July 2015

Characterization of Protein-Protein Interactions for Therapeutic Drug Design Utilizing Mass Spectrometry

Alex J. Johnson Mr
University of Massachusetts Amherst

CHARACTERIZATION OF PROTEIN-PROTEIN INTERACTIONS FOR THERAPEUTIC DRUG DESIGN UTILIZING MASS SPECTROMETRY

A Thesis Presented

by

Alex Johnson

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

Master of Science

May 2015

Molecular and Cellular Biology Graduate Program

CHARACTERIZATION OF PROTEIN-PROTEIN INTERACTIONS FOR THERAPEUTIC DRUG DESIGN UTILIZING MASS SPECTROMETRY

A Thesis Presented

by

ALEX JOHNSON

Approved in style and content by:

_____
Igor Kaltashov, Chair

_____
Lynmarie Thompson, Member

_____
Richard Vachet, Member

_____
Barbara Osborne, MCB Director

ACKNOWLEDGEMENTS

I would like to acknowledge and thank the people who have supported me throughout my higher educational career. I had the privilege of working along side Son Nguyen in the lab as he welcomed me aboard this project. Son showed me the ropes as we progressed through this tedious and exciting research. I want to thank him specifically for his patience when experiments went astray and for making each day in the lab an enjoyable one. I wish him and his new family the best of luck. Igor Kaltashov has been my mentor for many years now. When I first met him, I was ending my sophomore year after deciding to start my chemistry major and I needed his permission to enroll in his Analytical course, since I did not have the prerequisites needed at the time. After walking around the lab trying to find his office, I saw amazing instruments that caught my eye and stuck in my mind when it came time to choosing a senior research project. I want to thank Igor for taking me and introducing me to the wide world of analytical chemistry. Guanbo Wang helped me along with my senior research and taught me the ins-and-outs of conducting research in a lab. I want to thank Guanbo for making my senior research memorable and for making my decision to go back to UMass Amherst for my graduate career quite easy. Once again Igor took me in and continued to be a great help. I would like to thank my committee members, Professor Thompson and Professor Vachet, for their willingness to be a part of my time at UMass. I would like to thank Professor Osborne for introducing me to the greatest journal club at UMass, and that is excluding the delicious pizza every class. Lastly I would like to thank all of my friends in the Kaltashov lab, MCB program, and at the University of Massachusetts. You have played a role in shaping my future and I wish you all the best with your future endeavors.

ABSTRACT

CHARACTERIZATION OF PROTEIN-PROTEIN INTERACTIONS FOR THERAPEUTIC DRUG DESIGN UTILIZING MASS SPECTROMETRY

MAY 2015

ALEX JOHNSON, B.S., UNIVERSITY OF MASSACHUSETTS AMHERST

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Igor A. Kaltashov

The number of transferrin based therapeutics progressing to clinical trials remains disappointingly small despite promising capabilities of transporting therapeutic payloads to cancer cells and across the blood brain barrier. This meager success record is largely due to the complexity and heterogeneity of all protein conjugation products that generates difficulties for their analytical characterization. Discussed in this work, transferrin is conjugated to lysozyme as a model therapeutic to deliver this bacteriostatic protein to target central nervous system infections. In this work ESI- and MALDI-MS were used to characterize the modification sites at lysine residues in hopes of characterizing heterogeneity within the conjugate. Identification and quantization of modification sites using MS on tryptic digested samples proved difficult with poor signal to noise ratios and missing peptide fragments. The use of an ^{18}O labeling method that exchanges both C-terminal oxygen atoms with ^{18}O provided more reliable results, but still proved difficult to observe all needed peptide fragments. MALDI-MS allowed for verification of ESI-MS results, but was found unhelpful with full characterization due to abundant overlapping of isotopic labeled peaks. Hoping to create an ideal 1:1 binding ratio between the two proteins, a site-specific modification method using kinetically controlled conditions was used and was confirmed that the method, although capable of producing 1:1 conjugated species, actually created different isomers with separate binding frequencies at each lysine. Online-IEC helped with the identification of isomers and started the initial work of correlating modification sites with bioactivity of the proteins. It was determined that lysozyme has a high chance of being modified at lysine 33 and 116, with a possibility of also being highly modified at lysine 97. More work is needed to complete the characterization, especially with transferrin, but the experimental approaches developed in this work prove to be promising. This work aims at delivering an optimized framework for analytical characterization of protein and antibody conjugates to guide the development of future biopharmaceuticals.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
BACKGROUND/INTRODUCTION

A quite extensive library of promising novel therapeutic proteins and antibodies has been developed over the years. Improved ways in which to deliver these innovative agents has also been in the forefront of pharmaceutical research. Delivery of drugs can be enhanced by either changing the formulation or by changing the molecular structure of the medicine. Formulation manipulation may include colloidal systems, such as microspheres, and continuous release mechanisms, such as osmotic pumps, which can provide assistance compared to standard formulations. Modifying the molecular structure of a therapeutic, such as chemically adhering poly(ethylene glycol) [PEG] to the drug, can strengthen and protect the essential drug properties.[1] While becoming potent and specific therapeutic agents, intrinsic limitations have delayed these novel drugs advancements. These shortcomings include susceptibility to destruction by proteolytic enzymes and immunogenicity.[1,2,3,4] In order to overcome these limitations, PEG has been used to shield against these effects and create a conjugated protein.

1.1 Polyethylene Glycol

Frank F. Davis, a Professor of Biochemistry at Rutgers University, first conceived PEGylation in the late 1970s. As Davis was between grants, he became interested in developing a procedure whereby selected bioactive proteins could be utilized for human therapy. Davis decided to spend months in the university libraries to quench this new interest, until finally after reading a medical article his thought light bulb turned on.[5, 6] PEGylation has to be equipped with a reactive end group in order to attach to another

molecule. This attachment is achieved by targeting accessible amino residues, such as lysine or the N-terminus on the other molecule, and acts as a linker between two proteins.[4,7] The shielding effect of the polymeric chains surrounding the protein of interest provides steric hindrance, preventing degrading enzymes from attacking and destroying the hopeful drug conjugate. PEG is FDA approved for usage as a vehicle or as a base in foods, cosmetics, and pharmaceuticals. PEG shows little toxicity and is disposed of intact by the body either through the kidneys, when PEG is less than 30 kDa, or in feces, when PEG is greater than 20 kDa.[3]

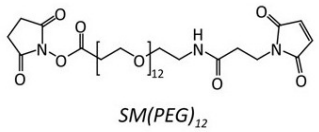


Figure 1: Illustration of the structure of SM(PEG)12, which acts as the linker in a conjugate.

Linking a carrier protein to a drug by using PEG introduces new challenges based on the heterogeneity and complexity of the conjugation, leading to difficulties in analytical characterization. The association of the two proteins can cause negative conformational changes, resulting in the interference with receptor binding, which can lead to a decrease in effectiveness in bioactivity. If binding a protein to another protein near either of their activation sites, the molecule may fail to be able to perform its normal function. One problem that may start to arise by attaching proteins using PEG is the possibility of attaching too many. This brings about the concept of DAR, which stands for drug-to-antibody ratio. This ratio is as it sounds; the average number of drug molecules covalently attached to the antibody or protein of interest. This ratio is important because drug designers need to anticipate how many drugs are attached to their

carriers to keep the efficacy and patient safety in check. It has been noted that the length, location, and number of PEG's attached to a specific protein can change its biological activity from 6% to 40%.[7,8] Understanding the dynamics and characterization of conjugated therapeutics is extremely important and essential in optimization of drug design.

1.2 Transferrin

There are many creative combinations being developed in therapeutic research these days. The research described in this thesis is specifically looking into the combination of Transferrin (Tf) conjugated to Lysozyme (Lz) via PEG linker as a model drug. Tf is a promising drug carrier due to its ability to safely deliver metals or therapeutic proteins throughout the body and across physiological barriers. Transferrin is a glycoprotein containing 679 amino acid residues and has a molecular weight of 79 kDa, remaining stable by its 19 intra-chain disulfide bonds. Transferrin has two evolutionary related lobes, the N-lobe (336 amino acids) and the C-lobe (343 amino acids), which are linked together by a short spacer sequence. Each of these lobes contains two domains comprising of a series of α-helices, which overlay a central β-sheet backbone. The interactions of these domains form a deep, hydrophilic metal ion-binding site.[9]

Iron is the second most abundant metal in the earths crust and plays an extremely important role in living organism. Iron ions have the ability to strongly bind oxygen and play a central role in DNA replication. However, free iron can be toxic, promoting free radical formation resulting in oxidative damage to tissues. The two lobes of transferrin are capable of binding one iron atom; therefore each transferrin molecule (apo-Tf) can transport one (monoferric Tf) or two (diferric Tf) iron atoms. Transferrin is vital in living

organisms to ensure that iron is transported in a redox-inactive form since ferric iron (Fe^{3+}) is the predominant state under aerobic conditions.[10] The binding site for both the N- and C- lobes have four conserved amino acids (N-lobe – Asp-63, Tyr-95, Tyr-188 and His-249) arranged in an octahedral arrangement that, with the help of two oxygen molecules donated by a carbonate molecule and surrounding amino acids, stabilize the iron atom.[9] Surrounding amino acids Lys-206 and Lys-296, which are located on opposite domains of the N-lobe, are shown to form hydrogen bonds when in the "closed" iron-bound state and when the pH is reduced to 5.5, the bond is broken and the iron is released from transferrin. [9] Iron loaded transferrin will seek out its particular transferrin receptor (TfR) on the surface of cells.

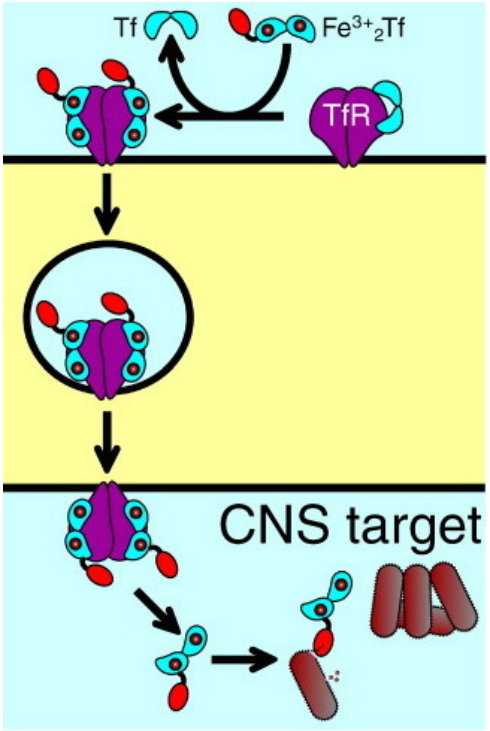


Figure 2: Illustration of the RMT pathway. Iron binds to Tf in a diferric fashion in order to bind to its receptor. Once bound, the complex is engulfed in an endocytic vesicle and fuses with an endosome, where it is then transcytosed to the brain.[11]

1.3 Transferrin Receptor

The transferrin receptor is a cell membrane associated glycoprotein involved in the cellular uptake of iron by internalization of Tf through transcytosis. The transferrin receptor is ubiquitously expressed at low levels on normal cells and expressed at greater levels on cells with a high proliferation rate such as the basal epidermis, intestinal epithelium, and cancerous cells. Much Tf therapeutic research is based on attaching toxins to Tf is hopes of targeting the cancerous cells that show high levels of Tf activity.[12] Numerous studies have shown elevated expression of TfR on cancerous cells compared to normal cells, which could be explained by the need for iron as a cofactor of the ribonucleotide reductase enzyme involved in DNA synthesis of rapidly dividing cells.[13] There are at least two types of transferrin receptors that Tf can bind to. Transferrin receptor 1 (TfR1) is expressed on a wide array of cells ranging from red blood cells to erythroid cells to the blood-brain barrier. Transferrin receptor 2 (TfR2) is expressed as two transcripts that are mostly expressed on the liver cells and other variety of cells at low levels.[9] The research conducted will focus on the intake of Tf using TfR1's across the blood-brain barrier.

1.4 Blood-Brain Barrier

The blood-brain barrier (BBB)[14,15] protects the central nervous system by regulating brain homeostasis through the control of entry of compounds into the brain. The barrier is formed by the endothelial cells of the brain capillaries and acts as a natural defense restricting clearance of harmful compounds to the brain, while allowing nutrients essential for normal metabolism to reach brain cells. Without this barrier, uncontrolled

neural activity would run rampant with the frequent small fluctuations brought on by changing levels of substances such as hormones and ions.[16] Virtually, 100% of large molecules (mw > 1 kDa) and approximately 98% of all small molecules are denied passage through the blood-brain barrier.[17] This fact is one of the few aspects limiting the development of new central nervous system (CNS) medicines and is quite discerning since one out of every three individuals will acquire a CNS condition during their lifetime.[18] Current methods of treating CNS conditions are invasive, dangerous, and sometimes ineffective. Such treatments are direct intracranial injection, intraventricular administration, and BBB disruption.[14] The research conducted concentrates on using a non-invasive procedure to help treat CNS conditions by utilizing the transport systems that allow larger molecules, such as transferrin, to enter into the brain.

1.5 Receptor-Mediated Transcytosis

Receptor-mediated transcytosis (RMT) allows the transport of larger molecules to pass through the blood-brain barrier by means of vesicular trafficking mechanisms of the endothelium. This transport system, depicted in figure 2, can be utilized to carry conjugated therapeutic agents non-invasively into the brain that would otherwise be excluded by the blood-brain barrier. Iron is one of the essential nutrients needed for a healthy brain. First coming into contact with transferrin to form diferric Tf, this fully closed state Tf seeks out a transferrin receptor for binding on the apical plasma membrane. Once Tf is bound to the TfR1, the encapsulation of the Tf-TfR complex in a clatherin-coated vesicle is initiated by the invagination of the membrane. With normal endocytosis, the decrease in pH in the endosome triggers a conformational change in

transferrin, subsequently releasing the iron and increasing the affinity of apo-Tf for the receptor. The Tf-TfR complex is returned to the cell surface, releasing Tf to once again find iron to repeat the process.[13,14] With transcytosis, the receptor is not recycled back to the cell surface, but instead passes through the BBB along with the conjugated Tf. The process of transcytosis is not completely understood, but experiments have shown that up to 50% of transferrin receptor-mediated uptake is carried out by transcytosis.[19] While this process sounds ideal for a non-invasive therapeutic delivery system, Tf is a difficult delivery vector since its receptors are frequently saturated with endogenous transferrin that circulates throughout the bloodstream at a concentration of 25mM; meaning competition for the receptor would take place between a drug conjugated Tf and natural Tf. There could be ways of inhibiting the release of Tf, thereby enhancing the transcytosis process. It has been noted that a low fraction of about 1-4% of injected doses actually reach the brain, but this low percentage is still sufficient enough to treat CNS condition.[14]

1.6 Lysozyme

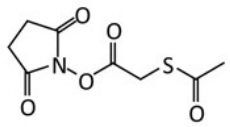
A particular protein of interest in which to bind to transferrin for a model drug conjugation is lysozyme (Lz). Lysozyme is alluring in this linkage with Tf for its bacteriostatic properties against gram-positive and gram-negative bacteria.[20] Lysozyme is a lytic enzyme discovered in 1921 by Sir Alexander Fleming, the same scientist who received the 1945 Nobel Prize for the introduction of penicillin. Lysozyme was the first enzyme containing the varied amino acids to be sequenced in 1963 and the first whose x-ray crystallographic analysis and reaction mechanism was achieved.[21] Being the first is

why lysozyme is one of the most characterized and modeled enzymes in biochemistry. Lysozyme can be found throughout the body, with particularly large concentrations in human serum; but there is no lysozyme found in cerebrospinal fluid (CSF).[22, 23, 24] Studies suggest that detectable levels of lysozyme in CSF correlate to assorted CNS pathologies when the BBB is disrupted and becomes leaky, allowing free access for species otherwise denied passage.[22, 25] There are almost 100 lysozymes sequenced, but the major type used in research today is chicken-type, which comes from hen egg white and is easily obtained in high quantities. Lysozyme is a small compact protein weighing 14,307 Da consisting of 129 amino acid residues held together mostly by four disulfide links. The structure has a mix of alpha and beta folds with a deep active site cavity. Lysozyme uses this active site to hydrolyze the β-1, 4-glycosidic bond between N-acetylmuramic acid (MurNAc/NAM) and N-acetylglucosamine (GlcNAc/NAG) residues, the components that make up a bacterial cell wall.[21, 26]

Since the introduction of antibiotics, the number of harmful pathogens with resistance to these medicines has increased, leading to the creation of so called superbugs. These superbugs have brought attention to scientists to seek out new antimicrobial medicines to destroy the pathogens in a new manner. In consideration are large biologically influenced host defense systems, such as amphiphilic peptides and bacteriostatic proteins.[26, 27] Lysozyme is considered one of these bacteriostatic proteins with its ability to lyse cell walls.

1.7 N-Succinimidyl S-Acetylthioacetate

PEG is used in order for transferrin to conjugate to lysozyme, but in order for the conjugation to work, there must be another linker attached to the protein of interest to link to PEG. Lysozyme can use N-Succinimidyl S-Acetylthioacetate (SATA) to complete the conjugation. SATA is a reagent used to introduce protected sulfhydryls into a protein when thiols are unavailable or absent. It works by covalently forming an amide bond from the reaction of the N-hydroxysuccinimide ester with primary amines of the protein of interest. Deacetylation, also known as deprotection, then occurs using hydroxylamine to generate a sulfhydryl to be used in the cross-linking.[28, 29] This allows the completion of SATA to link with PEG for the conjugation of transferrin to lysozyme.



SATA

Figure 3: Illustration of the structure of SATA, the reagent used to modify Lz.

