EXPRESSION AND CHARACTERIZATION OF THE MEMBRANE PROTEINS TCR α AND γ-SARCOGLYCAN WITH THE QUANTIFICATION OF *IN VITRO* GLYCOSYLATED γ-SARCOGLYCAN

By

MICHAEL SHAUN HARRIS

Bachelor of Science in Chemistry
Missouri State University
Springfield, Missouri
2017

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2022

EXPRESSION AND CHARACTERIZATION OF THE MEMBRANE PROTEINS TCR α WT AND γ-SARCOGLYCAN WITH THE QUANTIFICATION OF *IN VITRO* GLYCOSYLATED γ-SARCOGLYCAN

Dissertation Approved:

Dr. Gabriel A. Cook

Dissertation Adviser

Dr. Allen W. Apblett

Dr. Toby L. Nelson

Dr. Christopher J. Fennell

Dr. Junpeng Deng

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Dr. Gabriel Cook because without out him I may not have even went to graduate school at Oklahoma State University. He has always been welcoming and had an open door where I could always ask him questions to further my academic studies and research. He was always supportive, a great mentor, and advisor who made coming into lab everyday exciting with his enthusiastic attitude towards his research. I would also like to thank my committee members Dr. Apblett, Dr. Fennell, Dr. Nelson, and Dr. Deng for their guidance in the classroom and outside of classroom settings. Their guidance, comments, and ideas have made me a better scientist inside and outside of the laboratory setting. Another group of people I need to thank are those in the Molecular Biology Recombinant DNA and Protein Core Facility, specifically Dr. Hartson who assisted in mass spectrometry studies and has run several LC-MS/MS experiments and trained me on data analysis of these experiments for my research.

I would like to thank specifically Dr. Leshani Liyanage, a former lab member, for her help in the and training me in different techniques and instrumentation in the lab. Her help along with Dr. Cook's help in training has been immeasurable into my success and growth in the lab. I would also like to thank other lab members past and present including Austin Anderson, James Bryce, Tiwa Ogunleye, Ryan Mclntire, Daniel Reed, Michael Jamaleddine, Allye Gardner, Rylee Barber, Darci Kuck, Rachel Dolan, Megan Smith, Libby Caldwell, Bobby Sanford, Sophia Darrow, Alyssa Hickerson, Jacob Solomon, Emily Forster, Moriah Thompson, and Juliana Arndt for the experiences and knowledge shared together. I would also like to thank other graduate students in the chemistry department who have helped me along the way including Miguel Leal, Kwame Glinton, Jennifer Minnick, Ardalan Hayatifar, Erik Lantz, Alex Shumaker, Habeeb Al-Mashala, Colton Calvert, and Thad Stancil.

I want to thank my family and friends as well including my parents David and Peggy Harris and my two brothers Ryne and Cody and sister-in-law Mary for their unrelenting support through my time in graduate school. Thanks also to my friends Josh and Lauren Harkless and Sam and Jordyn Miller who have always been supportive and awesome to meet up with when possible and hangout with to help relieve stress from the arduous work of graduate school. All of these people have been some of the most instrumental people through these graduate school times, and I cannot thank them enough especially my mother who is the reason why I decided to further my knowledge in chemistry in the first place.

Lastly, I would like to thank my undergraduate research advisor Dr. Gautam Bhattacharyya and undergraduate advisor Dr. Gary Meints who pushed me to graduate school and prepared me for the struggles of graduate school. They nurtured my knowledge for chemistry and pushed me and helped mold me into the chemist I am today.

Name: MICHAEL SHAUN HARRIS

Date of Degree: MAY, 2022

Title of Study: EXPRESSION AND CHARACTERIZATION OF THE MEMBRANE PROTEINS TCR α WT AND γ-SARCOGLYCAN WITH THE QUANTIFICATION OF *IN VITRO* GLYCOSYLATED γ-SARCOGLYCAN

Major Field: CHEMISTRY

Abstract: Membrane proteins are a continually growing area of research in science as they are prevalent in a multitude of different disease states in humans. Membrane proteins are responsible for different biological processes in the cell such as cell-cell interactions, signal transduction, and transportation. However, very little is known about their structure and function because of their difficulty to work with them. Not only are membrane proteins difficult to express in high quantities, but they also contain highly hydrophobic portions which require detergents for solubilization to study these proteins. Expressing and purifying these proteins are only the first part in studying them, as most of them have post-translational modifications that can lead to structural and functional changes. N-linked glycosylation is a post-translational modification that can play a major role in the structural and functional changes of these membrane proteins by the attachment of a sugar and forming a glycoprotein. Therefore, using a novel *in vitro* technique, a chemoenzymatic reaction using an N-glycosyltransferase to attach a sugar moiety at a consensus sequence of Asn-X-Ser/Thr was performed. The attachment of the sugar moiety can influence physiological and biological properties of the protein by affecting their folding, modulating interactions with other biomolecules, and modifying their functions at the cellular level. We are specifically interested in the properties of membrane glycoproteins, which are key components in several different disease states. Therefore, γ-Sarcoglycan and TCR α, are the proteins being studied because they are linked to different disease states in humans and contain the consensus sequence for N-linked glycosylation. The ability to express and purify these proteins will be used to elucidate their structures through NMR spectroscopy and eventual determine changes in the proteins caused by glycosylation.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

