). *Protein science* 11.(4): 933.
16. Wu W, Huang W, Qi J, Chou Y, Torng E, Watson JT. 2004. 'Signature Sets', Minimal Fragment Sets for Identifying Protein Disulfide Structures with Cyanylation-Based Mass Mapping Methodology. *Journal of proteome research* 3(4): 770.
17. Dobson CM, 2003. Protein Folding and Misfolding. *Nature* 426(6968): 884.
18. Narayan M, Welker E, Zhai H, Han X, Xu G, McLafferty FW. 2008. Detecting Native Folds in Mixtures of Proteins that Contain Disulfide Bonds. *Nature biotechnology* 26(4): 427.
19. Singh P, Shaffer SA, Scherl A, Holman C, Pfuetzner RA, Goodlett DR. 2008. Characterization of Protein Cross-Links Via Mass Spectrometry and an Open-Modification Search Strategy. *Analytical chemistry* 80(22): 8799.
20. Yoo C, Pal M, Miller FR, Barder TJ, Huber C, Lubman DM. 2006. Toward High Sequence Coverage of Proteins in Human Breast Cancer Cells using on-Line Monolith-Based HPLC-ESI-TOF MS Compared to CE MS. *Electrophoresis* 27(11): 2126.
21. Suckau D, Resemann A, Schuerenberg M, Hufnagel P, Franzen J, Holle A. 2003. A Novel MALDI LIFT-TOF/TOF Mass Spectrometer for Proteomics. *Analytical and bioanalytical chemistry* 376(7): 952.

22. John H, Forssmann WG, 2001. Determination of the Disulfide Bond Pattern of the Endogenous and Recombinant Angiogenesis Inhibitor Endostatin by Mass Spectrometry. *Rapid communications in mass spectrometry* 15(14): 1222.

23. Krause E, Wenschuh H, Jungblut PR. 1999. The Dominance of Arginine-Containing Peptides in MALDI-Derived Tryptic Mass Fingerprints of Proteins. *Analytical chemistry* 71(19): 4160.

24. Padliya ND, Wood TD. 2004. A Strategy to Improve Peptide Mass Fingerprinting Matches through the Optimization of Matrix-Assisted Laser desorption/ionization Matrix Selection and Formulation. *Proteomics* 4(2): 466.

25. Fukuyama Y, Iwamoto S, Tanaka K. 2006. Rapid Sequencing and Disulfide Mapping of Peptides Containing Disulfide Bonds by using 1,5-Diaminonaphthalene as a Reductive Matrix. *Journal of mass spectrometry* 41(2): 191.

26. Foley SF, Sun Y, Zheng TS, Wen D. 2008. Picomole-Level Mapping of Protein Disulfides by Mass Spectrometry Following Partial Reduction and Alkylation. *Analytical biochemistry* 377(1): 95.

27. Shaklai N, Garlick RL, Bunn HF. 1984. Nonenzymatic Glycosylation of Human Serum Albumin Alters Its Conformation and Function. *The Journal of Biological Chemistry* 259(6): 3812.

28. Ho JX, Holowachuk EW, Norton EJ, Twigg PD, Carter DC. 1993. X-ray and Primary Structure of Horse Serum Albumin (*Equus caballus*) at 0.27-nm Resolution. *European journal of biochemistry* 215(1): 205.

29. Petitpas I, Petersen CE, Ha CE, Bhattacharya AA, Zunszain PA, Ghuman J. 2003. Structural Basis of Albumin-Thyroxine Interactions and Familial Dysalbuminemic Hyperthyroxinemia. Proceedings of the National Academy of Sciences of USA 100(11): 6440. (PDB:1hk1)
30. UniProtKB/Swiss-Prot **P02768** (ALBU_HUMAN)
31. UniProtKB/Swiss-Prot **P35747** (ALBU_HORSE)
32. UniProtKB/Swiss-Prot **P02769** (ALBU_BOVIN)
33. Burns JA, Butler JC, Moran J, Whitesides GM. 1991. Selective Reduction of Disulfides by Tris(2-Carboxyethyl)Phosphine. *Journal of organic chemistry* 56(8): 2648.
34. Thermo Fisher Scientific Inc. 2009. TCEP·HCl Tris(2-Carboxyethyl) phosphine Hydrochloride. Thermo Fisher Scientific Protein Research Products. <http://www.piercenet.com/Products/Browse.cfm?fldID=02051012>
35. Caldwell HK, Lee HJ, Macbeth AH, Scott Young III W. 2008. Vasopressin: Behavioral Roles of an "Original" Neuropeptide. *Progress in Neurobiology* 84(1): 1.