1.8 Solving the problems

As mentioned earlier, designing a novel drug is quite exciting, but developing this therapeutic to reach final stages can be difficult. The issues of premature degradation by proteolytic enzymes, immunogenicity, and delivery mechanism can be solved by the use of PEG to conjugate drugs to carrier proteins. Using conjugation to resolve those issues starts to create new difficulties, such as steric hindrance and the dug-to-carrier ratio, which in turn may cause new problems with receptor binding affinity, bioactivity, and efficacy. This study searches to develop and provide important feedback through

experimentation for the refinement and optimization of conjugate drug design. The model therapeutic used in this research is transferrin PEGylated to lysozyme using SATA as a linker.

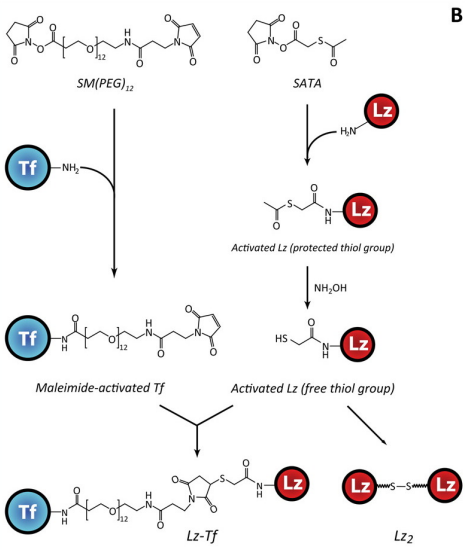


Figure 4: The schematic for creating the Lz-Tf conjugate. Tf is PEGylated and Lz is modified with SATA. Side products that could be created are dimers.[26]

Having the ability of transferrin to deliver lysozyme across the BBB could lead to the development of therapeutic strategies with the goal to cure infections in the central nervous system in a non-invasive fashion. The ideal concept is to have the internalization of Tf through the binding of its receptor on the BBB in order to release Lz on the basolateral side of the barrier. Lysozyme will then target the pathogens and lyse the cell walls. To help find some answers to these new challenges, this study utilized the capabilities of a few instruments and methods to help characterize the number of bound drugs and their locations within the conjugation.

CHAPTER 2
INSTRUMENTATION

2.1 Mass Spectrometry

Mass spectrometry is a method in which a mass-to-charge (m/z) ratio is determined in order to reveal the mass of a particular sample being analyzed. A mass spectrometer is an instrument that produces ions using an ionization source to then separate the charged species according to their mass-to-charge ratio by a mass analyzer. A mass spectrum consists of a plot of ion abundance versus mass-to-charge ratio and is presented in terms of Daltons (Da) per unit charge. Mass spectrometric analysis involves three main components: an ionization source, a mass analyzer, and a detector.

Mass spectrometry analyzes samples in a charge gaseous phase, so an ionization source is needed. In the past, samples had to first be volatilized and then ionized to be analyzed, but this limited the types of samples that could be analyzed to low molecular weight compounds that were thermally stable.[30] Advancements in ionization techniques has allowed the analysis of large, non-volatile and thermally labile compounds to be investigated. The source mainly utilized in this research is an electrospray ionization source.

The detector inside a mass spectrometer measures the ions and transforms this information into an electrical signal that can be relayed to a computer for user analysis. The detector, which is usually an electron multiplier, works by having dynodes held at successively higher voltages. Once energetic ions strike the surfaces of the cathode and dynodes, bursts of electrons are emitted that are attracted to the next dynode down a chain of dynodes, until the last one is reached where a large amount of electrons appear

for every ion that strikes the cathode.[31] The relay of electrons is then analyzed and converted to a signal for the computer to interpret for the user. There are different types of mass analyzers that measure ions in different ways. The most significant aspects of a mass analyzer are its accuracy and precision. Precision of the measurement is associated with the resolution of adjacent peaks, usually defined as $m/\Delta m$, where m is the integer mass of the peaks being resolved and Δm is the mass difference between the two peaks.[30] The analyzers used mostly in this study were quadrupole coupled with time-of-flight.

2.2 Quadrupole

The most common type of mass spectrometer used in quadrupole. The guts of a quadrupole instrument are the four parallel cylindrical rods acting as the electrodes. Opposite rods are linked electrically, with one rod acting as the positive side of a variable dc source and the other acting as the negative terminal. In addition, variable radio frequency ac voltages, which are 180^{o} out of phase, are applied to each pair of rods.[31] Ions are accelerated through the middle of the four rods and at the same time the ac and dc voltages on the rods are increase while staying at a constant ratio. At any given moment, all of the ions except those having a certain m/z value strike the rods and lose their charge, becoming neutral molecules, allowing ions having a limited m/z value range to reach the transducer.[31]

2.3 Time-of-flight

A time-of-flight (TOF) spectrometer works by separating ions based on their velocities through the drift tube. Ideally, all of the ions are formed at the same time and place in the ion source and are then accelerated through a fixed potential into the TOF

drift tube.[30] Once within the tube, ions with the same charge accelerate at the same rate, with low m/z ions achieving higher velocities than the higher m/z ions, all at an inverse relationship of the square root of the ions m/z.[30] After accelerating through the tube; they hit the detector at different time points based on their flight path allowing the m/z to be determined. More sophisticated TOF instruments include a reflectron, which allows the ions to travel farther by forcing them to make an almost 180° turn allowing a longer flight path to increase resolution because of greater separation between close m/z ions.

2.4 Electrospray Ionization

As little as thirty years ago, the idea of making proteins or polymers "fly" by electrospray ionization (ESI) seemed as improbable as a flying elephant, but modern techniques have allowed Dumbo to take flight and become a standard part of mass spectrometry and one of the most important techniques used for analyzing biomolecules.[32] In the 1980's, this soft desorption ionization method was conceived from the minds of two chemists, John Fenn and Koichi Tanaka, while searching for a method that could become the hyphen between LC-MS. Electrospray ionization is performed under atmospheric pressures when a solution of sample is injected through a stainless steel capillary needle at low flow rates of around a few micro liters per minute. The needle sustains a current ranging from several kilovolts with respect to a cylindrical electrode that surrounds the needle. This voltage creates a charged spray of fine droplets that passes through a desolvating capillary; where evaporation of the solvent and attachment of charge to the analyte molecules takes place.[31] The droplets then subsequently shrink just before entering into the mass spectrometer.

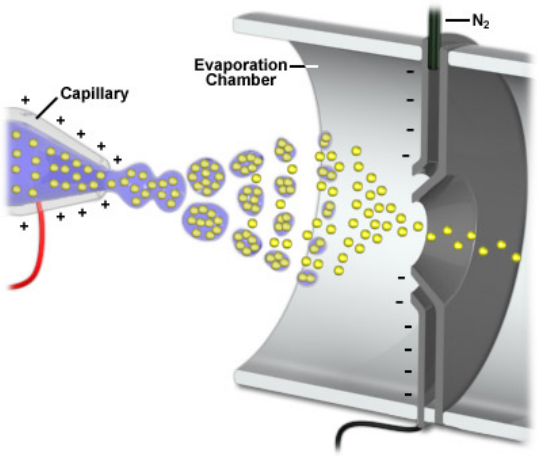


Figure 5: Electrospray ionization illustration. A voltage is transferred from the capillary to the sample. As the sample moves towards the mass spectrometer, desolvation occurs leading to multiply charged species to be analyzed.

As the sample proceeds to exit the needle, the liquid conforms into a conical meniscus known as a Taylor Cone. A Taylor Cone is essentially a buildup of charge at the tip of the needled caused by the created electric field, shaping the liquid into a cone figure developed by the surface tension and viscosity.[33] As the charge on the needle starts to increase, the cone begins to jet outwards until it starts to break at its weakest point, forming charged droplets

As the droplets move towards the mass analyzer, a progression of evaporation steps lead to smaller and smaller droplets (Figure 5). As solvent evaporation occurs from a charged droplet, the surface field increases until the Rayleigh limit is reached. At this point, the coulomb repulsion is of the same order as surface tension. The resulting instability disperses the droplet into a cluster of smaller droplets, sometimes known as a "coulomb explosion".[32, 34] The droplets continue to evaporate until ultimately each droplet is so small that it only contains a single solute molecule. The solute molecule preserves some of it's droplets charge to become a free ion as the rest of the solvent vaporizes.

Electrospray ionization mass spectrometry has provided scientists with the ability to analyze higher-molecular-weight molecules because of its ability to create multiple charged species. This method allows non-covalent biomacromolecular complexes to be ionized intact and has the ability to be combined with liquid separation techniques, such as ion exchange chromatography.[30]

2.5 Matrix Assisted Laser Desorption Ionization

Matrix assisted laser desorption ionization (MALDI) is another soft ionization method widely used to characterize large biomolecules. As implied in the name, MALDI uses a laser to bombard the sample to produce singly charge molecular ions. While ESI produces ions continuously, MALDI ions are produced by pulsed-laser irradiation of the sample, which is mixed and co-crystallized with a highly absorbing matrix on a steel plate. The more commonly used matrices are sinapinic acid, for protein analysis, and alpha-cyano-4-hydroxycinnamic acid (CHCA), for peptide analysis. The solvent used usually contains trace amounts of trifluoroacetic acid (TFA) to induce positive ionization. The matrix takes the energy from the laser pulse and transforms it into excitation energy for the sample, which creates a small plume of singly-protonated analyte and matrix ions from the steel plate. While under high vacuum, the ions usually go through a time-of-flight analyzer and are separated based on their mass-to-charge ratios. As with any method, there are favorable and flawed attributes. MALDI is terrific for conserving sample since it uses a pulse ionization instead of continuous flow, thereby little sample is wasted and high levels of sensitivity is achieved.[30] MALDI is also excellent in respect to time, allowing high throughput to analyze multiple samples in a short span of time. Some problems that come with the method has to do mostly with viewing ions in the spectrum.

The matrix causes a large amount of chemical noise at the low m/s range, hindering the ability to view low molecular weight samples. The method is quick and easy to interpret with its ability of producing singly protonated species, but this can cause a problem with performing peptide analysis or proteins of similar size because it can cause overlapping and difficulty identifying masses.

2.6 Analyzing a mass spectrum

Before the age of computers, scientists would physically calculate their results with pen and paper. Today, the ability of scientists to interpret their own results without the use of computer software is just as critical. A quick way to analyze a multiply charged mass spectrum is to use the given m/z values to calculate the charge state and approximate mass to help identify an unknown sample. To start the calculation, two adjacent peaks on the spectrum must be chosen, remembering that the difference between two peaks is one proton. The equation is as follows, where M = mass of sample, Z = charge state, and m/z is the peak value.

For the left peak, which has an extra proton:

$$\frac{M_L + (Z+1) \times 1}{(Z+1)} = m/z$$

Then solve for M in terms of Z:

$$M_L + Z + 1 = (m/z)\, Z + (m/z)$$

$$M_L = (m/z - 1)\, Z + (m/z - 1)$$

Next go to the right peak, which does not carry that extra proton, to later be compared with the M value for the first peak:

$$\frac{M_R + Z \times 1}{Z} = m/z$$

$$M_R + Z = (m/z)\, Z$$

$$M_R = (m/z - 1)\, Z$$

Then can compare both of our M values to solve for the charge state of the right peak:

$$M_L = M_R$$

$$(m/z - 1)\, Z + (m/z - 1) = (m/z - 1)$$

Solving for Z in the equation above once the the charge state is known, solving for the mass of the sample can be done by plugging in the Z value into the next equation:

$$M + Z = (m/z)\, Z$$

The mass will then be in units of Daltons (Da).

2.7 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) has been another key contributor in the advances of therapeutic research. A tandem mass spectrometer uses two analyzers separated by a collision cell where inert gas collides with the selected sample ions to bring about fragmentation and produce structural information for peptide mapping. The first analyzer isolates ions of a desired m/z, which are termed the parent or precursor ions. These ions are then induced to undergo chemical reaction that changes either their mass or charge, leading to dissociation. These resulting ions are known as product ions, and are then monitored by the second analyzer.[28] Three different types of bonds can fragment along the amino acid backbone: the NH-CH, CH-CO, and CO-NH bonds, as

seen in Figure 6A. This fragmentation creates a charged and neutral species, where the charged species is the one to be analyzed by the spectrometer. The most commonly cleaved bonds are the CO-NH bonds. The charge can attach to either of the two fragments depending on the proton affinity; therefore six possible fragment ions can exist. If the charge is retained on the N-terminal fragment, a, b, and c ions are created; and if the charge stays with the C-terminal fragment, x, y, and z ions are produced. The difference between two adjacent ions is equal to the specific amino acid residue weight. An example can be seen of a Lz fragment that was digested with trypsin and analyzed in Figure 6B. Observing that adding or decreasing the mass of each amino acid created b and y ions, the next peak will is that distance away.

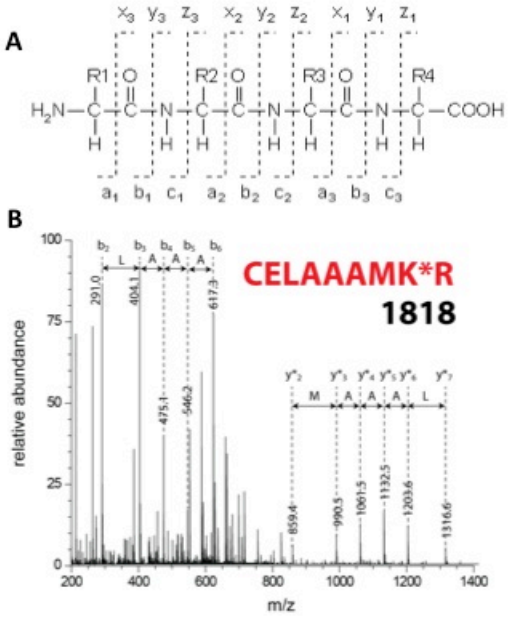


Figure 6: Illustrated are the fragmentations that can occur when using mass spectrometry (A) and an example of how MS^2 data is analyzed (B) using a peptide fragment of modified Lz. The asterisk between the lysine and arginine residue signifies a modified lysine residue.

2.8 Ion-Exchange Chromatography

Ion-exchange chromatography (IEC) is a practical technique used to separate species based on their charge. Ion exchange separation methods have been around since the first ion exchange resins were developed back in the 1930's, but the modern technique was not practiced until the middle of the 1970's, when high performance liquid chromatography showed anion or cation mixtures could be resolved on resin packed columns using the instrument.[31] Ion chromatography started with the development of separating rare earth cations with a cation-exchange resin for the research on the Manhattan project. The technique advanced to include the separation of other metals after World War II; which ultimately led to the detection and separation of amino acids and other ionic species found in complex mixtures.[31] Historically, ion-exchange chromatography was performed on small, porous beads that formed during emulsion copolymerization of styrene and divinylbenzene. These stable pores were then active towards ions by bonding functional groups to the cross-linking structure. While the packing performed the job of separating species, the process was slow; so two newer types of packing materials were developed. The first is polymeric bead packing, where the surfaces of the beads are coated with a synthetic ion-exchange resin and have a relative high capacity. The second type gives higher efficiencies and is prepared by coating porous microparticles of silica with a thin film of exchange resin.[31] There are two modes of IEC, cation exchange and anion exchange. In cation exchange, positively charged molecules are attracted and retained to the negatively charged solid support found in the column. The opposite is true with anion exchange, where the negatively charged species are retained to the positively charged stationary phase. The method

works by injecting the sample into the column and having a mobile phase of low conductivity wash through the initial species that are not retained by the stationary phase. The molecules are retained to the oppositely charged support until the mobile phase gains ionic strength and elutes the species out in order of their overall interactive strength with the functional groups on the stationary phase. In order for the mobile phase to elute the sample from the column, a type of gradient should be applied that varies the salt concentration or pH. By increasing the salt concentration, the molecules with weaker ionic interactions will elute out first, leaving the stronger ionic interactions to elute out later. By varying the pH, such as raising the pH to a more basic mobile phase, a cation exchange column will have the molecules become less protonated and less positively charged, allowing them to elute off of the negatively charged column. In principle, species can be separated based on their isoelectric point (pI), which is the pH at which the molecule holds no net charge, allowing it to have little to no interaction with the charged column.

CHAPTER 3

PAST STUDIES

The main source of background knowledge for conducting this research came from the past work of many scientists around the world. Research done on related lysozyme PEGylated therapeutic studies have been conducted by Lee et al. at the Korea Advanced Institute of Science and Technology and Maiser et al. at the Karlsruhe Institute of Technology.[7, 35] Research consisting of transferrin conjugation has been done by Peppas et al. at the University of Texas Austin.[36] This particular study came from the initial work of Son Nguyen in the Kaltashov laboratory at the University of Massachusetts Amherst.[26]

As aforementioned, identification of binding sites is critical in the characterization of conjugated therapeutic proteins. The study performed by Lee et al., presented a process for identifying PEGylated sites on their model protein lysozyme by using an avidin-biotin interaction and MALDI-TOF mass spectrometry. Using a PEG derivative, biotin-PEG-N-hydroxyl succinimide (NHS), conjugated to lysozyme; tryptic digestion was performed and fragments were separated using avidin affinity chromatography. Fractions of these peptides were analyzed by MALDI and RP-HPLC to determine the respective binding at the different lysine sites. In their work, Lysozyme and biotin-PEG-CO_2-NHS were dissolved in 0.1 M Tris-HCl, pH 8.0 and left at $4^{\circ}C$ for 10 hours. Acetic acid was added to terminate the reaction, adjusting the pH to 4.2, followed by the solution being dialyzed. The conjugated lysozyme was then analyzed by SEC using an isocratic mobile phase of 50 mM Tris-HCl, pH 8.0 with a flow rate of 0.8 mL/min with detection at 280 nm to collect mono-PEGylated lysozyme. The PEGylated lysozyme (6.3 mg) was then

dissolved in buffer and incubated with 6 M GuHCL and 10 mM DTT to reduce disulfide groups. Then to ensure there was no oxidation of the free thiol groups, 0.1 M iodoacetate solution was added and incubation occurred for 6 hours at room temperature. After dialysis, a solution of 50 mM Tris-HCl, 1 mM $CaCl_2$, and 2 M urea was added at a weight ratio of 100:1 lysozyme to trypsin. The group had trouble fully digesting the peptides because PEG was causing resistance to enzymatic digestion with its steric repulsion. To purify the digested conjugates, a column packed with monomeric avidin was used, where avidin binds to biotin with a K_d of 10^{-7}, allowing reversible binding of the biotinylated peptides under mild elution conditions (5 mM biotin). Separation of the biotin-PEGylated peptides was done RP-HPLC using a C-18 column with solvent A (0.1% TFA in H_2O) and solvent B (0.1% TFA in ACN). Analysis of the peptides was then conducted using MALDI-TOF mass spectrometry with a matrix of sinapinic acid. The study identifies four different PEGylation sites (K33, K96, K97, and K116) based on predicted m/z values using tryptic digest mapping patterns of native lysozyme. The sites at K96 and K97 produced the same tryptic digestion profiles, so the MALDI method at the time could not fully differentiate the mass peaks of the two sites, but based on past studies, they concluded that K97 was more reactive than K96.