CHAPTER I

RESEARCH OVERVIEW

This chapter will give a concise overview of this dissertation. Research objectives **(Figure 1.1)** were designed to look at membrane proteins and the specific post-translational modification (PTM) of glycosylation that is associated with many biological processes and human diseases. In the first section, the expression, purification, and preliminary characterization of T-cell receptor alpha Wild Type (TCR α WT) through Nuclear Magnetic Resonance (NMR) was performed for future membrane protein structure and function determination. The second section is focused on investigations into the expression, purification, and NMR studies, in different membrane mimetics, of γ-Sarcoglycan for future determination of the structure and function of the protein. The final section discusses the glycosylation of the membrane protein γ-Sarcoglycan and the approach to confirm and quantify the amounts of glycosylated γ-Sarcoglycan through trypsinolysis and mass spectrometry studies. This was the first time TCR α and γ-Sarcoglycan were recombinantly expressed, purified, and NMR studies were run on them. It is also the first time a membrane protein has been glycosylated *in vitro* through a chemoenzymatic reaction at a scale that could be quantified.

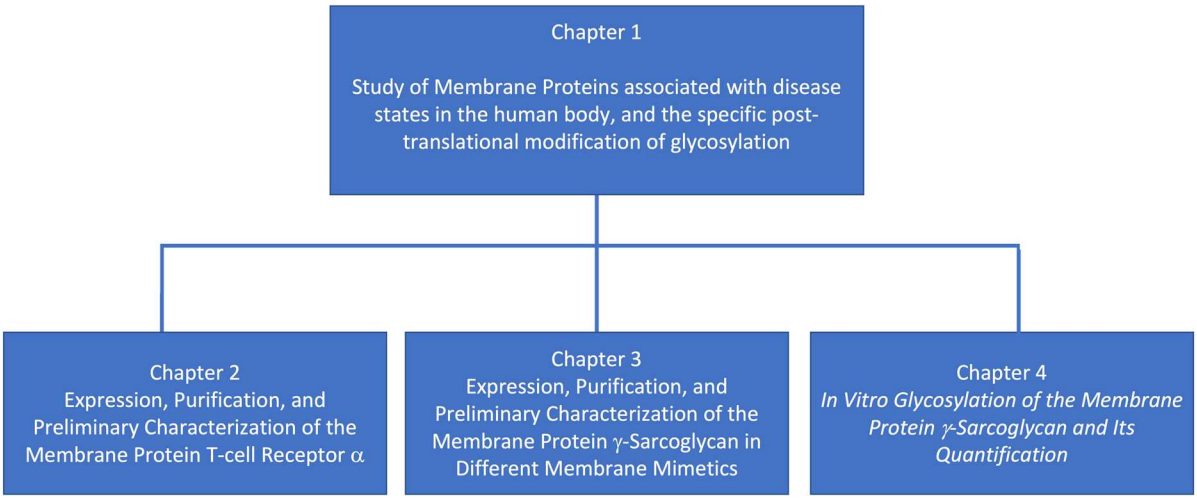

Figure 1.1 - Major research points in the dissertation

1.1 Introduction

Integral membrane proteins are proteins that are incorporated into or associated with cellular or organelle membranes. These membrane proteins are important as they have many different roles including cell signaling, transporting substances in and out of the cell, as receptors for hormones, and cell-to-cell interactions.[1] These integral membrane proteins are believed to account for 20-30% of genes in organisms.[2] However, the nature of these proteins are hydrophobic and amphiphilic which make them very difficult to study due to the challenges associated with overexpression, purification, and structural determination.[3] These proteins make up more than 40% of drug targets[4], yet very few membrane protein structures are actually known and represent less than 3% of available three-dimensional membrane structures.[5] In spite of all of these difficulties, it is imperative that we study them to find ways to combat the different diseases they cause. These studies could be helpful in finding novel therapeutic drugs in the future to combat these diseases.