Appendix A - Trypsin digested fragments

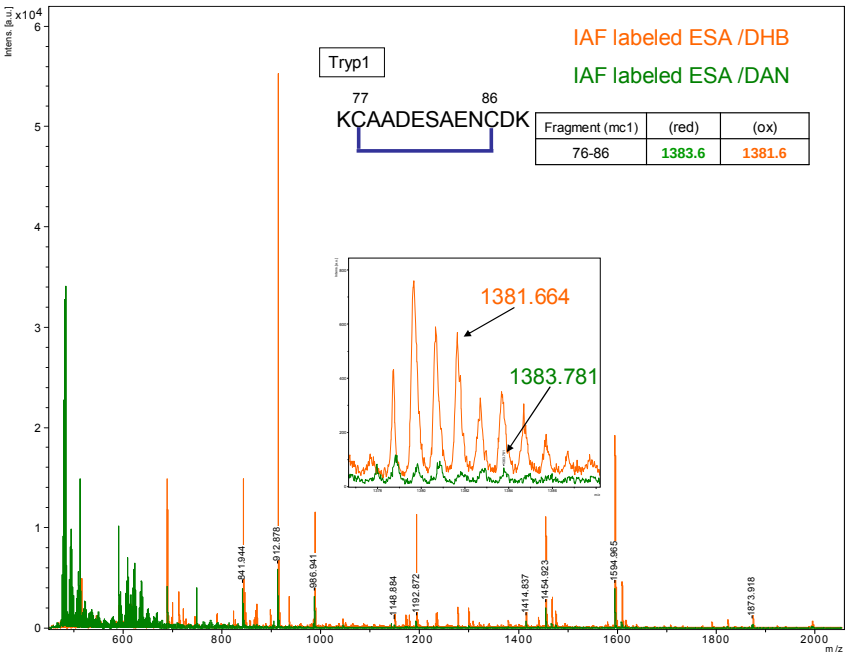

Figure A-1 MS spectra of Tryp1 (1 miscleavage)