Using the RP-HPLC, they were able to estimate the relative fractional amounts of the three major PEGylated fragments, revealing that the order of reactivity was K33 > K97 > K116. This study was a great stepping-stone for researchers looking to characterize conjugated therapeutic proteins, but Maiser et al. were not satisfied and took this research a step farther in trying and characterize all of the lysine binding sites.

The study done by Maiser et al. used MALDI-MS, as Lee had done before; but this time in the purification steps, ion exchange chromatography was utilized, allowing better separation of mono-PEGylated lysozyme. Lysozyme and mPEG-aldehyde were dissolved in 25 mM sodium phosphate buffer pH 7.2 containing 20 mM sodium cyanoborohydride as reducing agent at a protein to polymer ratio of 1:6. It should be noted that this group used a different PEGylation, where PEG-aldehyde readily reacts with N-terminal amine groups rather than the PEG-NHS mostly reacting with other amino groups, as in the case with Lee et al. This reaction was continuously shaken at room temperature for three and a half hours. The group used SEC to try and optimize the reaction time to produce higher yields of mono-PEG-Lz species. To collect mono-PEG-Lz samples, IEC was used with 25 mM sodium phosphate, pH 7.2 and 25 mM sodium phosphate with 500 mM sodium chloride, pH 7.2. Size exclusion chromatography was then run again to ensure only mono-PEG-Lz was collected in the fractions. Once satisfied with the collection of purely mono-species, IEC was used to try to separate the 7 possible isoforms on the conjugation. They separately ran the samples through a pH gradient from 10.5 to 11.5 and also a salt gradient from 0 mM to 500 mM NaCl. In their experiments, the salt gradient could only separate three isoforms. The pH gradient however worked splendidly for them, allowing separation of 5 out of the 6 isoforms, with the seventh being the N-terminal on lysine 1. Using the pH gradient allowed them to identify which lysine was being conjugated by means of calculating their individual isoelectric points (pI). The individual isoelectric points of the fragments were calculated by using a web-based tool called protein continuum electrostatics (PCE) that calculates the pK_a values of titratable groups in proteins by solving the Poisson-Boltzmann equation.[37] As seen in the

sixth figure of the article; based on the pI values, the modified residues elute in the following order: Lys 33, 1, 96, 97 & 116 together, and lastly 13. The linearity between the calculated pI value and the eluted pH value was within 99%.

In order to verify the binding site identifications, they investigated the mass of the PEG-peptide fragments. Using MALDI-TOF MS, four PEG-peptide fragments were detected from the fractions collected during IEC by determining the average mass of each peptide. The MS spectrum of each PEG-lys peptide shows peaks separated by 44 Da, which is caused from the PEG chains with a different number of monomer units attached. The mass over charge value of the peptide fragments were calculated by taking the m/z of the two fragments that would not be cleaved if PEG was bound; for example, with Lys_{33} which has peptides 22-33 (1325.63 m/z) and 34-45 (1428.65) flanking each side of the lysine residue, the addition of these peptides plus the m/z of the PEG (5,000 m/z) will give the correct m/z value.

It was not possible to prove through MS the third detected peak containing K96 because there were no detectable PEG mass peaks due to low concentrations. The group only observed K116 in the fourth peak, which is supposed to hold both K97 and K116 based on the same pI values. Validation of K13 could not be confirmed because they could not find the correct m/z value expected and even after some corrections for improper cleavage, the peptide still could not be validated. The group felt that their identification methods showed a good correlation and compared well with the results from a study done by Tilton et al, who used PEG-aldehyde and proposed a predominant N-terminal modification, which happened to have the largest abundance in their chromatogram as well.[38]

Transferrin studies done by Peppas et al. explored the conjugation of insulin to transferrin by forming disulfide bonds in order to improve insulin oral bioavailability through the use of receptor-mediated endocytosis. Using ESI-MS, the group studied the modification of insulin when conjugated to Tf, the stability of conjugated insulin to degradation, and site-specific modifications of insulin. A heterobifunctional cross linker called succinimidyl 3-(2-pyridyldithio)propionate (SPDP) that reacts with primary amines and sulfhydryl groups was used to conjugate insulin to transferrin. When SPDP reacts with the accessible amine groups, a mixed disulfide with free 2-pyridyldithio groups forms and reacts with free sulfhydryl groups from the other protein, resulting in heteroconjugation without unwanted cross-reaction products. While no unwanted cross-reaction products is great; the incorrect modification of insulin by DMMA, which blocks the N-terminal primary amines, allowing SPDP to react with the correct amines can diminish physiological activity. So the group used fluorescamine to quantify the number of free amino groups of insulin. The ideal insulin product should not react with any of the fluorophors, meaning that there are no free amino groups. While they showed that their reactions were producing what was wanted, this fluorescence assay couldn't tell the locations of the reaction sites, so mass spectrometry was used to solve this problem. Using ESI-MS, they showed that there were 1:1 and 2:1 DMMA:insulin species with no 3:1 ratio, which would signify nonspecific modification of insulin to the two N-terminal amines. Next they showed the 2:1:1 major product of DMMA/insulin/PDP using ESI-MS and decided to perform a tryptic digest on this species. There would only be one cleavage site for the insulin product, which produces a 43 amino acid peptide that contains both of the N-termini and an 8 amino acid peptide that contains the C-terminus and the PDP

attachment, if everything is bound to the correct sites. The smaller fragment's spectra show a mass shift of 197 Da, indicating the presence of the PDP attachment. While this shows that the PDP is clearly attached to the specific lysine residue since there is non-cleavage of that lysine, which would have produced a 7 amino acid peptide instead of the 8 amino acid peptide. The group further proved its binding location by performing an MS^3 experiment with the y_4-PDP and the y_2-PDP fragments, showing that the lysine was the sole site of PDP modification. To test the stability of the conjugate, degradation profiles were developed by using a trypsin solution and showed that the conjugate had increased stability compared to the insulin alone.

A bright PhD candidate named Son Nguyen, from the Kaltashov lab at the University of Massachusetts, accomplished the initial achievements for the research. His work explored the structure and interactions of lysozyme conjugated to transferrin as a model therapeutic that targets the central nervous system. The first portion of the study involved completing the chemistry of the conjugation. Following the classical scheme of conjugating Tf to a drug involves derivatizing the lysine side chains and amino terminus with sulfosuccinimidyl-4-[N-maleimidomethyl]cyclohexane-1-carboxylate (sulfo-SMCC) and then activating the payload, Lysozyme, by using Traut's reagent (2-iminothiolane hydrochloride). This scheme is suppose to ideally create a 1:1 stoichiometry so the functionalization of each protein is maintained with minimal hindrance. The actual scheme created is incredibly difficult to control to just two single reactive groups on each protein. With transferrin having 58 lysine residues and lysozyme having 6, the amount of different possibilities for binding between the two proteins is enormous, and trying to manipulate the proteins to only bind at two specific locations is a

daunting task. While being able to produce sufficient yields of the conjugate, this was only achieved if multiple activation groups were placed on both proteins. Creation of undesirable side reactions and homopolymerization due to these multiple attachments contributes to the heterogeneity of the conjugates. In order to try and minimize the unwanted products, Nguyen altered many variables and used mass spectrometry to optimize the yield of wanted conjugates.

The first variable looked at was the modification reagent. Traut's reagent started to create problems by forming dead end products of N-substituted 2-iminothilane (NSI) and with the reduction of intramolecular disulfide bonds. Using Traut's reagent with Tf lead to unexpected reduction of the native disulfide bonds leading to deleterious conformational changes in the higher order structure. This problem was seen in Lz as well, but was easily deterred by reducing the concentration and conducting the reaction on ice. As stated above, the reagent started to produce nonreactive products when the unstable thiol adduct broke down into a nonreactive 5-membered ring and reduced the number of free thiol groups; which can cause problems with downstream conjugation. Overcoming this obstacle was accomplished by substituting Trauts reagent with SATA, which was found to increase stability and reduce heterogeneity.

The next variables tested focused on trying to optimize reactions and reducing unwanted byproducts. Using ESI-MS, screening included changing the pH from 7.0 to 9.0, temperature from 0 to 37 °C, reaction times from 0.5 to 24 h, and ratios of protein to reagent from 2:1 to 1:4. The final optimized conditions for Lz were performed at pH 8.0, 0 °C, for 12 hours at an equimolar ratio. Transferrin was found to be optimized at pH 7.0, 25 °C for 90 minutes using 50 μM at a 1:2 ratio. While individually optimized, the Lz-Tf

conjugate was then screened and found to have the highest yield at pH 7 at 4 °C for half a day. Even after all of the optimization, a large number of other species can be seen in the mixture, so products must be separated from other products and reagents. SEC was thought of, but proved impractical at this stage because of the incremental mass increase of Lz-Tf over intact Tf.

A different way of separating the species is based on their pI values, with Tf having a pI value of 5.5-6.3 and Lz having its value at 11.0. Using weak cation exchange, the Tf, Lz and Tf-Lz species were all separated and after fraction collection and desalting, the ESI-MS confirmed each of the species. Amongst the details found by using ESI-MS, neither protein underwent partial or complete unfolding, but this does not totally mean that the proteins are unaffected when it comes to their abilities to properly bind to their partners. Native ESI-MS has allowed the research group to analyze Tf and its partner receptors and its relative binding affinity. From the spectrum below in figure 7, it is clear that the Lz-Tf conjugate is still capable of binding to the TfR, but at a lower affinity than intact Tf.

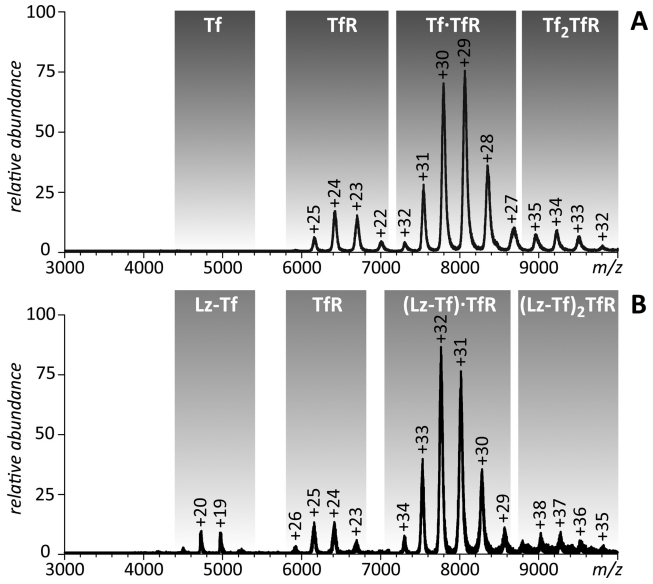


Figure 7: ESI mass spectra of Tf/TfR (A) and Lz-Tf/TfR (B) at neutral pH, 20 mM ionic strength.[26]

From the spectra, there is noticeably little to no ionic signal for the unbound Fe_2Tf, which is expected from the receptor-binding affinity of Fe_2Tf being in the sub-μM range. Then by observing the ionic signal that is detected from the unbound Lz-Tf in spectra B, it is concluded that the TfR binding affinity for the conjugate is in the low-μM range. Even though the affinity is lower, it still should be sufficient enough for the binding to occur and could become beneficial for the dissociation after passing through the BBB. From this experiment, it is certain that the conjugate will bind to the receptor, but the next experiment has to test whether the drug, or in this case the lysozyme, will still perform its proper function and exhibit its bacteriostatic properties.

In order to determine the enzymatic functionalization of Lz is still intact, a short trisaccharide NAG_3 is used as a surrogate substrate to show the substrate binding capabilities. Even with the non-specific binding of reagent near or far from the catalytic

site of Lz, the Lz-Tf conjugate still has the ability to bind NAG_3, which can be seen in the mass spectrum below in figure 8 by the presence of the protein-NAG_3 complexes.

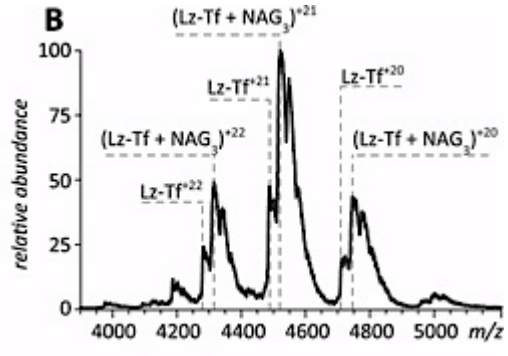


Figure 8: ESI mass spectra of NAG_3/Lz-Tf at near-native conditions.[26]

While the substrate binding capability experiment was promising, results from a photometric-based activity assay that measures the lysis of Gram-positive bacteria shows that there is a significant decrease in bacteriolytic activity of the conjugate. This assay can be seen in the article, where intact Lz and Tf are compared to conjugates as well as Lz dimers. It is hypothesized that the reduction of activity may be due to either change in electrostatic properties on the surface that may cause a conformational change near the catalytic site or that the large protein within the conjugation may be causing steric effects, making it difficult for the Lz to attack the cell wall. The results concluded that the Lz dimer is reduced 4-fold compared to intact Lz and after adjusting for having two catalytic sites, it was reported as 18.6%. This lysozyme dimer byproduct proves important to show that the while reagent modification of the protein contributes to the reduction in activity, it is not the sole contribution to decreased activity. This assay primarily showed that the inactivity is largely contributed to the steric restraint caused by the bulky anchor protein Tf. With this being the case, the thought of using a longer linker between the two proteins should decrease the hindrance and allow more movement for the Lz to attack cell walls

more efficiently. An amine-reactive $SM(PEG)_{12}$ was used to activate Tf and introduce free thiol groups to Lz still using SATA. With this new scheme, the conjugate was recognized by the TfR and had over a full order of magnitude higher Lz activity than the old scheme. In a perfect world, correlation between the location of linkers in the conjugation and the bioactivity of the conjugate can be monitored for the optimization during the design of future protein conjugates. Since there is likely a correlation between the location of the linker and the activity of the conjugate, identifying specific sites of conjugation by the use of mass spectrometry became the future focus of Nguyen's initial study and has become the current work detailed in this thesis.

The goal of this whole study is to provide important feedback through experiments to use in the refinement and optimization of conjugate drug development. To achieve this goal, SATA will be used to modify lysozyme in order to link with $SM(PEG)_{12}$ modified transferrin as a model conjugate and utilize different mass spectrometric approaches to characterize the binding. ESI and MALDI mass spectrometry was performed on these two proteins individually, as well as conjugated together, to identify the location of the linkers in this highly heterogeneous mixture. Initially, quantification was performed by tryptic and chymotryptic digestion to measure the intensity of each peptide fragment to determine the relative binding frequency. This proved to produce difficulties, so other methods were explored, which included site-specific modifications, ^{18}O labeled digestion, creation of calibration curves, and online/offline IEC. Various parameters were altered in order to achieve the best results possible with hopes that future experiments can follow this works progress.

CHAPTER 4
METHODS

4.1 Preparation of Lz-Tf conjugate

Lysozyme from chicken egg white (Sigma-Aldrich, St. Louis, MO) was modified on its primary amine groups by N-succinimidyl-S-acetylthioacetate (SATA; Pierce Biotechnology, Rockford, IL). The reaction was performed by incubating for 90 minutes at 37°C in 50 mM phosphate buffer pH 7.0 consisting of 100 μM Lz with varying ratios of SATA ranging from 1:0.5 to 1:16 protein to reagent. Ideally, a 1:1 ratio would be preferred. Activation of human transferrin (provided by Prof. Anne B. Mason, University of Vermont College of Medicine) was performed by reacting succinimidyl-([N-maleimidopropionamido]-dodecaethyleneglycol) ester ($SM(PEG)_{12}$; Pierce Biotechnology, Rockford, IL) for 90 minutes at 37°C in 50 mM phosphate buffer pH 6.0 consisting of 50 μM Tf and varying ratios of protein to reagent ranging from 1:1 to 1:4. Both reagents are prepared in DMSO and excess reagents were removed by using 5 kDa for Lz and 30 kDa for Tf centrifugal filters (Sartorius Stedim Biotech SA, Bohemia, NY). In order to deprotect SATA, for every 100 μL of modified protein solution, 10 μL of a solution of 0.5 M Hydroxylamine and 25 mM EDTA in 50 mM phosphate buffer should be added and allowed to incubate for 2 hours at room temperature. A 5 kDa filter can once again be used. The next step is to establish the noncleavable thio-ether linkage between the two activated proteins by incubating 50 μM of each modified protein in 50 mM phosphate buffer pH 7.0 overnight in the fridge at 4°C. The reaction between Lz and Tf can be seen in the figure 4 above.