With such little known about the structure and function of membrane proteins, it is a highly studied field. However, there are many challenges that comes with producing recombinant membrane proteins. The fact that it is very hard to express these membrane proteins in large quantities makes it very difficult to study as there is not enough protein to do some routine experiments that require just

nanogram or milligram amounts of protein. The hindrances needed to overcome when expressing these membrane proteins are protein aggregation, misfolding of overproduced membrane proteins, solubility issues due to their hydrophobic nature, and toxicity to the host cell which causes premature death. These issues are a few reasons that lead to minimal yields of the target protein.[6] Therefore, it takes a longer time to produce and characterize the needed quantities of protein which is usually in the milligram range.[7]

The aim of the introduction chapter is to give an overview for membrane proteins, recombinant protein expression and purification, use of membrane mimetics for structural studies, the characterization of membrane proteins using Nuclear Magnetic Resonance (NMR), and N-linked glycosylation. Through these techniques, different methods and experimental data was determined and described in the following chapters to further study membrane proteins *in vitro.*

1.1.1 Membrane Proteins

Membrane proteins incorporate themselves into or associate with membranes due to different interactions between the lipid membrane and protein. Examples of membrane proteins are integral membrane proteins and peripheral membrane proteins **(Figure 1.2)**.[8] Integral membrane proteins have hydrophobic surfaces which are incorporated into the hydrophobic portion of the lipid membrane while peripheral membrane proteins bind to integral membrane proteins through compatible binding sites or into a small portion of the cell membrane. Integral membrane proteins can be classified as either integral monotopic proteins, integral bitopic proteins or integral polytopic proteins depending on how they are associated within the lipid bilayer.[9] Integral bitopic and polytopic proteins are considered 'transmembrane proteins' as they transverse the lipid or cell membrane at least once. These transmembrane proteins have at least one highly hydrophobic segment (transmembrane domain) that usually consists of 19-23 amino acids which incorporates within the membrane due to hydrophobic effects.[8] A transmembrane protein is the an integral membrane protein which exists in

the membrane, and contains an extracellular domain, and intracellular domain. Commonly, these transmembrane domains consists of an α-helical structure which crosses the membrane.[10, 11] These proteins can also consist of beta barrels that have their hydrophobic residues pointing to the outside of the barrel to interact with the hydrophobic portions of the membrane.[8] Integral monotopic proteins exist on one side of the membrane and do not fully traverse the membrane, but they still have hydrophobic interactions with the membrane. Peripheral membrane proteins do not have hydrophobic interactions with the membrane, but rather temporarily associate with the membrane. Peripheral membrane proteins can attach to integral membrane proteins through different binding sites or can even interact with the lipid polar heads groups, but do not interact with the hydrophobic portion of the membrane.[12] The difference between the two being that integral membrane proteins are both hydrophobic and hydrophilic while peripheral membrane proteins are predominantly hydrophilic. The ability of transmembrane proteins to function both inside and outside of the cell membrane allows for them to transfer ligands or molecules across the cell membrane. These transmembrane proteins act as cell-surface receptors and transmit signals from the extracellular region of the cell membrane to the intracellular region of the cell.[13]

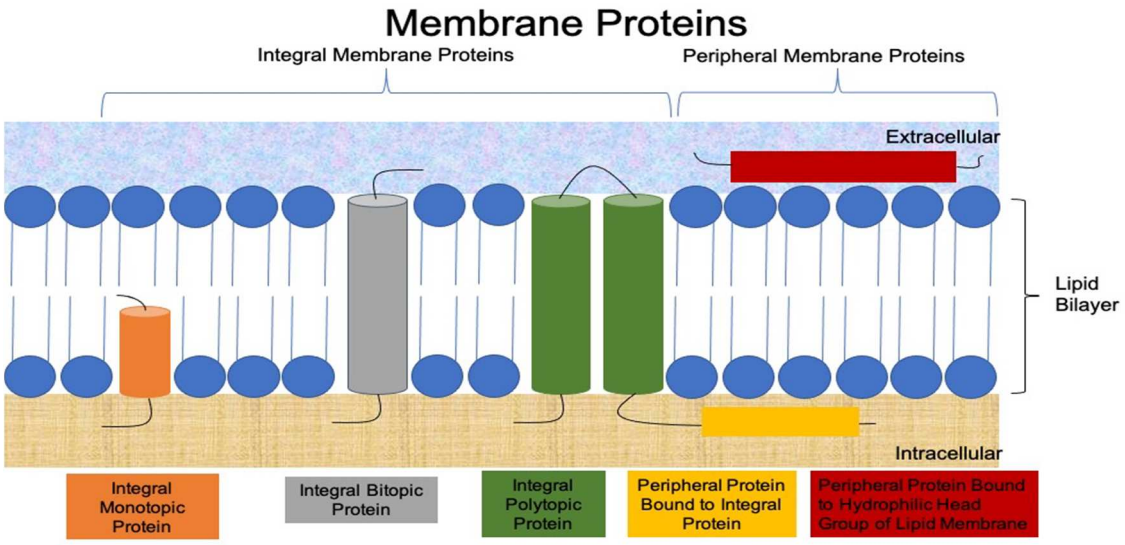

Figure 1.2 – Various membrane proteins and how they associate with the cell membrane and lipid bilayer.

Every year, studies and different developments within the membrane proteins field increases our ability to advance forward into new information about these troublesome proteins. To aid in these developments, this dissertation will give information on methods for expressing, purifying, and characterizing specific transmembrane proteins. These characterizations will lead to further knowledge on transmembrane proteins and will give helpful information in the future on ways to attack novel therapeutic drug discoveries to treat the diseases associated with these transmembrane proteins.