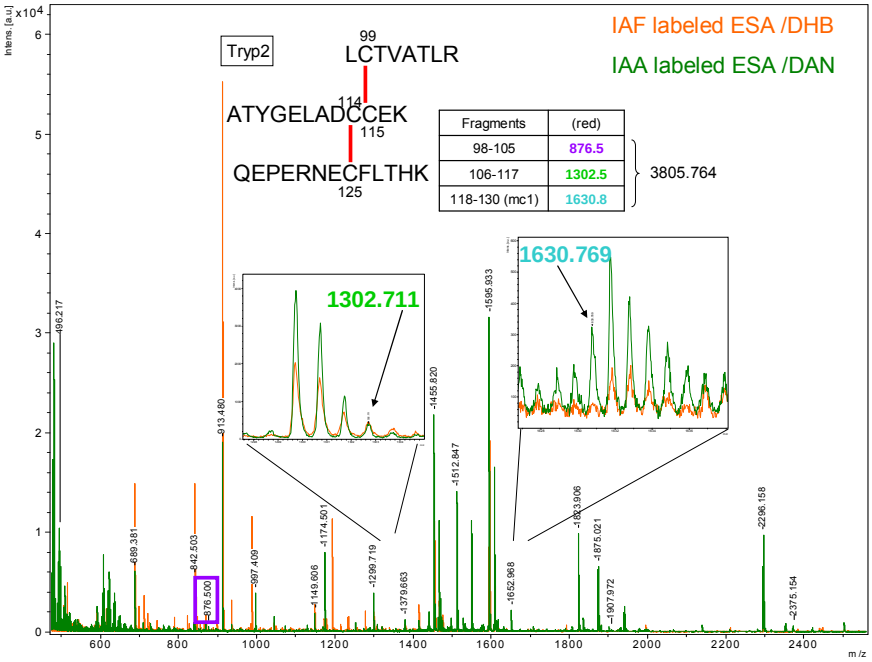

Figure A-2 MS spectra of Tryp2

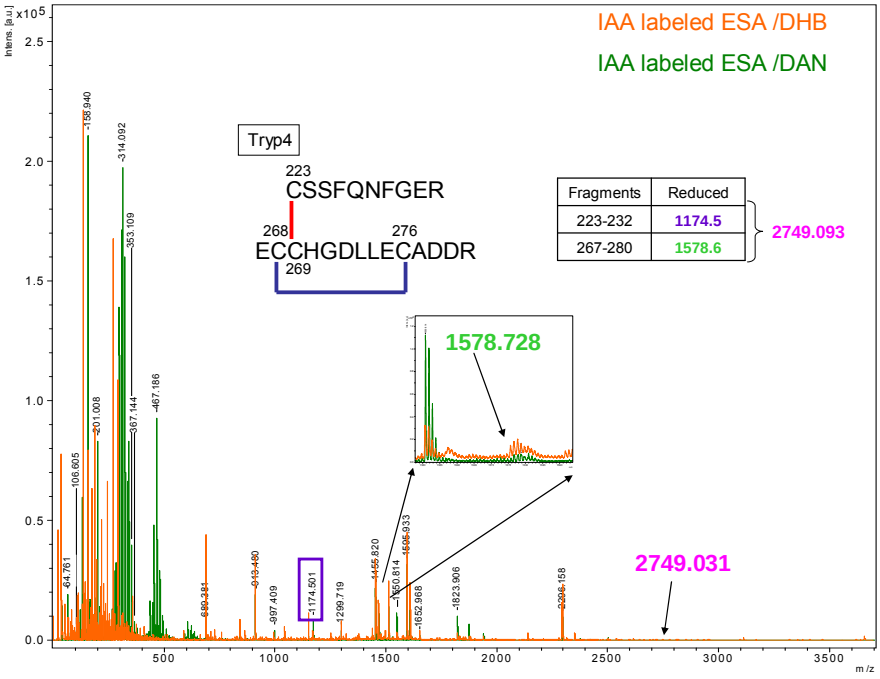

Figure A-3 MS spectra of Tryp4

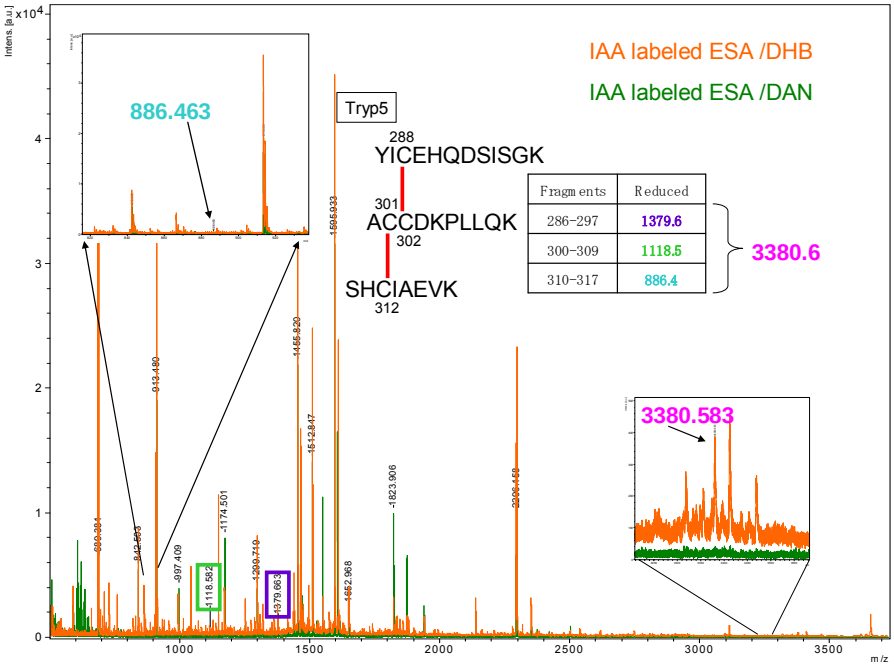