4.2 Mass Spectrometry

All electrospray ionization mass spectrometry measurements were performed on a QStar-XL (ABI/SCIEX, Toronto, Canada) hybrid quadrupole/time-of-flight MS fitted with either a nanospray or turbo spray source. An LC Packings Ultimate (Dionex/Thermo Fisher Scientific, Sunnyvale, CA) was used to combine LC/MS/MS. All MALDI mass spectrometry measurements were obtained using a microFlex TOF MALDI MS (Bruker Biospin Corporation, Billerica, MA). The matrix used for this method was α-cyano-4-hydroxycinnamic acid (CHCA) saturated in 50:50 (v/v) water and acetonitrile with 0.1% trifluoroacetic acid (TFA). The samples were initially mixed with a 1:1 ratio of sample to matrix and 1 μL was pipetted onto a MALDI stainless steel plate and allowed to air dry. Calibration studies varied the amount of matrix and sample mixed before plating. The mass spectrometer was run in positive, reflectron mode that was optimized and calibrated prior to running samples.

4.3 Ion-Exchange Chromatography

Weak cation exchange chromatography was performed on an Agilent 1200 (Agilent Technologies, Palo Alto, CA) HPLC system using a 4.6 x 100 mm PolyCATA column (5 μM, 1000 Å, PolyLC inc., Columbia, MD). This column consists of silica-based material with a bonded coating of polyaspartic acid. To separate Lz isoforms, a salt gradient was introduced with a starting buffer of 50 mM phosphate buffer pH 6.8 and buffer B consisting of 50 mM phosphate buffer with 0.4 M NaCl pH 6.8 using a flow rate of 0.4 mL/min. To separate Tf isoforms, a pH gradient was utilized by increasing the pH of 50 mM phosphate buffer from 5.5 to 7.5 using a flow rate of 1 mL/min.

4.4 ^{18}O labeling

This isotope labeling method utilizes trypsin's ability to exchange both C-terminal oxygen atoms with labeled $H_2{}^{18}O$ oxygen atoms. When tryptic digestion takes place, it cleaves the carboxy side of lysine or arginine, and replaces these carboxy oxygen atoms with labeled ones, allowing the ability to run two samples at one time since the labeling will not change the chromatographic properties of the peptide and creates a 2- or 4-Da mass shift from the isotopic oxygen. After performing the digest of modified Lz and Tf, the sample is nitrogen or air dried until they are completely depleted of water. Once assured the sample is water free, water is added back into the sample by adding 45 µL of $H_2{}^{18}O$ (Sigma-Aldrich Chemical Co., St. Louis, MO) to the control sample and regular $H_2{}^{16}O$ to the modified sample. Then an additional 0.5 µL of trypsin and 5 µL of DMSO are added to allow incorporation to occur while incubating at 37°C for a day. When it comes time for analyzing the samples, figure 9 shows that normal tryptic cleavage will produce two separate peptide fragments (blue) to act as the control peptides. Once the protein is modified, trypsin can no longer cleave the modified lysine residue, meaning the spectra cannot show the modified peptide (red*). By comparing the cleavable peptides to the control, in this figures case, 70% to the 100%, it is clear that 30% of the abundance is lost, signifying the lost percentage went to the modified species, giving the ability to predict the amount modified per peptide. Looking at the mass spectra in the figure, a mass shift between ^{16}O and ^{18}O samples can be observed. The "I" values seen are the relative intensities, with I_0 being the modified species, I_2 being one incorporation of ^{18}O and I_4 showing a 4 Da mass shift representing our unmodified species. The "M" values are the expected relative intensities calculated by the analyst software used. In order to

determine the percent modified, a formula is used to find the ratio of ^{18}O to ^{16}O. Once the value of this formula ratio is found, the control ratio serves as the r_o value and the modified ratio serves as the r_m value in the next percent modified equation, which uses the control ratio over the modified ratio to give a percentage modified.

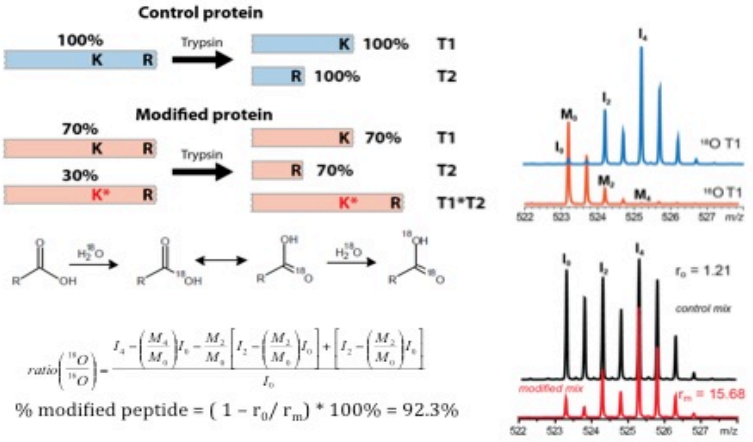


Figure 9: This figure illustrates the ^{18}O method. By using the control peptides, the percent modified can be calculated by introducing oxygen isotopes to modified species. By using the ratio of $^{18}O/^{16}O$, the r_o and r_m values can be determined to finish the percent-modified equation. To the right is an example of the 4 Da mass shift between ^{18}O and ^{16}O labeled species.

4.5 Site-Specific Modification

This site-selective method comes from an article by Weil et al. demonstrating their ability to control site modification.[39] With this method, they claim to have created a method that allows site-specific modification of the N-terminal of the protein RNase A by using kinetic control conditions. The main idea behind this method is to add small amounts of reagent slowly over a long increment of time. A depiction of their experiment can be seen in the first figure of the article. They start with their control spectrum with RNase A alone showing a single peak with the correct m/z value of 13682. The next step involves adding the entire volume of reagent at one time with a 1:10 ratio of protein to

reagent. The spectrum shows a poor resolution of up to seven modified species. Next by decreasing the ratio of protein to reagent, the spectrum becomes cleaner and shows fewer and fewer modified species. The true experiment starts when they use a 0.5:1 reagent to protein ratio and divide the reagent into 100 equivalent portions added stepwise over a course of 100 minutes. The reactions are further continued for 1 hour after the final injection of reagent to ensure modification has the time to take place. The reaction is finally quenched with the addition of ethanolamine solution and incubated overnight in order to deactivate any remaining unreacted reagents. By using affinity chromatography, the modified peak was extracted and once digestion, proved to represent the N-terminal, being the solely modified species.

A replication of this groups experiment was performed with Lz by using an automated syringe pump to add the SATA reagent to the protein solution at specific stepwise fashion. The first step is to create the Lz and SATA solution in separate vials, keeping in mind what ratio is desired and making slightly extra SATA solution to wash through the tubing connecting the syringe to the Lz solution. A 1 mL Lz solution was made using 50 mM phosphate buffer pH 7.0 and a 250 μL solution of SATA in DMSO was made. The method calls for adding 50 μL of reagent to the Lz solution, so the concentration of SATA had to be calculated differently each time a different ratio was desired. With the excess reagent solution, a thorough cleaning of the lines was performed beforehand to ensure that when it came time to injecting reagent, reagent was actually coming out of the line and not water or other unwanted solutions. Two samples with 1:4 and 1:2 ratios of Lz:SATA were prepared and the SATA solution was added all at once or at a 5 μL/min rate with a stir bar constantly spinning. Once the reagent was added, the

solution spun for an extra hour to ensure the modification would take place. The stir plate created heat during the longer addition steps, so placing the Lz solution on top of a weigh boat allowed the reduction of any heat coming from the magnetic stir plate. This experiment had a goal of creating a single species of mono-modified protein. Lastly, a 1:16 ratio was run at a 1 μL/min rate followed by immediate MS and IEC analysis to see how this increase in ratio and decrease in rate would affect the site-selection method.

4.6 Antimicrobial Activity Assay

The bioactivity of Lz and modified Lz was measured using resuspended dried cells of *Micrococcus lysodeikticus* (Worthington Biochemical Corp., Lakewood, NJ) as the substrate. The rate at which the cell walls were lysed was measured in 0.1 M sodium phosphate buffer pH 7.0 at 25 °C using the transmission at 450 nm. All measurements were performed in a 1 mL cuvette with a NanoDrop 2000c (Thermo Fisher Scientific, Rockford, IL) UV-vis spectrophotometer. Percent activity of modified Lz species is measured against control Lz species.

CHAPTER 5

RESULTS

There are few and far between transferrin based therapeutics despite its promising capabilities of transporting therapeutic payloads across the blood brain barrier. This rarity can essentially be blamed on the complexity and heterogeneity of conjugating this carrier to a drug, which causes stress when it comes time for analytical characterization. The focus of this study is on characterizing and optimizing the conjugation of transferrin and lysozyme by the use of linkers SATA and $SM(PEG)_{12}$. By characterizing the conjugate, the correlation between conjugation sites and efficacy of the drug can hopefully be determined. Transferrin alone has 58 lysine residues, which becomes the point of attachment in the conjugation; so trying to solve the optimal locations for the highest bioactivity or receptor binding is part of the goal for the overall large-scale project. Lysozyme has six lysine residues and the N-terminal amino group, making a total of seven locations for possible modification. Based on the model shown in figure 10, the locations of each lysine residue can be seen. As expected by the hydrophilic properties, all of the locations are located on the surface of the protein. The optimal location for modification is thought to be as far away from the catalytic site as possible so there is little interference with the bacteriostatic abilities of Lz. In order to try and solve the full conjugate problem, the challenge was broken into separate sections concentrated on solving the individual proteins first. This study goes through multiple methods in order to try to achieve every goal.

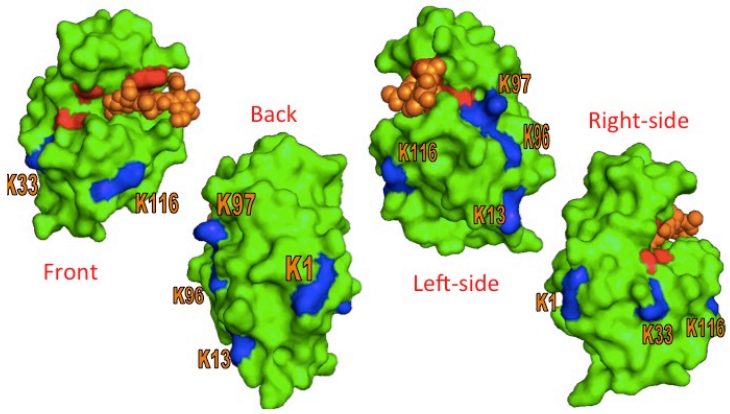


Figure 10: Lz surface showing locations of lysine residues (blue) and the activation site (red) that lyses bacterial cell walls (yellow).

5.1 Lysozyme Modification

The first series of experiments mimicked previous studies by performing a tryptic digest in order to calculate peak intensities of individual peptides relating to a percentage of modification. Lz is first modified by decorating the free thiol groups with various concentrations of SATA by taking 100 µM Lz in 50 mM phosphate buffer at pH 7.0 and incubating for 90 minutes at 37°C. Excess reagents were removed by using a 5 kDa centrifugal filter. Since the first portion of these experiments focus on the characterization of the individual proteins, Lz does not need deprotection of SATA for conjugation quite yet. With Lz modified, a tryptic digest should produce modified and unmodified peptides. The digestion is performed by incubating 14.3 µg of Lz in denaturing buffer, which consists of 6M guanidine hydrochloride in 100 mM ammonium bicarbonate pH 8, for 1 hour at 37°C. Once the reaction is cooled to ambient temperature, alkylation occurs after reacting the Lz solution with 500 mM Iodoacetic acid for 30 minutes in an ambient dark location. A buffer exchange of the protein from the matrix to 50 mM phosphate buffer is performed through the use of centrifugation. Finally 0.5 µg of

trypsin is added and the solution is incubated overnight at 37°C. To ensure that SATA is properly attaching to Lz, a large ratio of reagent to protein was used. The specta in figure 11 show that Lz is indeed being modified with SATA by having a mass addition of 116 units. As the concentration of SATA is increased from 400 μM (1:4) to 3200 μM (1:32), the spectra show a correlated increase in Lz modification; to the point where all seven locations appear modified and the predominant species shifts from being singly modified to having 5 modified sites per Lz protein. It also appears that as the SATA ratio increases, singly or doubly modified Lz species disappear; which are the desired products needed in order to determine the specific sites SATA prefers to bind to. It should be noted that the smaller peaks to the right of each SATA peak are salt adducts from the buffer used.

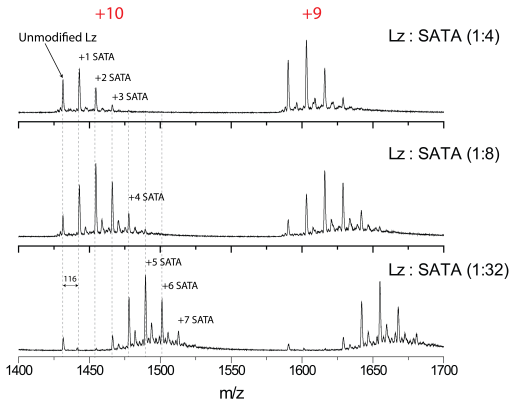


Figure 11: The MS spectra of different Lz:SATA ratios, showing the amount of modification that takes place as reagent is increased.

With the assurance that SATA will bind with Lz, the task of determining at which location does binding preference occur becomes the next focus. Through tryptic digestion, Lz will be chewed into smaller peptide fragments with cleavage points at lysine and arginine residues. When Lz is modified with SATA, the lysine residues are incapable of being cleaved; leaving specific peptides that can be used to calculate the

amount of modification at each individual lysine residue contained within the sequence. Utilizing a peptide sequence map to calculate correct masses of peptides that are cleaved at arginine sites and contain a modified lysine, Lz will have modified peptides 1-5, 6-14, 22-45, 74-97, 97-112, and 115-125. Unmodified Lz can still be cleaved at both arginine and lysine residues. Each modified and unmodified peptide has a specific masses based on the amino acids and can be correlated to peaks on a mass spectra. After calculating the masses of those modified peptides; a mass addition of SATA (116 units) will give the correct modified mass. Using the intensities can give a percentage of modification by using the modified intensity divided by the total intensity of the modified and unmodified (no 116 mass addition peptide). Based on the mass spectra, the mass addition of the modified peak was not adding up to the expected 116 units, but instead was seen at +132 units. An additional oxygen that was not expected can account this for. In figure 12, the modification scheme of Lz with SATA is depicted. Starting with the top left corner, SATA attaches to the primary amine group of Lz. Below that is the deprotection of the sulfhydryl group in preparation of binding Lz-SATA with the modified Tf. The sample is prepared for tryptic digest by adding iodoacetic acid for alkylation. The right half of the figure shows that when iodoacetic acid is added to a cysteine, which acts in this case as modified Lz, alkylation occurs and a mass addition of +132 appears on the mass spectra.

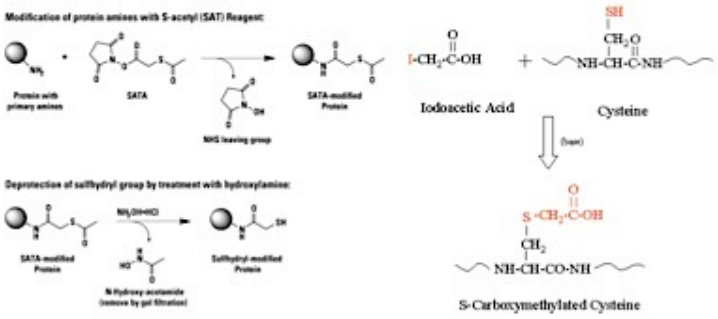


Figure 12: The scheme of how Lz is modified by SATA and activated by deprotection. When performing tryptic digestion, alkylation occurs with Iodoacetic acid. An example is given by using a cysteine residue.

5.2 XIC Estimation of Lysozyme

After complete tryptic digestion, the Lz-SATA samples are ready to run on the LC-MS and peptides are verified by MS2. Then by using the peak intensity values, the percentage of modification per lysine residue can be computed. The XIC of the expected modified peptides for 1:4 and 1:8 Lz:SATA can be seen in figure 13 The blue peak is the unmodified peptide (mass of peptide without SATA) and the mass addition of 132 units is the red modified peak. Typically the unmodified peaks should not be observed because they would be the cause of a miscleavage, since trypsin would cleave at the lysine residues; but since the peptide desired has the mass of the unmodified plus SATA, it was decided to use these unmodified peaks to determine the baseline amount of miscleavage that happens naturally. After going through the data, there was difficulty observing peaks over 2500 m/z and some of the modified peaks had low detection, causing problematic situations when trying to calculate the percentages since just a small change in intensity lead to a large change in percentage. In the 1:8 Lz:SATA K1 spectra, observed is another mass addition of 132 from the already modified, meaning that there was a doubly

modified version of that peptide, which could be thought of as an extra modification somewhere else in the peptide. The first lysine residue does happen to have two modification sites as stated above; one at the N-terminal and the other on the lysine residue.