1.1.2 Recombinant Expression and Purification of Transmembrane Proteins

The most important part of studying transmembrane proteins is producing them in quantities large enough to run different experiments for determination of the protein structure. However, it is difficult to obtain sufficient amounts due to the fact that transmembrane proteins are usually present at low levels in biological membranes.[6] This shows the importance of having a planned out methodology that has been tested and optimized to get appreciable amounts of these proteins to study through biophysical techniques. Some important factors when optimizing expression are selecting a suitable overexpression system for recombinantly expressing the transmembrane proteins, labelling the protein with different isotopes, which purification techniques to use, and finding model membrane systems to study the protein in question. A common host for expressing these membrane proteins recombinantly is Escherichia Coli.[14] These methods are well established and have become the most popular expression platform of proteins. Many of its advantages include unparalleled fast growth kinetics, high cell density cultures for exponential growth, rich media from inexpensive components, and plasmid transformation in as little as five minutes.[15, 16]

One problem with the expression of transmembrane proteins is that overexpression on the membrane of a cell can affect the normal proliferation and homeostasis of the cell leading to minimal expression and cell death.[15, 17] One way to combat these problems is the utilization of a fusion proteins like TRP

leader which can aid the expression by directing the protein into inclusion bodies. This allows for high quantities of protein expression by avoiding premature cell death. Instead it leads to an easier purification step by using chemical or enzymatic cleavage to remove the fusion protein and isolate the protein of interest.[18]

After the expression of the membrane protein of interest, purification steps are then taken to isolate the protein. Upon isolation, further steps will be taken to analyze the structure and function of the protein. The recombinant expression and purification of TCR α WT and four of its mutants (Chapter2), γ-Sarcoglycan, Membrane Scaffold Protein (MSP), Tobacco Etch Virus (TEV) (Chapter 3), and N-glycosyltransferase (Chapter 4) will be described in depth throughout this dissertation.

1.1.3 Membrane Mimetics

NMR experimentation is an important tool for structure and dynamic determination of membrane proteins *in vitro*. However, for an accurate determination of the structure and dynamics of a membrane protein, a membrane mimetic is needed to represent a more native environment. This allows for the protein to be studied in its near-native structure in the cell membrane.[19] Therefore, after expression and purification of a target protein, these proteins can be reconstituted into these different membrane mimetics which are prepared using lipids, detergents, and even Nanodiscs from membrane scaffold proteins (MSP). Different types of membrane mimetics have different effects on the stability of the protein and on their characterization experiments through NMR. Therefore, it is important to study and test multiple membrane mimetics to determine which membrane will work best for your target protein.

Membrane mimetics are often made from detergent and lipids because they contain both a hydrophilic polar head group and a hydrophobic nonpolar tail group. **(Figure 1.3 A & B)** Having both a nonpolar and polar region is what makes detergents amphipathic and allows them to be suitable as membrane mimetics. In the monomer state, a detergent is typically cone shaped as the

hydrophilic head group occupies more space than the single alkyl chain group. **(Figure 1.3 A)** Lipids, in the monomer state, are typically cylindrically shaped unlike detergents as they consist of two hydrophobic alkyl side chains that occupy a space similar in diameter to the that of the hydrophilic head group. **(Figure 1.3 B)**[20]