Figure A-4 MS spectra of Tryp5

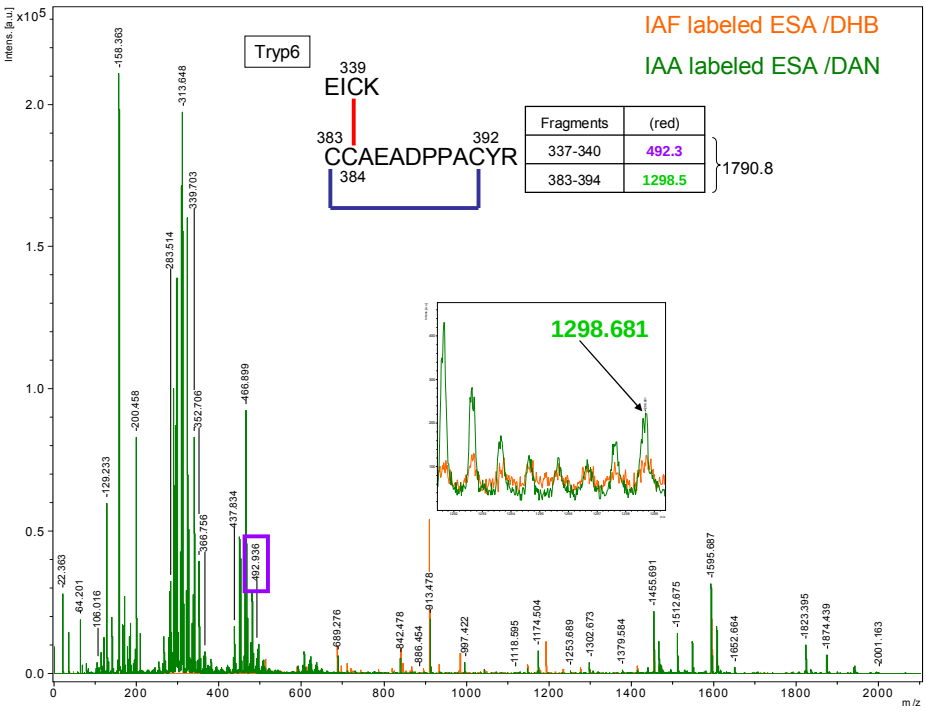

Figure A-5 MS spectra of Tryp6

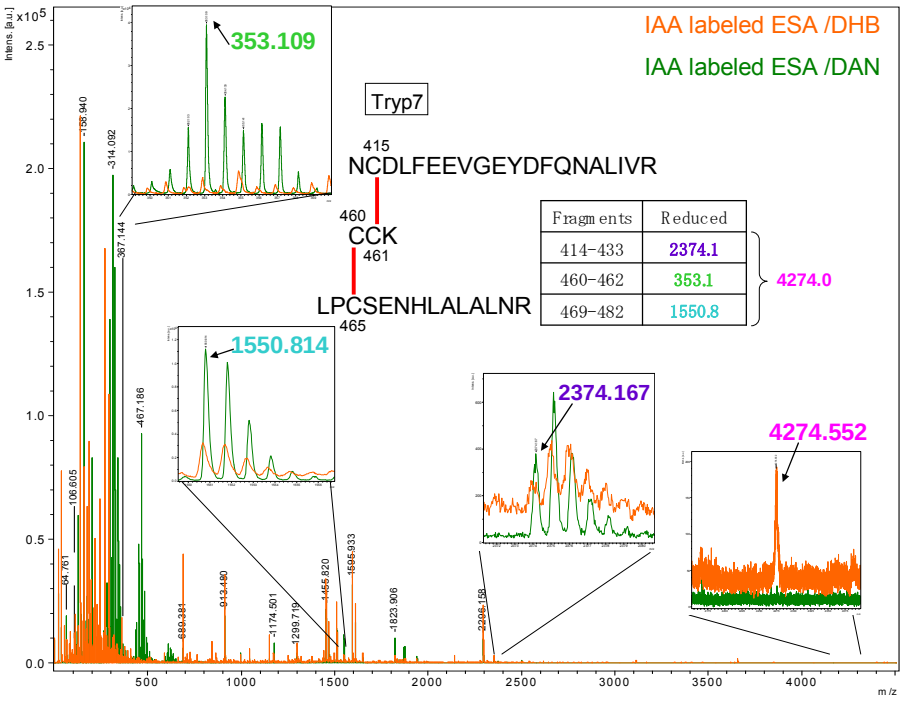