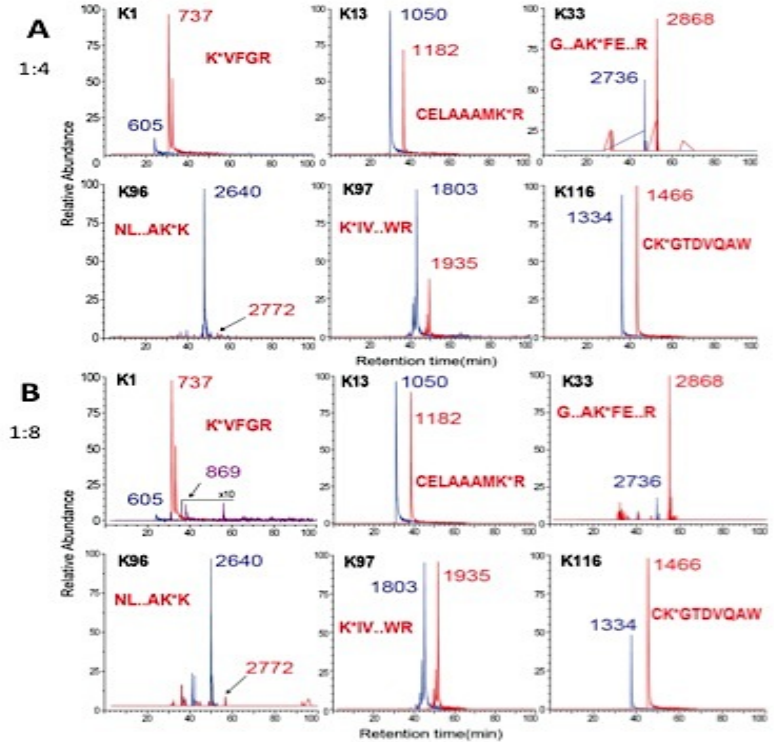


Figure 13: The XIC for 1:4 (A) and 1:8 (B) Lz:SATA. The blue peaks depict the unmodified species and the red peaks are modified species.

While the spectra do not look ideal, the percentages were calculated and as seen in figure 14, the K1 peptide followed by K33 had the highest percentage of modification. An interesting observation happens when the unmodified peak has a similar relative intensity to the modified peak. When the peaks have similar relative intensity, the peptide has a difficult time becoming modified since the unnatural miscleavage is just as common as the modified peptide. The spectra that have their red modified peaks with greater relative intensity over the unmodified are the peptides with the higher chance of

being modified. This hypothesis of relative peak intensity can be looked into later on when the IEC experiments begin. Since it was difficult to observe m/z values over 2500 and some smaller peptides with a higher charge, the thought of using chymotrypsin came to mind to see if this could help observe the missing peptides.

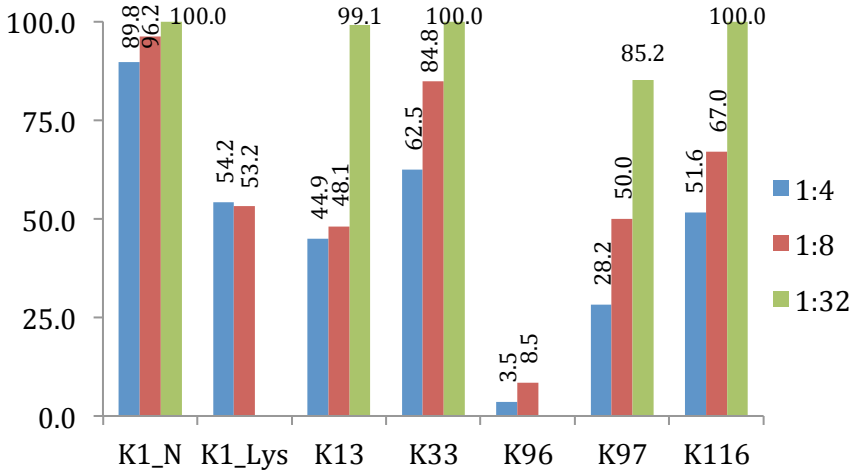


Figure 14: The calculated percent modified for each individual lysine residue at 1:4, 1:8, and 1:32 Lz:SATA ratios under tryptic digest.

Chymotrypsin cleaves peptides at residues containing aromatic ring such as tryptophan, tyrosine and phenylalanine. Below in figure 15 the modified lysine containing peptides that will be produced from the cleavage are seen.

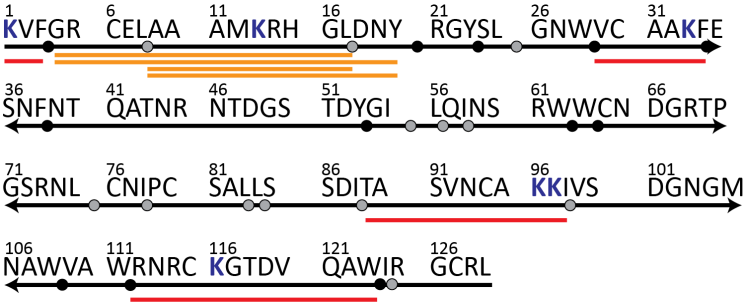


Figure 15: Lz peptide map for chymotryptic digest. The red underlined peptides are ones observed and the yellow are peptides not observed but needed.

The samples were run on the LC-MS and after verifying the peptides; their intensities were used to calculate the percent of modification for each modification site. The red underlines from figure 15 above indicate the modified peptides that were verified; while the yellow lines indicate peptides not observe in the spectra. The main troubles this method had were the inability to find a peptide containing lysine 13 and separate peptides for lysine 96 and 97. Based on what was found, figure 16 shows the XIC of Lz:SATA at a 1:4 ratio using chymotryptic digest and the percentages calculated from these spectra.

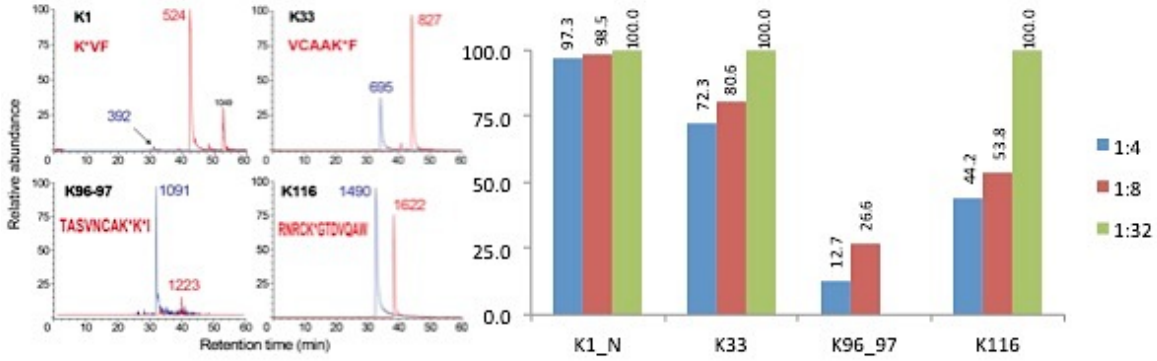


Figure 16: The XIC of 1:4 Lz:SATA (left) and the percent modified from the lysine residues observed with chymotryptic digest (right).

Desperate for quality results for the missing peptides, the thought of combining trypsin and chymotrypsin came into play. While at first this seemed like a decent idea and results looked promising, it was quickly realized that the combination was not going to work since trypsin and chymotrypsin will cleave one another and possibly create more havoc than good.

5.3 Transferrin Modification

Transferrin is the next part of the puzzle in this conjugation, but has proved to be the more troublesome piece. Transferrin binds to $SM(PEG)_{12}$ in similar fashion to Lz and SATA, with the only real difference being the length of the linker. As seen in figure 17 this time the mass difference between modified and unmodified peaks is 750 m/z.

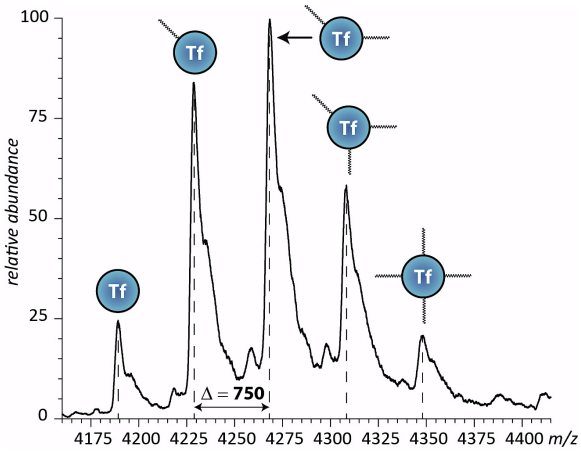


Figure 17: Mass spectrum of Tf and modified Tf, showing the mass difference from adding $SM(PEG)_{12}$.

5.4 XIC Estimation of Transferrin

Transferrin has proved too difficult of a protein for this simple tryptic digest method to figure out each percentage of modification. In fact, when looking at the peptides needed to solve this goal, there appears to be too many lysine residues in close proximity to one another. Based on the data deciphered, only a handful of peptides could be used; whose sequence can be found in the red underline in figure 18 of the Tf peptide map. The green lines show peptides that could not found based on their m/z values, but still contain a lysine residue needed to achieve the overall goal. Other problems arise, indicated with the blue and purple underlines, which either could not be seen simple

because of the large size or because of peptides with the same masses as another. Peptide 42-50 has the same mass as peptide 594-602 and peptide 103-113 has the same as 297-308. The yellow underlines are peptides observed, but were too difficult to verify because of low intensity values.

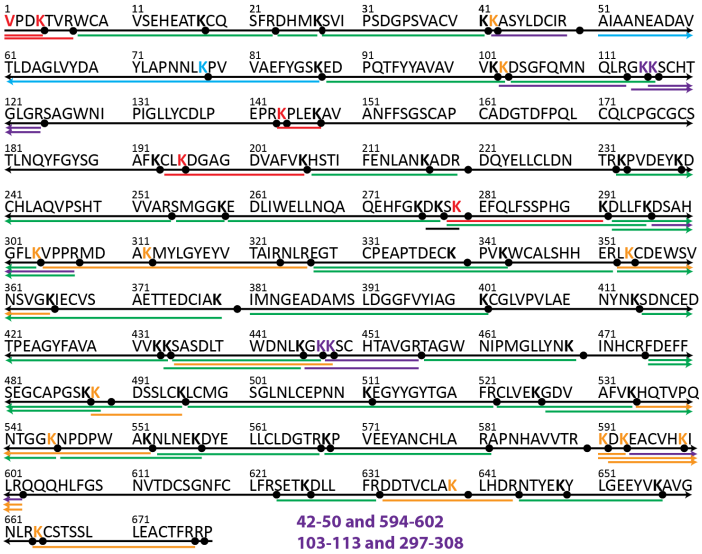


Figure 18: Peptide map of Tf. The red underlined peptides are peptides confirmed by MS^2, green underlined are peptides not observed, yellow underlined are peptides that were observed but not confirmed, and the purple underlined peptides are examples of problems faced by having same masses.

From the four peptides confirmed from the 1:2 Tf:$SM(PEG)_{12}$, it appeared that the first lysine containing peptide was highly modified based on the mass spectra. With the difficulties in both proteins, a new method was implemented, where the incorporation of an oxygen isotope is used.

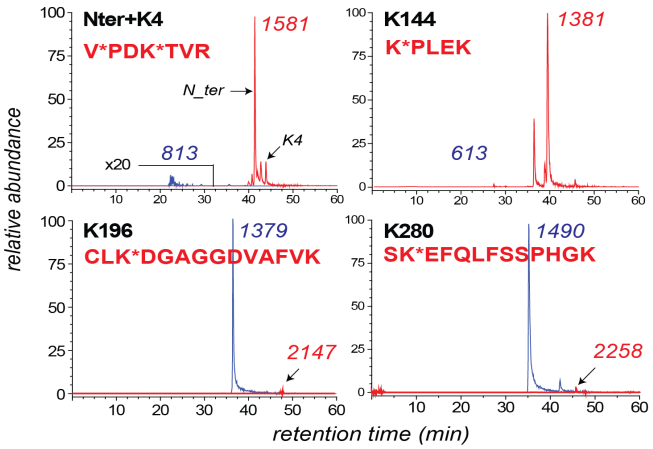


Figure 19: XIC of 1:2 modified Tf peptides. The blue peaks are unmodified and the red peaks are modified Tf.

5.5 ^{18}O Labeling

The new method involves the incorporation of ^{18}O into both carboxyl oxygen atoms of the C-terminal amino acid mediated by trypsin. This is performed in the same way as the original tryptic digest method, but after digestion, the samples go through a nitrogen/air gas drying. Once the samples are completely dry and depleted of water, they are reconstituted with either 45 µL of $H_2{}^{18}O$ to the control sample or regular H_2O to the modified sample. Control peptides are those that when properly cleaved, do not contain any lysine residues to be modified. The addition of 5 µL of DMSO and 0.5 µL of trypsin to both control and modified samples was done to compensate for trypsin losing 80% of its cleaving ability after a day. Afterwards, the solution is incubated at 37°C for a day. The initial method was changed during the reduction step to exclude the denaturing buffer of Gd.HCl and have ammonium bicarbonate instead. The incubation was also altered to 50°C because during the protein purification step to eliminate excess reagents, samples were losing protein with the centrifugal filters. By eliminating the Gd.HCl, there

is no need for the use of centrifugal filters in the digestion method, thereby possibly giving better results.

5.6 ^{18}O Labeling of Lysozyme

^{18}O labeling of Lz started by finding the control peptides (Figure 20). Running the Lz samples through the nanoLC-MS, the control peptides give the average r_o value to use in the calculations for determining the percent modified.

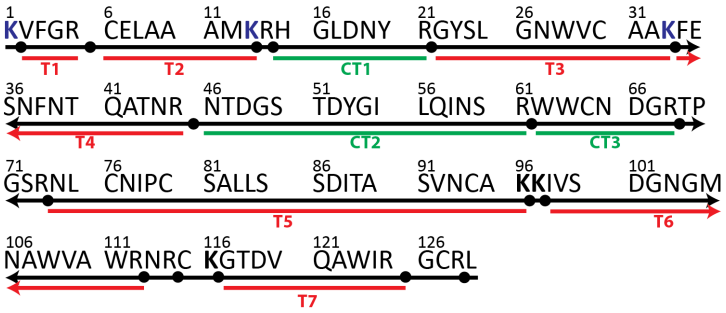


Figure 20: Peptide map of Lz. Red underlined peptides contain lysine residues and are needed for the r_m values. Green underlined peptides are control peptides not containing a lysine residue and serve as the r_o value.

With a 1:4 ratio of Lz:SATA sample (figure 21), the three control peptides were averaged to give a value of $r_o = 1.09 \pm 0.11$. This number comes from using the ratio equation shown in the methods section and then averaging the three calculated values. To produce the denominator of the equation, the r_m values for the seven modified peptides were calculated. The larger the r_m value, the more modification is possible at the specific lysine in that peptide. The same experiment was done with a 1:8 ratio of Lz:SATA. For the control peptides, the r_{o2} (peptide 46-61) was noticeably lower than the other two, so this was accounted to be an error and determined to be an outlier. With the outlier gone, the

calculated r_o value was 2.87 ± 0.1. Both Lz ratio samples agree in having the same peptide fragments as highly modified. Comparing the new isotopically labeled results with the old percent modification values seen from the XIC estimates in figure 22, there are some clear discrepancies when it comes to which lysine has more capability of being modified.

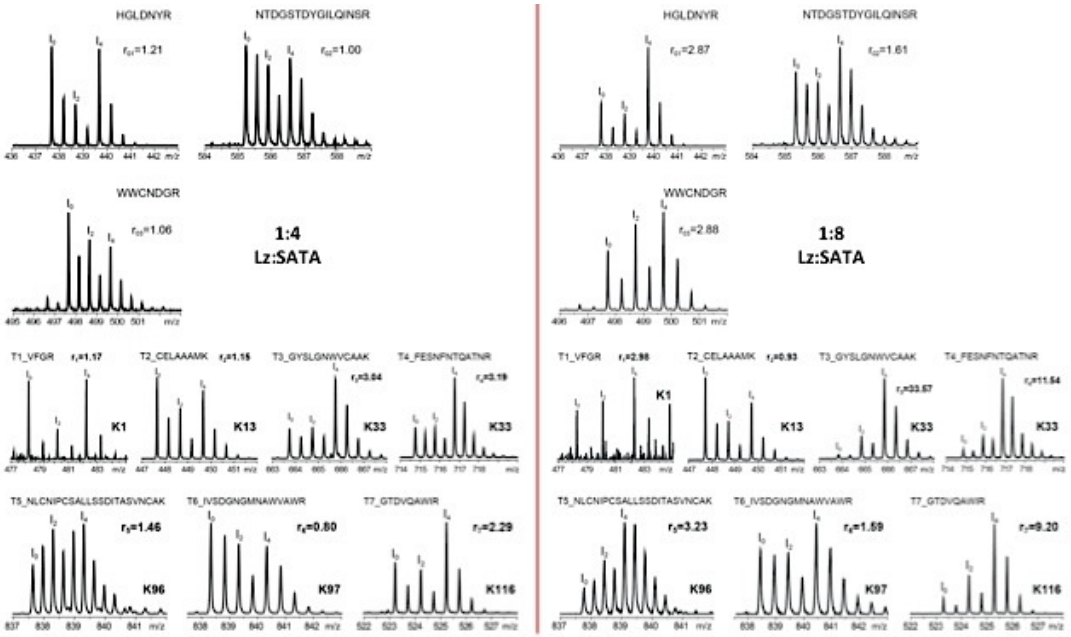


Figure 21: The mass spectra of 1:4 (left) and 1:8 (right) Lz:SATA using the ^{18}O labeling method. The top spectra represent the control peptides and the bottom spectra represent the modified species. Their calculated r-values are shown above each spectrum.