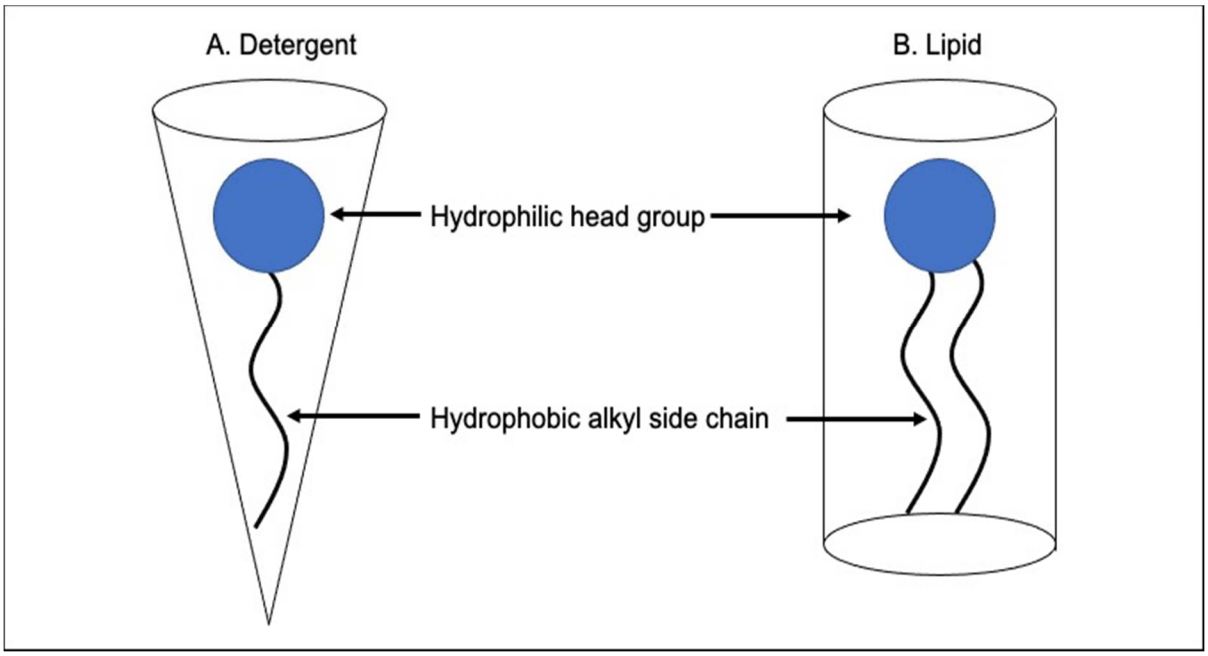

Figure 1.3 – Depiction of monomeric forms of A) a detergent and B) a lipid which shows the similarities of the structure of the hydrophilic head group and hydrophobic alkyl side chain(s), but the difference in their overall shape in space.

The detergent and lipid monomers aggregate in aqueous solutions as the hydrophilic head groups coordinate together while the hydrophobic effects bring the alkyl side chains together with high affinities. These hydrophobic alkyl side chains interact with each other to minimize their contact with the water, while the hydrophilic heads interact with water. The way the detergents and lipids interact give a suitable environment for protein incorporation and some even give optimal conditions for experimental studies.[21] Depending on the detergent or lipid used, the different membrane mimetics consist of micelles, bicelles, liposomes, and Nanodiscs **(Figure 1.4)**. Each of these have a unique assembly structure and can interact differently with each individual protein. Therefore, steps need to be taken to test and determine which specific membrane mimetic is the best for the target protein.

When determining a mimetic to use, it's important to investigate different factors such as the solubilization of the protein, ability to produce and allow the protein to be in its native state, retain enzyme activity, and gives a well-resolved NMR spectra.[22]

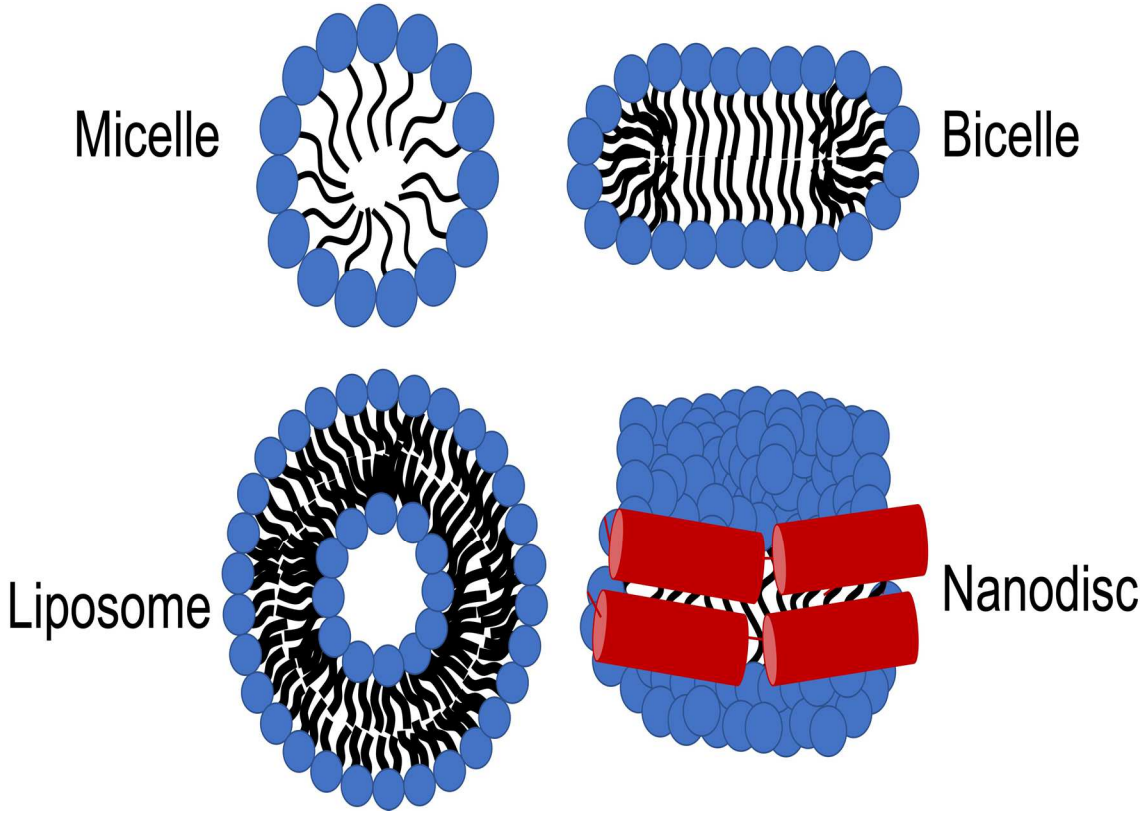

Figure 1.4 – Visual representation of a micelle, bicelle, liposome, and nanodisc which are common membrane mimetics used for studying integral membrane proteins.