Figure A-6 MS spectra of Tryp7

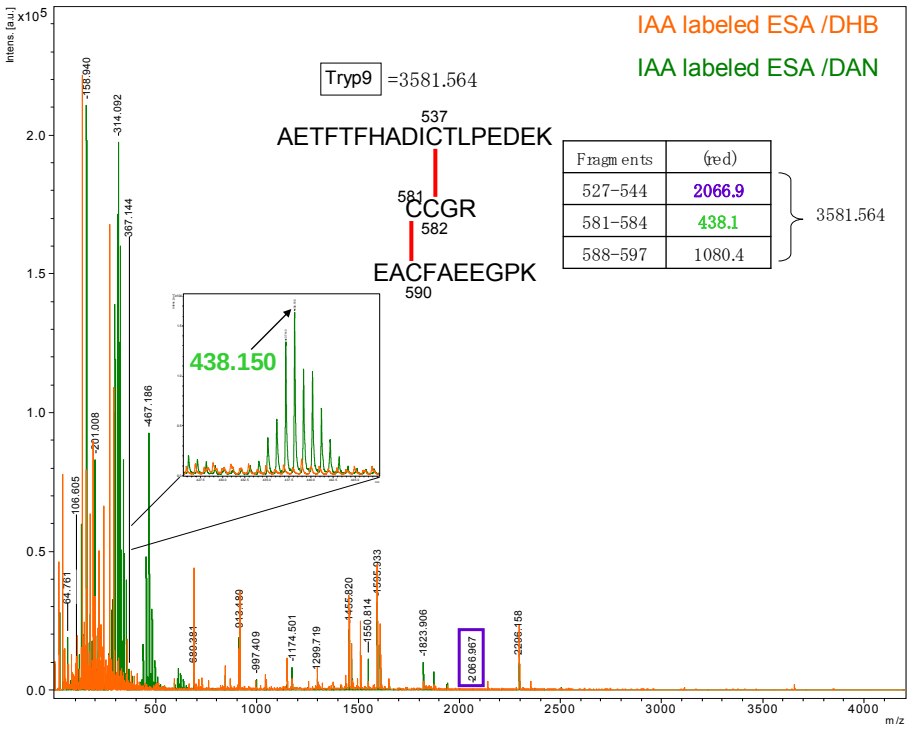

Figure A-7 MS spectra of Tryp9

Appendix B - Lys-C digested fragments

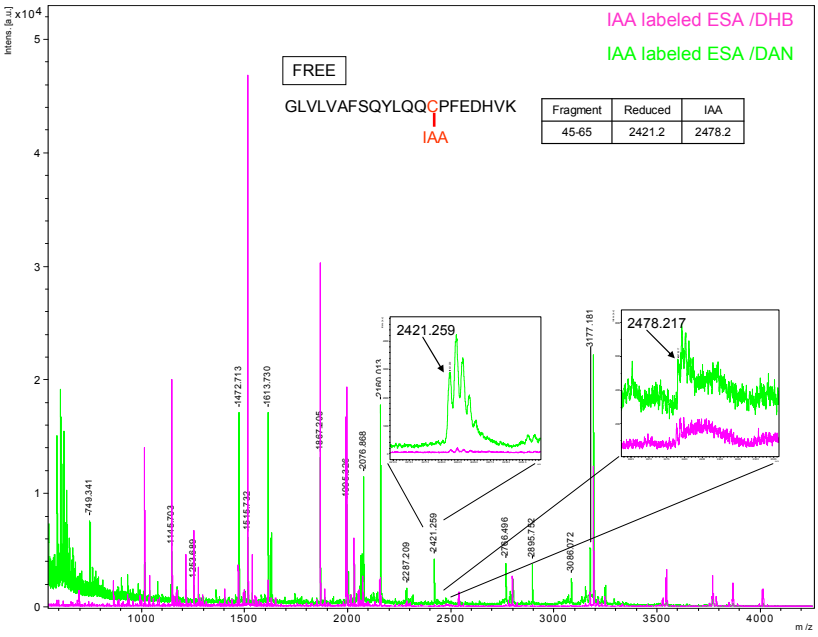