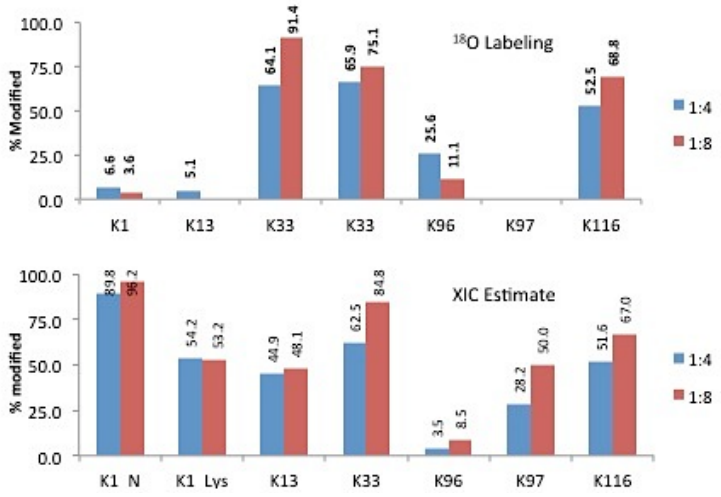


Figure 22: The percent modified for ^{18}O labeling method (top) and the XIC estimate method (bottom) are shown for 1:4 and 1:8 Lz:SATA ratios.

5.7 ^{18}O Labeling Calibration Curve

The ^{18}O values indicate that there is actually negative modification at lysine 97. Knowing that this cannot be fact, gave rise to the possibility that the method is peptide dependent and varies with the degree of ^{18}O incorporation depending on the amino acid residues. For this reason, the next step taken was to create individual calibration curves for each peptide; that way each peptide is its own control, leaving the variable of peptide dependence out of the equation. Each peptide will have two controls, the first being the simple ^{16}O and ^{18}O alone and the second being the 1:1 ratio, which gives the r_o value. The r_o and r_m values ideally should be equal to 1, making the percent-modified equation equal to 0% modified. The r_o value is equal to the 1:1 ratio, or the 0% modified. The r_m value is the percent modified value in question. The goal of creating a calibration curve is to find an equation for the line of best fit and use this equation to predict the percentage of modification.

5.8 ^{18}O Labeling Calibration Curve of Lysozyme

The same ^{18}O labeling procedure was followed, but the amount of ^{16}O species added to the sample was varied in order to create a standard calibration curve. After calculating the r_0 and r_m values, as seen in figure 23, the plot of these values creates an equation for the line of best fit. With the equation of a line (y = mx + b), the y value is the r_o value and the x value is the percent modified. In the graph shown in figure 24, each peptide can have its own linear fit line and equation.

		0 %	10 %	20 %	60 %	80 %	90 %	Mod Lz
Peptide Sequences		1:1	0.9:1	0.8:1	0.4:1	0.2:1	0.1:1	1:1
VFGR	2-5	0.58	0.80	0.97	1.59	4.14	5.15	0.93
CELAAAMK	6-13	1.74	1.93	2.38	4.24	11.99	17.74	2.16
(R)HGLDNYR	15-21	1.51	1.62	1.87	2.61	5.72	7.31	1.04
GYSLGNWVCAAK	22-33	2.08	1.25	1.95	2.54	5.97	6.31	5.34
FESNFNTQATNR	34-45	2.23	2.19	2.85	3.81	8.15	10.07	5.21
(R)NTDGSTDYGILQINSR	46-61	4.39	2.82	4.30	5.45	10.40	14.94	1.53
(R)WWCNDGR	62-68	1.20	1.21	1.47	2.15	4.39	5.49	1.40
NLCNIPCSALLSSDITASVNCAK	74-96	9.88	28.25	22.43	11.82	15.77	45.45	8.52
IVSDGNGMNAWVAWR	98-112	11.46	12.39	15.84	10.59	18.24	35.00	3.19
GTDVQAWIR	117-125	1.92	1.63	2.18	3.20	8.74	12.43	5.37

Figure 23: Standard calibration curve values for Lz. The r-values are observed as a percent modified and the last column is 1:1 modifed Lz. The peptides highlighted in yellow are the control peptides.

For examples, the top equation in the graph is for peptide 2-5 and the bottom equation in the graph is for peptide 117-125. Then by taking the normal 1:1 modified Lz, the r_o value can be calculated. For example, by taking the peptide 2-5's r_o value of 0.93

and plugging this value in for y; solving for x gives an answer of 28.2% modified. There will obviously be some error for each individual peptide when using the equation; for instance, the 60% modified r_0 value for peptide 2-5 ends up giving a result of 57.4%, leaving an error within 5%.

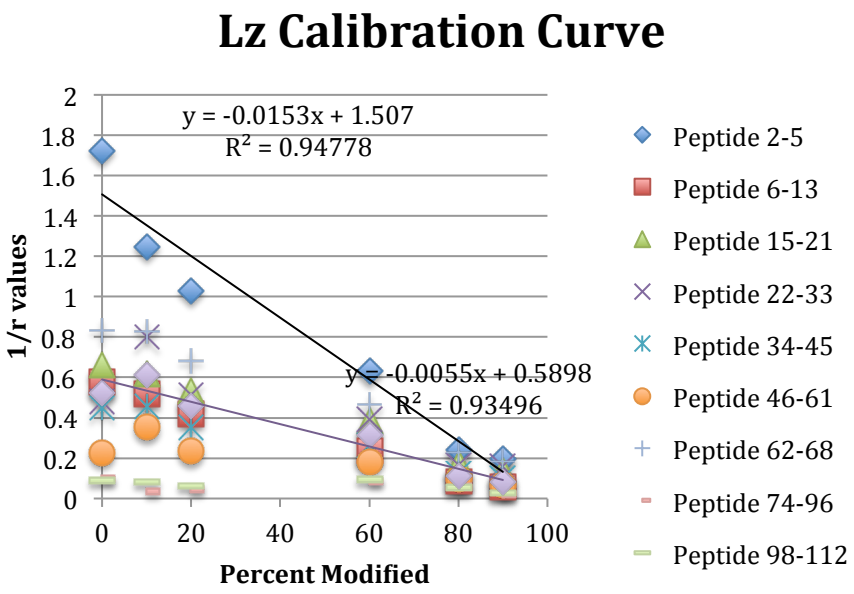


Figure 24: Graph of the Lz calibration curve. After finding a line of best fit, the equation is used to determine the percent modified for each peptide.

The standard equations created by the calibration curve from the ^{18}O method gives comparable results with values calculated by using the old percent modified equation of dividing the ^{18}O r value by the ^{16}O r value (figure 25). There is still a problem with lysine 96 and 97 when it comes to their ion counts being low and causing large errors in the calculations.

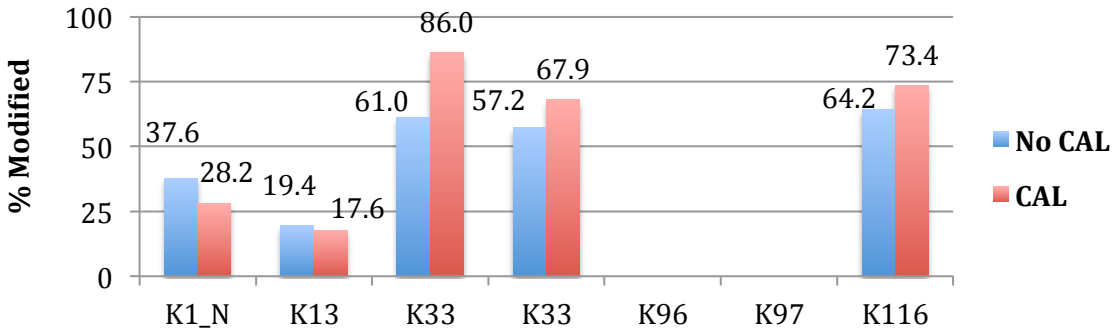


Figure 25: The percent modified for 1:8 Lz:SATA using the standard calibration curve. The blue trace shows the calculated percent modified by using the percent-modified equation, while the red trace utilizes the equation of a line to calculate the percentage.

5.9 ^{18}O Labeling of Transferrin

The same procedures were performed on Tf, but once again this large protein proved difficult and will need to be worked on more extensively in the future. It may be possible to use this method, but it will take a lot of man-hours to go through different MS^2 data to decipher which lysine is being modified when there are multiple residues in a single peptide. The beginnings of the Tf portion of the study found r-values for several peptides, but many more could not be calculated.

Plotting the few calculated r-value into a graph to find a line of best fit gives some hope for solving each peptide in the future. Using the equation of a line, the percent modified for the few peptides are compared with the percent modified values without using the calibration curve (figure 26). Once again, some of the values come out as a negative modification and this can be explained the same way as in Lz, where the specific peptides had low intensities, so the small variations in interpreting the data results create large changes in value.

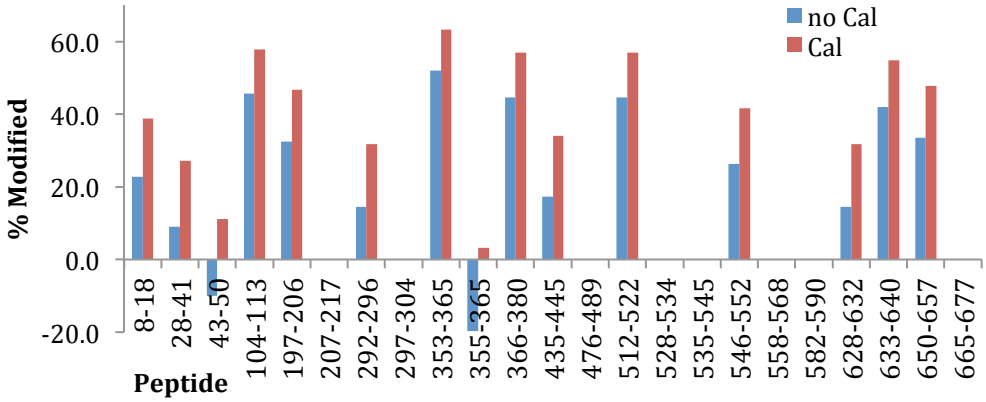


Figure 26: Percent modified for each observed Tf peptide. The blue trace shows the values calculated without the calibration curve and the red trace utilizes the line of best-fit equation to calculate the percent modified.

5.10 MALDI-MS

ESI-MS is not the only method capable of solving problems; MALDI-MS has the ability to produce similar results. Even though the project is aimed at specifically using ESI-MS, the use of MALDI-MS can help verify results. To prepare the ^{18}O labeled calibration curve samples, the amount of ^{16}O was varied once again while the ^{18}O amount was kept at a constant.

Ratio	Stock A (uL)	^{18}O sample (uL)	Matrix (uL)	Total Volume (uL)
1:1	5	1	0	6
0.9:1	4.5	1	0.5	6
0.8:1	4	1	1	6
0.4:1	2	1	3	6
0.2:1	1	1	4	6
0.1:1	0.5	1	4.5	6

Figure 27: Preparation for ^{16}O sample to ^{18}O sample for the MALDI calibration curve. The stock consists of a 1:4 ^{16}O sample to Matrix.

A stock solution consisting of 5 µL of ^{16}O sample and 20 µL of matrix (CHCA) was prepared. Samples had a total volume of 6 µL with varying amounts of matrix and stock

added, keeping the ^{18}O sample constant by adding only 1 μL. The 1:1 ratio was completed by adding 5 μL of stock and the constant 1 μL of ^{18}O, leaving no room for any additional matrix. The 0.9:1 ratio took 4.5 μL of stock and 1 μL of ^{18}O sample, leaving the remaining 0.5 μL consisting of matrix. The mass spectra for the control 1:1 mixtures can be seen in figure 28; where it is clear that the Lz is capable of being used in this method and Tf has some problems of overlapping peaks.

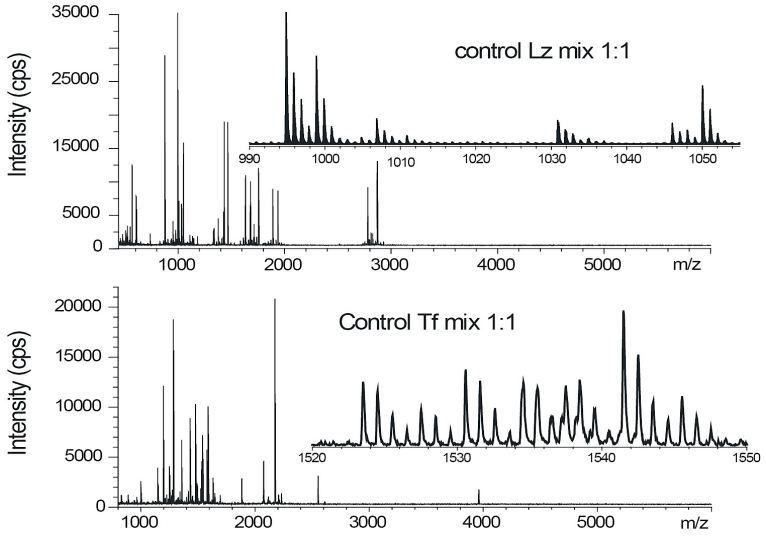


Figure 28: The MALDI-MS spectra for Lz and Tf control species using the ^{18}O labeling method. The Tf inset shows the troubles of overlapping peptides.

Figure 29 shows the Lz calibration curve produced by the MALDI instrument and the percent modified calculated using the equation created with the linear trend line. There was some difficulty observing the larger peaks, which turned out to be the peptides most desperately needed since they contain the 96 and 97 lysine residues that had been giving trouble before. As before with the oxygen labeling method, lysine 33 and lysine 116 seem to have the most probability of being modified.

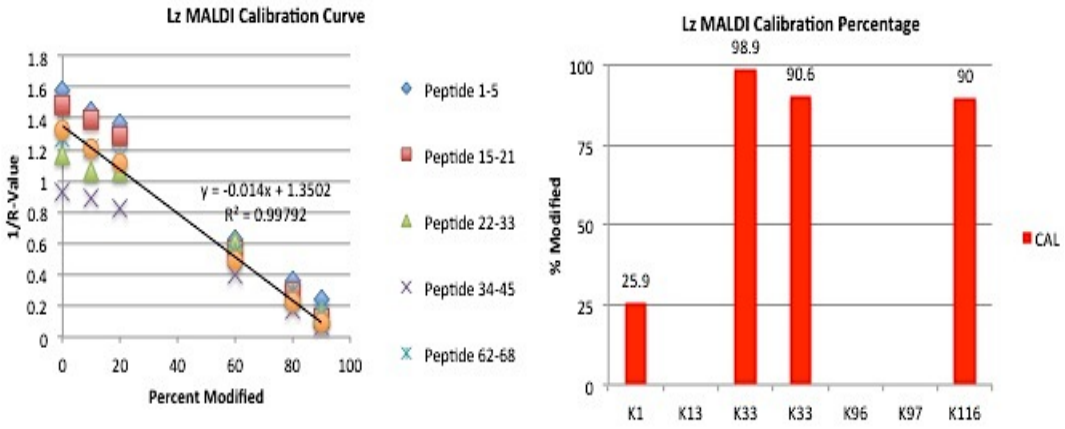


Figure 29: The calibration curve for Lz using MALDI-MS is depicted on the left, and the calculated percent modified using the line of best fit is shown to the right.

When it was time to run Tf samples with MALDI, only a handful of peptides could be fully interpreted because of overlapping peaks and low intensities. Comparing the values for the experimental percent modified against the theoretical percent modified shows nice linearity (figure 30. There is a possibility that if the instrument was highly optimized specifically for Tf, more interpretable peptides would appear; but strictly for time concerns, Tf samples were left for future work.

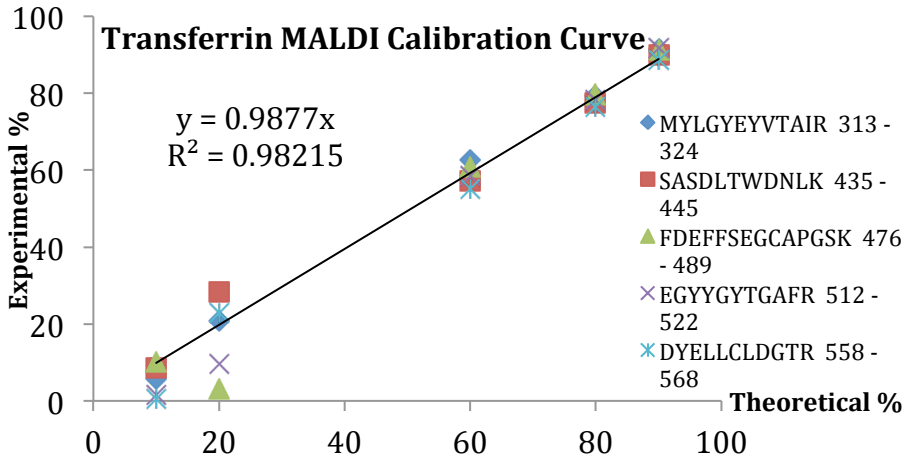


Figure 30: Calibration curve for the few observed peptides of Tf. We arranged the data to show the theoretical % modified to the experimental % modified and are able to observe a relatively linear line.

Although MALDI proved to be superior when it came to run time and precision, with each sample only taking about 1 minute for MALDI and about 100 minutes for ESI-MS, the method could not resolve as many peptides as ESI-MS and would take time consuming optimization for each type of sample. With the difficulties faced up to this point of not being able to interpret all of the peptides, the experiment took a different route after discovering the research done by Weil et al. The thought that by taking this site specific approach could remove the heterogeneity by only allowing a single site to be conjugated, revealing the most prevalent location for binding. This would eliminate the need to probe each lysine residue since each protein would only have one specific linker site.