The formation of these different membrane mimetics is dependent on the detergent or lipid being used. Micelles are formed in detergents that are dissolved in aqueous solutions at specific concentrations above the detergent's critical micelle concentration (CMC). These detergents have hydrophobic tail groups that repel water so they will structure themselves in water as shown in **Figure 1.4** to create a micelle where the hydrophilic head group is exposed to water, but the tail groups are not.[21] When looking at different detergents there are three different types and are classified by the polar head group. These three types are ionic, non-ionic, and zwitterionic. Ionic detergents have a charged polar head group whether it be positively charged (cationic) or negatively

charged (anionic). These detergents are considered harsh detergents as they are effective in solubilization of membrane proteins, but can also be a denaturant of proteins.[6] Examples of ionic detergents are sodium dodecyl sulfate (SDS) and sodium deoxycholic acid. Non-ionic detergents have a hydrophilic head that is uncharged and a tail that either consists of a polyoxyethylene like Triton-X or a glycosidic tail group like n-dodecyl-B-maltoside. However, these detergents do not interfere with protein-protein interactions, so they are not considered denaturing.[23] Zwitterionic detergents have characteristics of both ionic and non-ionic detergents as they have a neutral charge like non-ionic detergents. However, like ionic detergents, these detergents can be denaturing as they disrupt protein-protein interactions.[24] Examples of zwitterionic detergents used in this dissertation are 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) and 1,2-diheptanoyl-sn-glycero-3-phosphocholine (DHPC).

Other important characteristics of detergents and lipids are their CMC, micelle size, molecular weight, and the aggregation number of the detergent or lipid. Depending on the protein, looking at these characteristics may provide a starting point for determining the best detergent to use in characterizing the protein. These characteristics of detergents and lipids used throughout this dissertation are shown in **Table 1.1**. Upon determination of the detergents and lipids suitable for a protein NMR sample can be prepared for characterization using the membrane mimetics shown in **Figure 1.4**.

Table1.1 – Properties of detergents/lipids discussed throughout this dissertation.[20, 25, 26] The physical properties of detergents and lipids listed below are important when considering ways to optimize experimental conditions for structure and functional characterization of membrane proteins

Detergent/Lipid	Type	Molecular Weight (g/mol)	CMC (mM)	Micelle Size (Da)	Aggregation Number
SDS	Ionic	288.4	7-10	18,000	62
DM	Non-Ionic	482.6	1.8	33,300	69
DDM	Non-Ionic	510.6	0.1-0.6	50,000	98
Triton X-100	Non-Ionic	647	0.2-0.9	80,000	100-155
DHPC	Zwitterionic	481.5	1.4	N/A	N/A
Fos-Choline-12	Zwitterionic	351.5	1.5	19,000	54
Fos-Choline-16	Zwitterionic	407.5	0.013	72,500	178
CHAPS	Zwitterionic	614.9	6-10	6,150	10

An important factor in studying integral membrane proteins is the solubilization of the protein. Not only are the properties of the detergent and lipid characteristics, but also determining a suitable pH, temperature, and ionic strength is also important for the purification and characterization techniques. Making sure these conditions are optimized can mitigate chances of the protein unfolding.[27] A further look into the structure of each detergent/lipid above is shown in **Figure 1.5**. The membrane mimetics studied were all formed using these specific detergents and lipids.