Figure B-1 MS spectra of IAA labeled ESA_{ox}

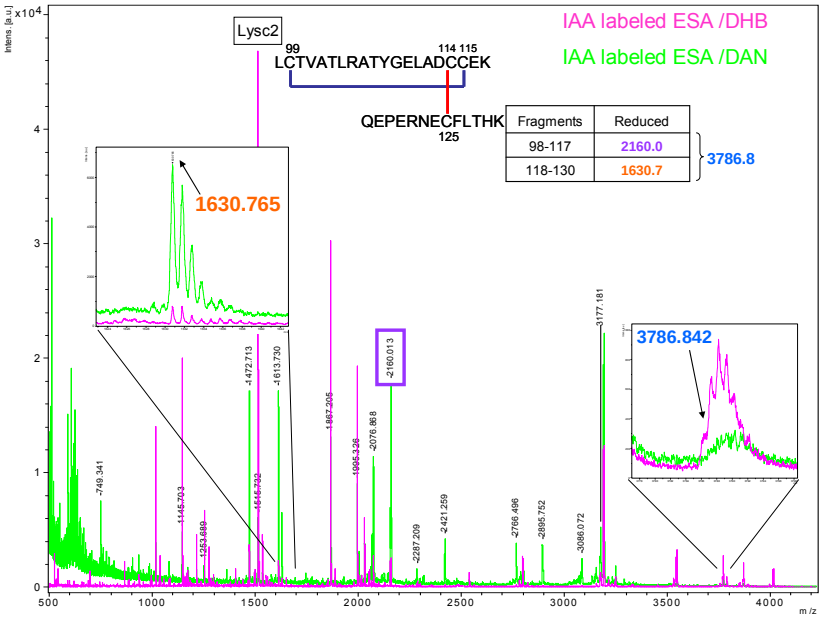

Figure B-2 MS spectra of Lysc2

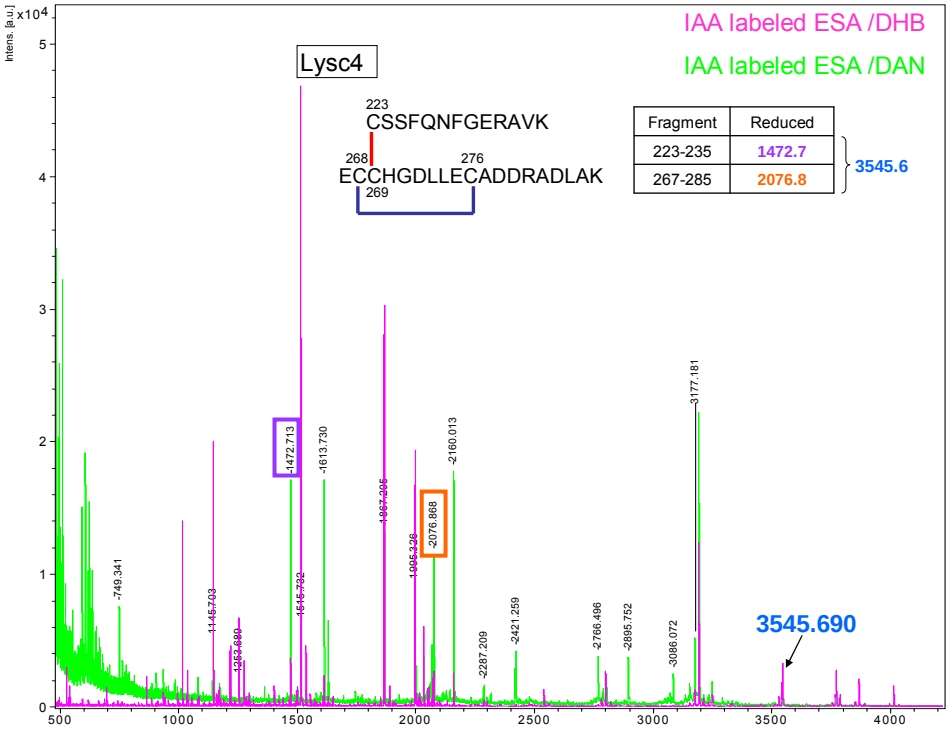