5.11 Site-Specific Modification

In the paper, the authors claim to be able to specifically modify the N-terminal of the protein RNase A by using kinetic controlled conditions. In short, by injecting small portions of the reagent over time intervals, the concentration of reagent throughout the mixture is minimal, leading to sole binding at the optimal site. As mentioned in the methods section, the group showed how their approach worked by injecting at a ratio of 0.5:1 reagent: protein at 100 different time intervals to achieve their mono-biotinylated RNase A. Following their idea, an automated syringe pump was used to slowly inject SATA into a solution of Lz at certain time intervals. To start the experiment, a 1:4 Lz:SATA ratio (4eq) was created and the full amount of SATA was added to the Lz solution at one time (1-Portion). This is actually the normal procedure to create the modified Lz samples. From the MS, there is at least a tri-modified species observed (left

spectra figure 31). The next step was to halve the ratio and again add the entire amount of reagent at one time, leaving the MS to show at least a 2-SATA modified species. With a 1:4 ratio, the amount of SATA added to the protein solution was split into portions to try and replicate the results of the authors. In total, there is 50 μL of reagent to add to the Lz solution, so the experiment added 5 μL every minute for a total of ten times. There is a noticeable difference between the two mass spectra of the control 1:4 sample performed at the start and this step-wise sample, with less locations being modified. The same step-wise procedure was performed on a 1:2 ratio and once again the mass spectra shows less heterogeneity took place. With the MS data showing that there is still the presence of doubly modified species, it was decided to add less reagent per minute. To ensure there would still modification, a 1:16 ratio of Lz to SATA was prepared and had 1 μL of the reagent portion added every minute. The MS spectra in figure 31 shows that this slow addition of reagent did create a singly modified species, but unlike the original experiments, there was not sufficient yield of this site-specific species created even with an increased ratio of SATA. The experiments were performed on the bench top at room temperature; so there is a possibility that if the addition of reagent was done in a cold room that there could be a higher yield of mono-modified species.

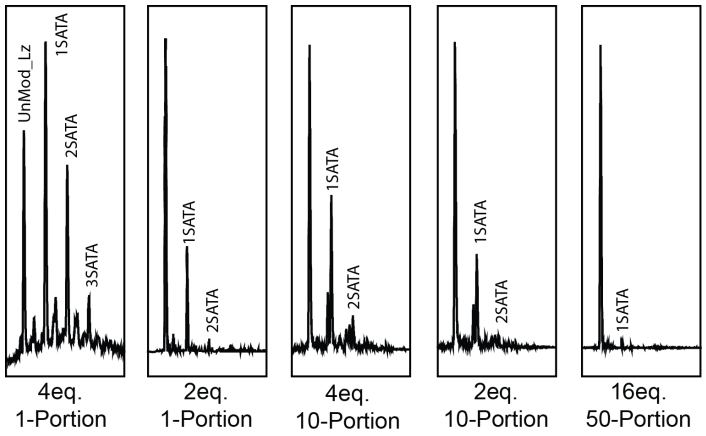


Figure 31: The MS spectra for the site-selective method. The farthest left box shows a 1:4 Lz:SATA ratio where all of the reagent was added at the same moment. Next is the same situation except a 1:2 ratio was used. Still moving right we see in the middle box the 1:4 ratio, but this time the reagent was added in 10 equivalent portions. The next box is the same situation except with a 1:2 ratio. The last box is the using a 1:16 ratio and added reagent in 50 equal portions.

5.12 Ion Exchange Chromatography of Lysozyme

After performing the site specific method and seeing that adding the reagent slowly in small volumes can create a singly modified species, the next step is to check and see if the species was actually a single lysine modified or if there were different lysine residues modified. Instead of digesting this site-specific species, weak ion exchange chromatography was used to show us if it is solely one site-specific species or if there are multiple site-specific species created based on the chromatogram peaks. By running a salt gradient, the chromatogram clearly shows that the site-specific species was in fact different lysine residues modified. While this was a sad outcome in regards to not being able to modify a single lysine residue and leaving the others unmodified; an opportunity to see which species is more abundant based on the peak areas could brighten the situation. From figure 32 the 1:2 Lz:SATA site-specific sample had fewer species than the 1:4 ratio, so to be able to collect and identify these species, a larger ratio of

reagent would be wise, so the smaller peak species can be observed and collected to run offline ESI-MS to identify which elution peak corresponds to which lysine residue.

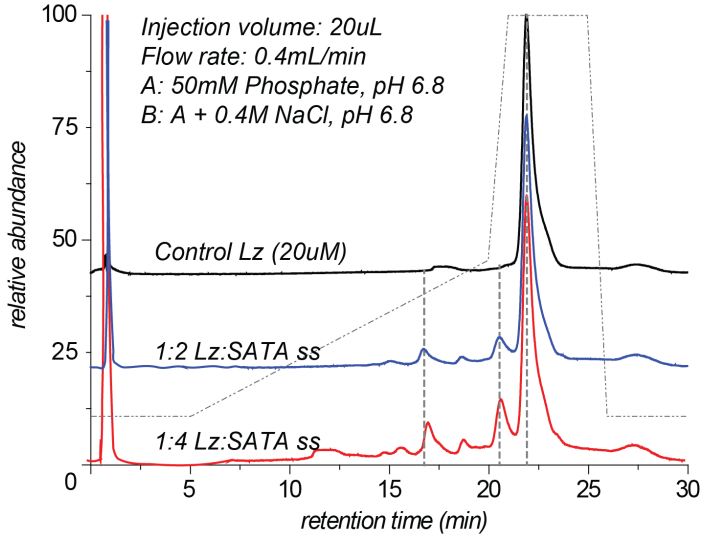


Figure 32: Weak ion exchange chromatogram of Lz samples. The black trace is control Lz, the blue trace is 1:2 Lz:SATA using the site-specific method, and the red trace is a 1:4 ratio using the site-specific method.

Since the site-specific experiments did not produce the same results as the previous lab, the next step was to collect fractions after isoform separation with IEC. When optimizing the conditions for the column by running multiple ratios of Lz to SATA samples; a noticable effect occurred when the pH of the buffer used in the modification was changed from 7.0 to 6.0. A lower yield of modified Lz was produced when we dropped the pH compared to the normal pH, meaning pH 7 is the optimal pH to try to collect fractions. Also shown in figure 33 are three different ratios of Lz:SATA and this shows how increasing the amount of reagent produces more multiply modified species. To the right of the chromatograms in the figure below are the MS spectra of each varied ratio sample, showing what is predominately produced from the increased ratio of reagent. The MS proves that the lower retention time peaks appear to be the more multi-

modified Lz species, since the 1:3 ratio samples mostly showed singly modified and the 1:12 ratio showed more triply modified species.

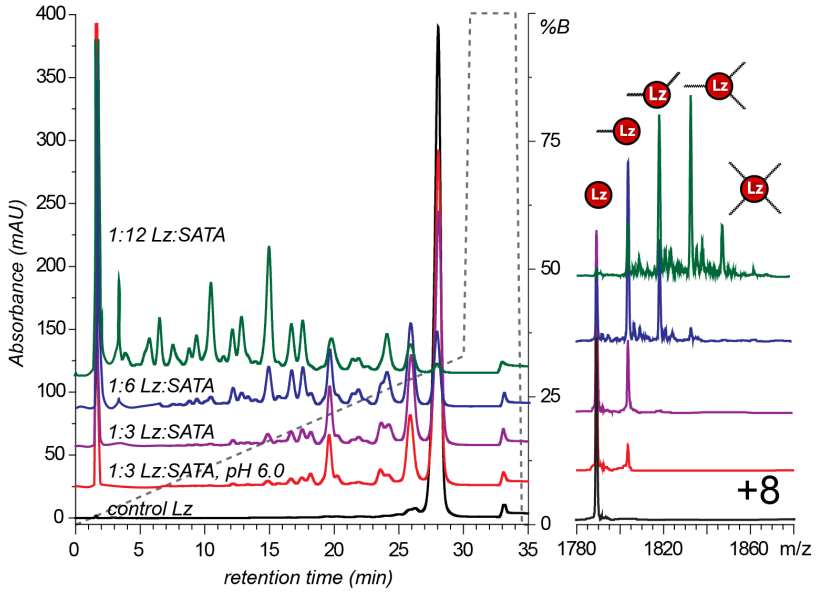


Figure 33: The left chromatogram depicts the IEC of different Lz species. The right shows the MS spectra of the same Lz samples depicting the Lz-SATA species present.

The MS spectra allowed us to make the observation that the lower retention time peaks were higher modified species, but there still was a need to identify what peaks actually correspond to which species. The long way of doing this is by collecting each peak and running offline MS to find out what is what, but an easier way is to run online MS and based on the m/z value, the spectra will tell exactly what species is predominantly in each elution peak. Going with the middle ground, a 1:8 Lz:SATA sample was run with the IEC-MS. The black trace in figure 34 shows the IEC chromatogram of the Lz-SATA sample. The large spike at the start is the solvent front and is mostly made from excess reagent and buffer. The red trace above the chromatogram is the MS spectra. The spectrum starts after the solvent front so damage

and salt build up is minimized on the MS instrument. Below both are the MS spectra based on what modified species is observed.

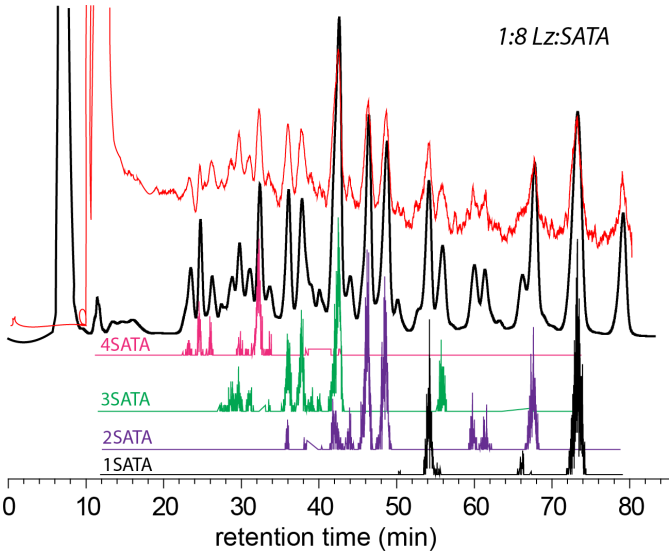


Figure 34: This figure shows the IEC chromatogram in black and the MS spectra in red. Below are the different species of modified Lz, showing what peaks correspond to which modified species.

As expected, the higher modified species elute out earlier than the mono-species. The data also shows that some peaks have overlapping species. With knowing where the mono-species are located on the IEC chromatogram, it is possible to collect these specific fractions in order to discover which lysine residue corresponds to which peak. This will provide the site-specific species that could not be created through the other methods. By identifying the specific location, the possibility of correlating each lysine residue with bioactivity can create a more potent and exact conjugate drug. Five fractions were collected, starting with F0 being the unmodified Lz sample and moving towards the lower retention time fractions 1-4. The most abundant modified peaks are fractions 1 and then 3, so hopefully by using the LC-MS each peaks corresponding lysine residues can be found, eluting to which site is the more likely binding sites on Lz. Fraction 1 contained

traces of both K96 and K97. Both may be found in the same peak because there is only a small difference in surface change when modified at these locations. There is a strong chance that the doubly modified species located between F1 and F2 (figure 35), around 24 minutes, is Lz doubly modified with both K96 and K97, since the distance from unmodified to F1 to this doubly modified peak appears to have equal distances from each other. Fraction 2 consisted of K13 modified Lz and fraction 3 contained lysine 33 modified Lz. Fraction 4 contained the K1 modified Lz species. The method was unable to observe even trace amounts of K116 in any of the collected fractions, leading to the belief that it is found somewhere else by co-eluting with multi-modified species. The results differ from Maiser et al. with regards to the location of the lysine residues as they elute off of the weak cation exchange column, but this could be explained by the different chemistry performed when modification occurs. Their lab was working with PEG linkers while the Lz in this experiment is modified with SATA. While both linkers attack similar residues; there is a difference in mass and surface charges. The results are somewhat similar to Lee et al. when it comes to K33 then K97 followed by K116 being the most predominate modification sites.

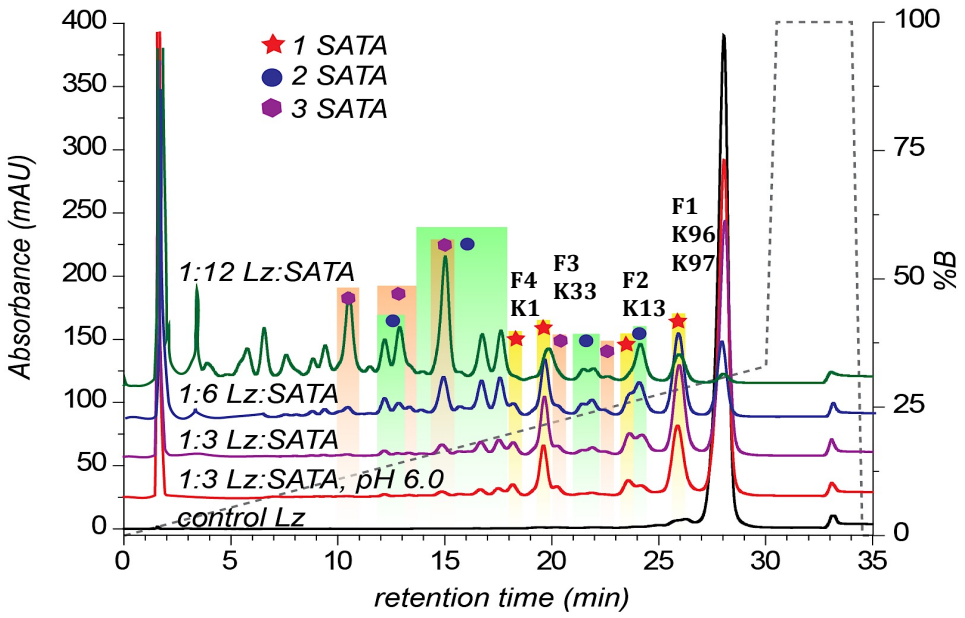


Figure 35: Shown is the chromatograms using IEC of the different Lz species after MS identified which species is present in which peak. Fraction 1 contained K96 and K97 modified species, F2 contained K13 species, F3 contained K33 species, and F4 contained K1 modified Lz species.

5.14 Bioactivity of Singly Modified Species

With the idea that modifying certain lysine residues may lead to different bioactivity of the protein, a photometric-based activity assay similar to what had been done before in the early experiments of this project was run. This experiment was performed more to see if the modification sites did produce any sort of difference; so the more probable sites were used since those are the sites most likely used in the future conjugation process. The three peaks were fractions: 0, 1, and 3. The initial results from the bioactivity assay used the unmodified collected peak as the control Lz sample and found that K33 had 49.9% activity and K96/K97 had 54.6% activity compared to the 100% unmodified sample. Dividing the rate by concentration and then comparing this value to the control will give the overall percent of bioactivity. In figure 36, the graph shows the concentration of each Lz species determined by UV and their rate of cell wall

lysis based on the slope of the line. The rate given for control Lz was 0.1928 a.u./min and the concentration of 5.6 μM was used. The rate for K33 was 0.1482 a.u./min and the concentration of 8.7 μM was used. The last fraction, with a concentration of 6.1 μM, had a rate of 0.1146 a.u./min.

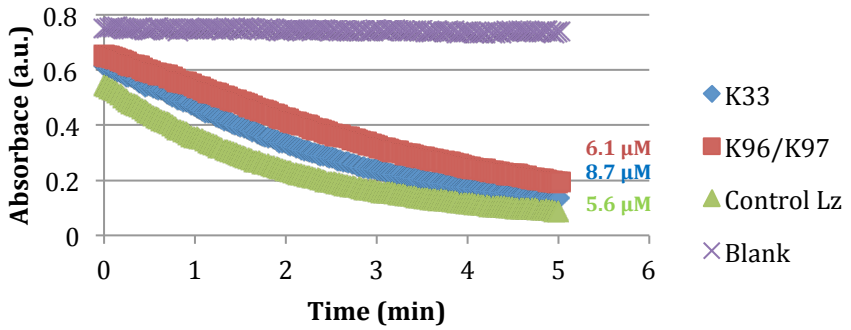


Figure 36: The bioactivity assay shows unmodified Lz in green, K33 modified Lz in blue, and K96/K97 modified Lz in red. Comparing rates to the control Lz, the percent activity can be calculated by taking rate over concentration.

If these species were conjugated to Tf, then these activity values would decrease significantly. Future experiments will hopefully be able to determine where K116 is located and separate the K96 and K97 species to allow a full set of modified lysines for bioactivity testing. While these experiments will be conducted in the future to better characterize the novel drug, the near future will work on separating the isoforms of Tf-SM(PEG)$_{12}$ samples.

5.15 Ion Exchange Chromatography of Transferrin

When running control Tf to observe the elution time of unmodified species, the chromatogram showed two peaks instead of one. The first thought that came up was the possibility that the transferrin was losing iron somehow and turning into Apo-Tf. A

sample of Apo-Tf was prepared by washing the stock Holo-Tf with 1mM EDTA in 3% acetic acid. The apo and holo control samples still showed small shoulders, meaning that the control still was not purely apo or holo (figure 37), but it was encouraging to see the second peak was indeed the other form of Tf.

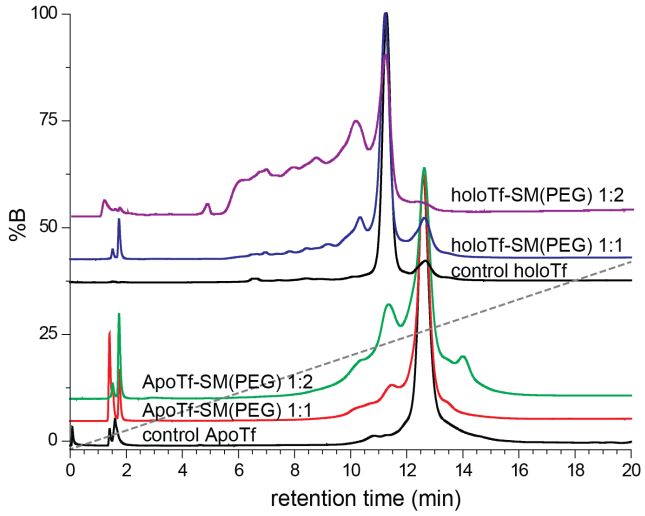


Figure 37: The chromatograms of apo-Tf and holo-Tf using IEC are shown. Ratios of 1:1 and 1:2 protein to reagent show modified isoforms to be determined by MS in the future.