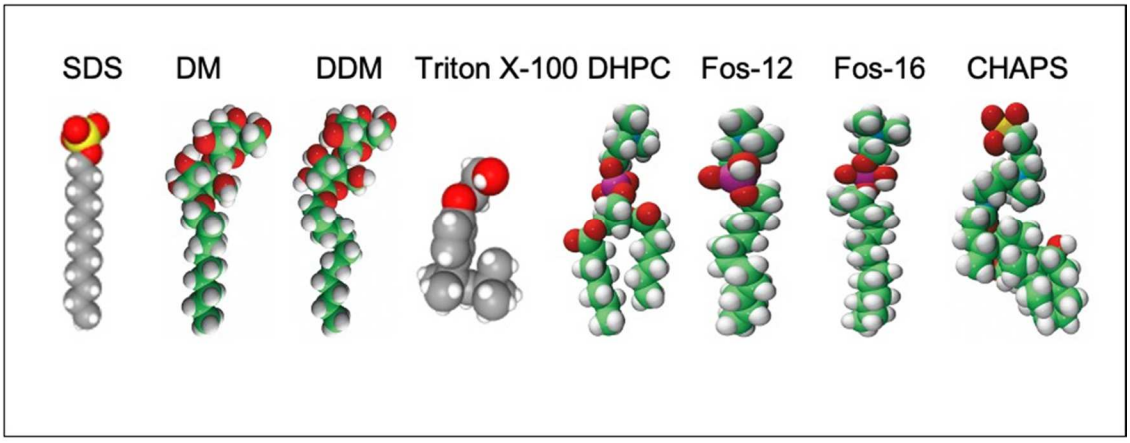

Figure 1.5 – 3D representation of the lipids/detergents used in the making of membrane mimetics.[28-30]

1.1.4 Solution State Nuclear Magnetic Resonance (NMR)

Membrane proteins structures, protein-protein interactions, protein-ligand, and protein dynamics can be studied using solution state NMR.[31] Atoms that are placed in a magnetic field, with non-zero spin, make transitions between different energy levels at specific resonance frequencies. These frequencies can then be differentiated and studied due to small differences in the chemical environment that surround these atoms.[32] These chemical shifts of atoms due to the environment are studied and are used as the most prominent feature of NMR spectroscopy.[33]

Depending on the size of the protein, you can use different atoms and NMR experiments to assign the peaks to specific amino acids. Smaller proteins, usually consisting of 100 amino acids or fewer, use a two-dimensional (2-D) homonuclear experiment which deals with ^{1}H-^{1}H correlations to assign chemical shifts. When studying proteins that consists of more than 100 amino acids, spectra can become more complicated due to overlapping chemical shifts and increased line width in the spectra. Therefore, heteronuclear NMR experiments are used to study these large proteins while specifically looking at correlations between ^{1}H, ^{13}C, and ^{15}N atoms. Studying these chemical shifts are done by different heteronuclear experiments to help determine structure and functions of larger proteins.[32, 33] Among different techniques to study proteins through NMR, solution state NMR continues to be the most insightful for integral membrane proteins as we can obtain structural and functional information on them in conditions that mimic their native environment.[5] Not only can solution state NMR be used for structural and functional determination, it can also be used for different studies dealing with dynamics, folding, lipid binding, ligand interactions, and ligand and drug binding to these integral membrane proteins.[34]

The most common experiment in membrane protein NMR is a 2D experiment called ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectroscopy. Through scalar coupling, magnetization is transferred from a proton to the ^{15}N nuclei evolved and transferred back to the proton for detection by the instrument. Data that can be gleaned from these experiments are the backbone

amide groups and the side chains on the amino acids tryptophan, asparagine, and glutamine as they all contain a ^{15}N labelled nuclei. Each amino acid having only one backbone amide, allows for further studies as each signal or peak corresponds specifically to one amino acid residue. Not only are there 2D experiments, but also 3D experiments that can be used for structure determination. In this dissertation the experiments used deal with ^{1}H, ^{13}C, and ^{15}N atoms. The four 3-D experiments run for structural determination were HNCA, HN(CO)CA, HNCO, and HN(CA)CO. In HNCA experiments, magnetization is passed from ^{1}H to ^{15}N on the backbone of the protein. From there, it is transferred via N-C_α J-coupling to the $^{13}C_\alpha$. The magnetization is then passed back to ^{15}N and ^{1}H for detection while the chemical shift is evolved on all $^{1}H^{N}$, $^{15}N^{H}$, and $^{13}C\alpha$ resulting in a 3D spectrum.[35] A visual representation of a HNCA experiment is shown in **Figure 1.6A**. HN(CO)CA experiments are a result of the magnetization being passed from ^{1}H to ^{15}N and then onto ^{13}CO. The magnetization is then transferred to the $^{13}C\alpha$ and the chemical shift is evolved. This is then transferred back through the ^{13}CO to the ^{15}N and ^{1}H for detection. The chemical shift is evolved for all the atoms except the ^{13}CO. This allows for a spectrum similar to that of an HNCA, but it is selective to the previous residue and a visual representation of an experiment is **Figure 1.6B**.[36]

HNCO experiments deal with the magnetization transfer from ^{1}H to the ^{15}N atom then selectively to the ^{13}C carbonyl. This is done through $^{15}N^{H}$-^{13}CO J coupling and the magnetization is then passed back via ^{15}N atom to the proton for detection. The chemical shift is evolved on all three atoms creating this 3-D spectrum.[37] In HN(CA)CO experiments magnetization is transferred from the ^{1}H atom to the ^{15}N atom then to the $^{13}C\alpha$ atom through J-coupling. It is then transferred to the ^{13}C carbonyl via $^{13}C\alpha$-^{13}CO J-coupling. The magnetization is then transferred through ^{13}CO to $^{13}C\alpha$ to the ^{15}N and lastly to the ^{1}H. The chemical shift evolves on all the atoms besides the $^{13}C\alpha$. This transfer occurs on either side of the ^{15}N atoms so two carbonyl groups are observed in the spectrum. Since the coupling with the $C\alpha$ immediately attached to the nitrogen is stronger the carbonyl attached to that $C\alpha$ will have a stronger signal than the carbonyl directly attached to the nitrogen atom.[38] The

magnetization transfer and evolution of the signal is shown below for both HNCO and HN(CA)CO as **Figure 1.6C and Figure 1.6D** respectively.