Figure B-3 MS spectra of Lysc4

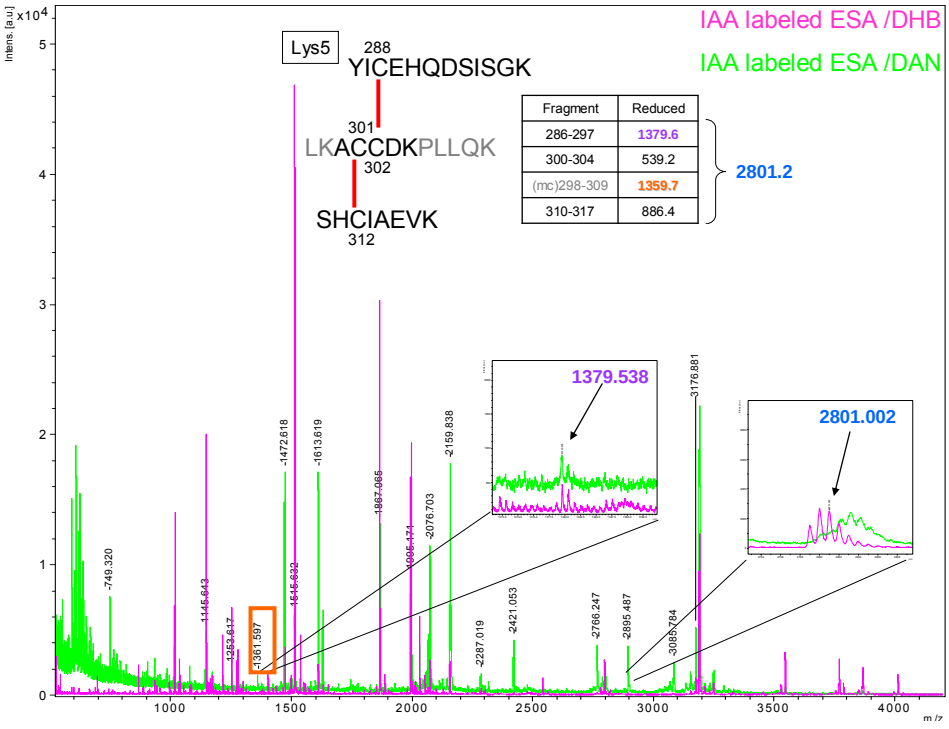

Figure B-4 MS spectra of Lysc5

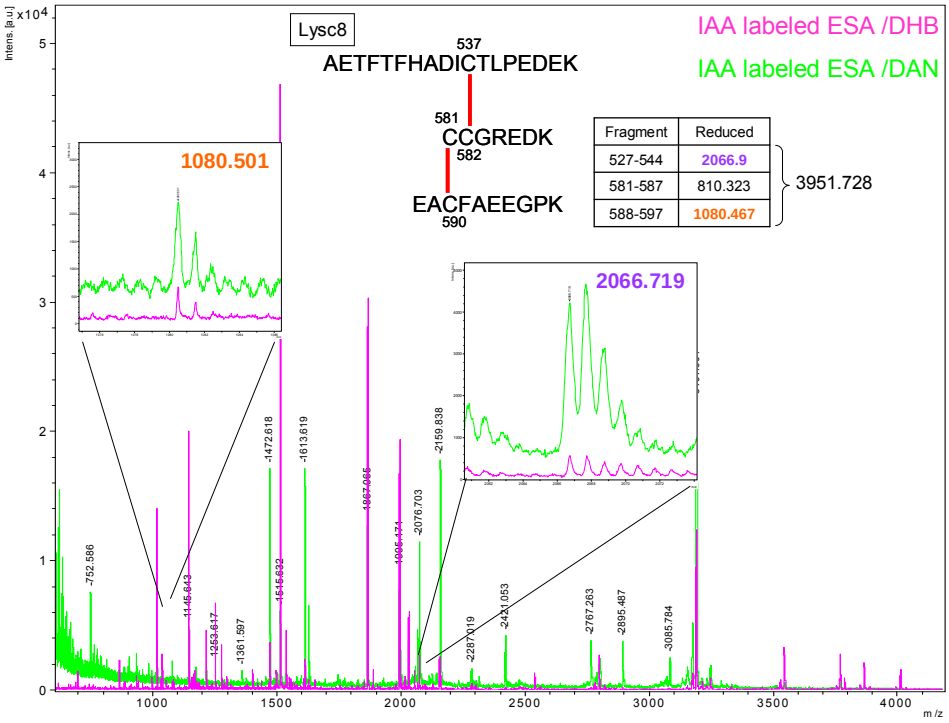

Figure B-5 MS spectra of Lysc8