From the chromatograms, it appears that holo-Tf has major isoforms, as seen with the multiple unresolved peaks in the 1:2 ratio sample. Apo-Tf looks as though there are not as many modified isoforms based on both chromatograms. If Tf is similar to Lz, then there can be the assumption that the first peak before unmodified species is a singly modified Tf. At this stage, it is unsure whether holo or apo Tf will produce a better result. It would be interesting to run a future experiment of using apo-Tf and once modification has occurred, replace iron back into the solution and use the mass spectrometer to see what comes of this.

CHAPTER 6

CONCLUSION & FUTURE

The amount of CNS infections appear to be sky rocketing, while the amount of small-molecule medicines capable of penetrating the BBB appear to be minuscule. With the fear of super bugs, antibiotic resistant pathogens, and the drastic side effects of inflammation immune responses ever growing, the urgency to develop novel biopharmaceuticals can no longer be ignored. Effective medicines starting to push their ways to the front lines of pharmaceutical research are chemically conjugated therapeutics, where a drug is bound to a transporter allowing specific delivery to target cells. The main reason why some of these awesome drugs have not yet reach any markets is because of their extreme complexity and heterogeneity that complicates the ability to characterize and create progress through the drug developmental processes. This work demonstrated that ESI-MS has the ability to provide characterization for a Lz-Tf conjugate.

The ability to determine the exact modification site for each of the conjugated proteins can hopefully later lead to the ability to correlate conjugation sites with bioactivity of the drug. Throughout the work, multiple methods were performed to try and predict the locations of modification for Lz and Tf. The first method performed was a tryptic digestion of modified Lz and Tf to run on nanoLC-MS-MS. Peptide fragments were searched for on analyst software and their intensities were used to determine a percent modified compared to the unmodified peptide peaks. This method had worked for others, but the current experiments had problems of low intensities and missing peptide fragments needed to truly calculate a trustworthy answer. The results showed that Lz was

mostly modified at the N-terminal and least likely modified at lysine 96. To try and overcome these problems, chymotrypsin was used as an alternative to digest samples, but once again had difficulty coming to a consensus answer. Transferrin showed a great deal more of trouble when digested because there were a great number of lysine residues that had the possibility of being modified and finding peptides that contained only one lysine residue was extremely difficult.

A new method was employed to help characterize the modified proteins with more accuracy. This ^{18}O labeling method involved the normal digestion steps, but included a little extra by incorporating the isotopic water into both carboxyl oxygen atoms to create a specific mass shift that can be detected by the mass spectrometer. This method allowed for the comparison of two samples in a single run, which helps a great deal with limiting variability between samples. This method relies on the fact that incorporating the isotopic species will not change the chromatographic properties of the peptides. By using peptides that did not contain any lysine residues, the amount of incorporation that takes place in modified and unmodified control peptides was calculated by using the ratio of ^{18}O over ^{16}O. Using the r_o values of the control peptides, it was discovered that lysine 33 of Lz appeared to be the most modified species, followed by K116. This differed from the original XIC estimates greatly, which had the N-terminal being the most heavily modified. Even this isotopic incorporated method had difficulty in observing clearly resolved spectra for all of the lysine residue peptides though. In noticing that the 1:8 Lz:SATA control peptides did not all have relatively similar r_o values, the belief that there is a possibility that the oxygen labeling is peptide dependent came about.

The way to get around peptide dependent labeling is to use each peptide as an individual control, eliminating the need to compare different peptides to each other. By creating a standard calibration curve, a line of best fit can be set to allow the percent modified to be calculated by simply inputting the r_m value into the equation of a line for each specific peptide. This method gave a little more comfort in providing percentages of modification for each lysine residue, even though there was still difficulty with finding the K96 and K97 peptide masses. These calibrated results showed that K33 and K116 are highly modified compared to the other lysine residues. Trying this method out with Tf showed once again that large proteins with dozens of lysine residues in close proximity creates problems when finding proper peptide masses.

Lysine residues 96 and 97 had been showing problems when it came time to finding their modified peptides. MALDI is another instrument utilized for observing peptide masses, but is only capable of producing +1 charged state species and this characteristic lead to problems with overlapping peptides. Lysozyme is a small enough protein that overlapping peptides did not create an issue, but once again the larger peptide fragments containing K96 and K97 were elusive. With the use of a calibration curve, the MALDI results related nicely with the ESI results by having K33 and K116 being the most probable sites for modification, but MALDI had each of these peptides modified by at least an extra 10%. MALDI could prove helpful when it comes to characterizing smaller proteins like Lz with the ability to run numerous samples extremely quickly (almost 100x quicker) when compared to the LC-MS method used in this study, but the MALDI method proved little help when it comes to larger proteins because of the many overlapping peptides observed in Tf samples.

The difficulty of finding each peptide in the mass spectrometer lead to a new method of trying to create a single modified site by introducing the reagents in small portions over long periods of time. This site-selective method proved to work for other groups and would have hopefully produced just a singly modified lysine species to finally prove which residue has the greatest chance of being modified. After slowly adding SATA every minute for 100 minutes to complete a ratio of 1:16 Lz:SATA, the results produced were disheartening. Although the MS showed that singly modified species was indeed being created, the yield of this species was dismal and running the samples on IEC proved that there was not just a single lysine modified, but in fact multiple isoforms created.

Observing that the site-selective method still created isoforms of Lz, it was decided to take advantage of that property to collect the separate isoforms. Utilizing the ability of weak cation exchange chromatography to separate isoforms of the modified species, online IEC was used to determine where the mono-modified species were located and those peaks were collected to try and help in characterization. Discovering where the singly modified species were located, the samples were run offline through the nanoLC-MS to determine which lysine residue corresponded to which chromatographic peak. Four singly modified peaks were identified using the MS spectra, with K33 and K96/K97 being the most abundant modified species for Lz. Peptides containing K96 and K97 could not be separated because of similar surface profiles when modified and there was no detection of the peptide containing K116, which the past results showed to be a dominant species, so this was discouraging to not have complete results.

Collecting specific lysine modified species allowed the ability to run bioactivity assays for each Lz species to see if modifying different lysine residues changed the bioactivity significantly. When compared to the unmodified species, it was found that K33 and K96/K97 had about 50% of the ability to lyse bacterial cell walls. The values were expected to be lower than the unmodified Lz species because the surface of Lz is being distorted and SATA may move in the way of the activation site cavity, hindering the ability to properly grab the cell wall and destroy the bacteria. This was a preliminary experiment using the highly modified species, where hopefully in the future these species can run with the Tf portion attached to get a full table of how this modified site correlates to the amount of bioactivity. Once Tf is attached, the bioactivity will surely decrease as Tf creates a much larger hindrance on Lz's activity.

IEC experiments with Tf were conducted with the hopes of seeing how many isoforms are created, allowing an estimate of how many peptides truly need to be searched for on the LC-MS. Initial results appear to show that apo- and holo-Tf are modified differently, and although the drug will be using a holo-form to bind to the receptor, studies on apo-form could prove insightful.

As seen through the overview of results, this project still has a lot to do in the future. The continued work on characterizing Tf and then the actual Lz-Tf conjugate remains the biggest project remaining. The possibility of utilizing new and longer linkers could prove to increase the bioactivity of Lz, as the hindrance from Tf would likely decrease as more space is created between the two proteins. The stability of the conjugate is yet another study to be performed in the future. Knowing exactly how long the conjugate will stay bound and if the activity or receptor recognition decreases over time

is essential. None of these decreases are expected to happen since the conjugate is forming disulfide bonds, so the conjugate should stay stable over a long period of time. Understanding how much of the conjugate actually reaches the brain and any other locations within the body is critical as well. Recent efforts to create large amounts of conjugate has allowed the study to move forward with rat injections to determine the drugs migration pattern and if enough can pass through the blood-brain barrier to support its role in effectively killing targeted bacteria. Initial studies seem to show that while the amount entering the brain is small, this concentration is enough to perform the task asked of it.

A bevy of mass spectrometric-based methods have been performed in this work to help characterize the conformational integrity of the model bacteriostatic protein lysozyme and it's transporter protein transferrin. The steps taken to achieve the goal of determining modification sites of both proteins can hopefully one day be used to create more advanced conjugate biopharmaceuticals. Till that day, the work discussed in this thesis will continue to be optimized to provide a model system for future medicines.

WORKS CITED

1. Harris, J.; Martin, N.; Modi, M. Pegylation. *J. Clinical Pharmacokinetics* **2001,** 40, 7, 539-551.
2. Pasut, G.; Veronese, F.M. Polymer–drug conjugation, recent achievements and general strategies. *Progress in Polymer Science* **2007,** 32, 8–9, 933-961.
3. Harris, J. Milton; Chess, Robert B. Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discov.* **2003,** 2, 214-221.
4. Moosmann, A; Blath, J; Lindner, R; Muller, E; Bottinger, H. Aldehyde PEGlyation kinetics: A standard protein versus a pharmaceutically relevant single chain variable fragment. *Bioconjugat Chem.* **2011,** 22, 1545-1558.
5. Davis, Frank F. The origin of pegnology. *Advanced Drug Delivery Reviews* **2002**, 54, 4, 457-458.
6. Davis, F.F.; Abuchowski, A.; Es, T.; Palczuk, N.C.; Chen, R.; Savoca, K.; Wieder, K. Enzyme-polyethylene glycol adducts: modified enzymes with unique properties. *Enzyme Engineering* **1978**, 169-173.
7. Maiser, B; Kröner, F; Dismer, F; Brenner-Weiß, G; Hubbuch, J. Isoform separation and binding site determination of mono-PEGylated lysozyme with pH gradient chromatography, *Journal of Chromatography* **2012,** 1268, 102-108.
8. Abzalimov, R. R.; Frimpong, A.; Kaltashov, I. A. Structural characterization of protein-polymer conjugates. I. Assessing heterogeneity of a small PEGylated protein and mapping conjugation sites using ion exchange chromatography and top-down tandem mass spectrometry. *Int. J. Mass Spectrom.* **2012,** 312, 144–154.
9. Gomme, Peter T.; McCann, Karl B.; Bertolini, Joseph. Transferrin: structure, function and potential therapeutic actions. *Drug Discovery Today* **2005**, 10, 4, 267-273.
10. Kaltashov, I. A.; Bobst, C. E.; Zhang, M.; Leverence, R.; Gumerov, D. R. Transferrin as a model system for method development to study structure, dynamics and interactions of metalloproteins using mass spectrometry. *Biochem. Biophys. Acta.* **2012,** 1820, 417-426.
11. Kaltashov, I. A.; Bobst, C. E.; Nguyen, S. N.; Wang, S. Emerging mass spectrometry-based approaches to probe protein-receptor interactions: Focus on overcoming physiological barriers. *Advanced Drug Delivery Reviews.* **2013**, 65, 8, 1020-1030.
12. Lao, B. J.; Tsai, W.-L. P.; Mashayekhi, F.; Pham, E. A.; Mason, A. B.; Kamei, D. T. Inhibition of transferrin iron release increases in vitro drug carrier efficacy. *J. Controlled Release* **2007**, 117, 403-412.
13. Daniels, T. R.; Delgado, T.; Rodriguez, J. A.; Helguera, G.; Penichet, M. L. The transferrin receptor part I: Biology and targeting with cytotoxic antibodies for the treatment of cancer. *Clin. Immunol.* **2006,** 121, 144-158.
14. Jones, A. R.; Shusta, E. V. Blood-brain barrier transport of therapeutics via receptor-mediation. *Pharm. Res.* **2007,** 24, 1759-1771.
15. Gabathuler, Reinhard. Approaches to transport therapeutic drugs across the blood-brain barrier to treat brain diseases. *Neurobiology of Disease* **2012**, 37, 1, 48-57.

16. Friden, P.M.; Walus, L.R.; Musso, G.F.; Taylor, M.A.; Malfroy, B.; Starzyk, R.M. Anti-transferrin receptor antibody and antibody-drug conjugates cross the blood-brain barrier. *Proc. Natl. Acad. Sci.* **1991**, 88, 4771-4775.
17. Pardridge, W.M. Blood-brain barrier delivery. *Drug Discov. Today* **2007**, 12, 54-61.
18. Regier, D.A.; Boyd, J.H.; Burke, J.D.; Rae, D.S.; Myers, J.K.; Kramer, M.; Robins, L.N.; George, L.K.; Karno, M.; Locke, B.Z. One-month prevalence of mental disorders in the United States. Based on five epidemiologic catchment area sites. *Arch. Gen. Psychiatry* **1988**, 45, 11, 977-986.
19. Fishman, J. B.; Rubin, J. B.; Handrahan, J. V.; Connor, J. R.; Fine, R. E. Receptor-mediated transcytosis of transferrin across the blood-brain barrier. *Journal of Neuroscience Research*. **1987,** 18, 2, 299-304.
20. Ibrahim, H. R.; Aoki, T.; Pellegrini, A. Strategies for new antimicrobial proteins and peptides: Lysozyme and aprotinin as model molecules. *Curr. Pharm. Des.* **2002,** 8, 671.
21. Imoto, Taiji. Lysozyme, *ELS*. **2009**, 1-6.
22. Mishra, O.P.; Batra, P.; Ali, Z.; Anupurba, S.; Das, B.K. Cerebrospinal fluid lysozyme level for the diagnosis of tuberculous meningitis in children. *J. Trop. Pediatr.* **2009**, 49, 13-16.
23. Brouwer, J,; Leeuwen-Herberts, T.; Otting-van de Ruit, M. Determination of lysozyme in serum, urine, cerebrospinal fluid and feces by enzyme immunoassay. *Clinica Chimica Acta* **1984**, 142, 1, 21-30.
24. Stoop, M. P.; Coulier, L.; Rosenling, T.; Shi, S.; Smolinska, A. M.; Buydens, L.; Ampt, K.; Stingl, C.; Dane, A.; Muilwijk, B.; Luitwieler, R. L.; Smitt, P.; Hintzen, R. Q.; Bischoff, R.; Wijmenga, S. S.; Hankemeier, T.; van Gool, A. J.; Luider, T. M. Quantitative proteomics and metabolomics analysis of normal human cerebrospinal fluid samples. *Mol. Cell. Proteomics* **2010**, 9, 2063-2075.
25. Rosenling, T.; Stoop, M. P.; Attali, A.; Aken, H. v.; Suidgeest, E.; Christin, C.; Stingl, C.; Suits, F.; Horvatovich, P.; Hintzen, R. Q.; Tuinstra, T.; Bischoff, R.; Luider, T. M. Profiling and identification of cerebrospinal fluid proteins in a rat EAE model of multiple sclerosis. *J. Proteome Res.* **2012**, 11, 2048-2060.
26. Nguyen, S. N.; Bobst, C. E.;Kaltashov, I. A. Mass spectrometry-guided optimization and characterization of a biologically active transferrin-lysozyme model drug conjugate. *Molecular Pharmaceutics*, **2013,** Online publication: 27-Mar-2013.
27. Hancock, R. E. W.; Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.* **2006**, 24, 1551-1557.
28. Pierce Protein Biology Products. SATA and Sulfhydryl Addition Kit. *Thermo Fisher Scientific*. http://www.piercenet.com/product/sata-sulfhydryl-addition-kit. Web.
29. Duncan, J.S.; Weston, P.D.; Wrigglesworth, R. A new reagent which may be used to introduce sulfhydryl groups into proteins, and its use in the preparation of conjugates for immunoassay. *Analytical Biochemistry* **1983**, 132, 1, 68-73.
30. Glish, G.; Vachet, R. The basics of mass spectrometry in the twenty-first century. *Nature* **2003**, 2, 140-150.

31. Skoog, D.A.; Holler, F.J.; Crouch, S.R. Principles of Instrumental Analysis, 6e; Thomson Brooks/Cole: Canada, **2007**, 923, 325.
32. Fenn, B. John. Electrospray Wings for Molecular Elephants (nobel lecture). Angewandte *Chemie Int. Ed.* **2003**, 42, 3871.
33. Kebarle, P. J. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J. Mass Spectrom*. **2000**, 35, 804.
34. Fenn, J.; Rosell, J.; Meng C. In electrospray ionization , how much pull does an ion need to escape its droplet prison. *J. Am. Soc. Mass Spectrom*. **1997**, 11, 1147.
35. Lee, Haeshin; Park, T.G. A novel method for identifying PEGylation sites of protein using biotinylated PEG derivatives. *J. Pharm. Sci.* **2002**, 92, 1, 97-103.
36. Kavimandan, N.J.; Losi, E.; Wilson, J.J.; Brodbelt, J.S.; Peppas, N.A. Synthesis and Characterization of Insulin-Transferrin Conjugates. *Bioconjugate Chem.* **2006**, 17, 1376-1384.
37. Bashford, Donald. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* **1992**, 224, 473.
38. Daly, S.M.; Przybycien, T.M.; Tilton, R.D. Adsorption of Poly(ethylene glycol)-Modified Lysozyme to Silica. *Langmuir* **2005**, 21, 1328-1337.
39. Chen, X.; Muthoosamy, K.; Pfisterer, A.; Neumann, B.; Weil, T. Site-Selective Lysine Modification of Native Proteins and Peptides via Kinetically Controlled Labeling. *Bioconjugate Chemistry* **2012**, 23, 500-508.