These 4 experiments, HNCA, HN(CO)CA, HNCO, and HN(CA)CO in conjunction with the HSQC experiments can help assign specific residues to different peaks on the spectra. Determination of the specific residues and how they lineup in the backbone can allow for secondary structure determination of the protein.

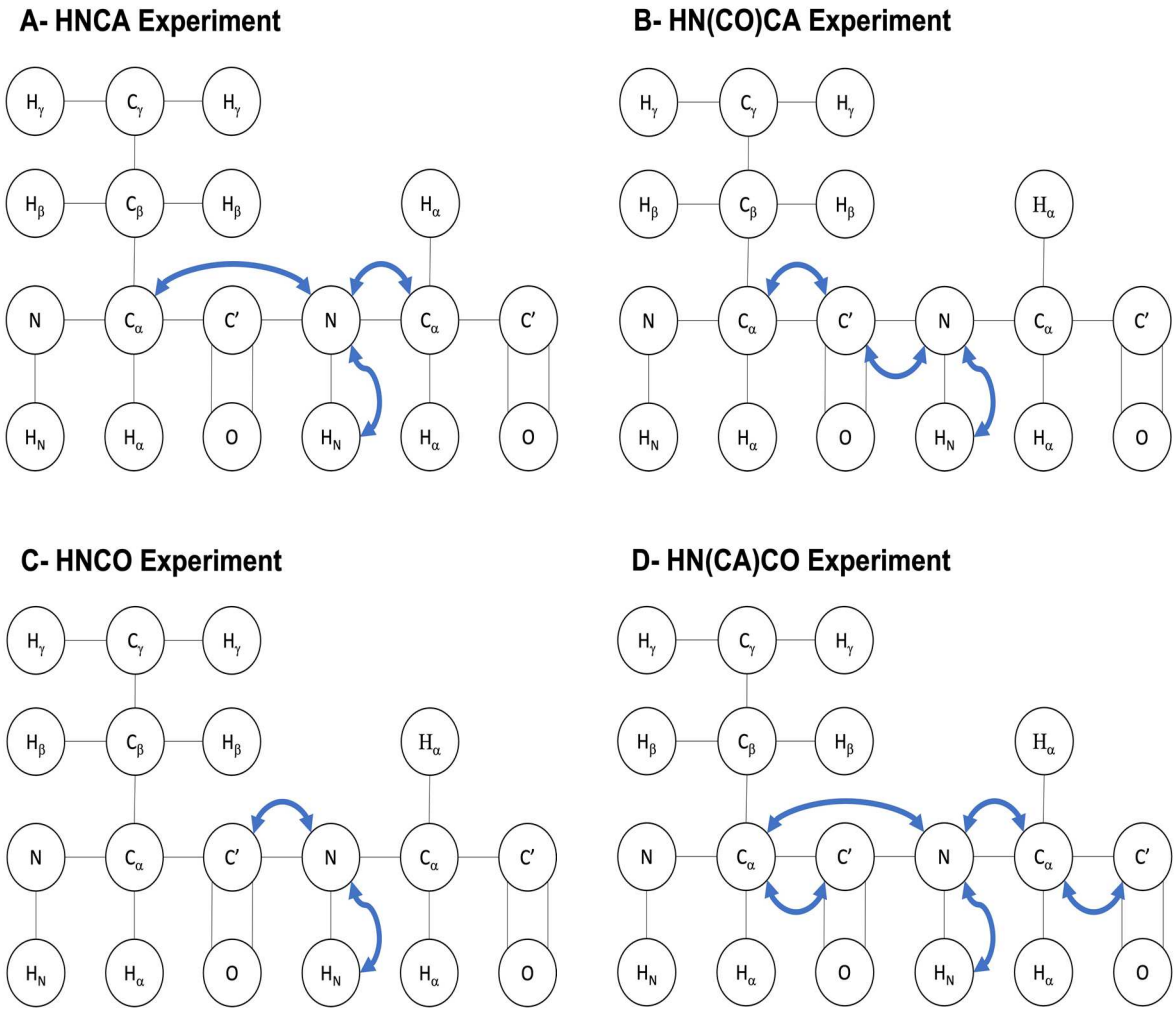

Figure 1.6 A, B, C, & D- Visual representations of the magnetization exchange from atom to atom in four different experiments. These experiments consist of A) HNCA NMR experiment B) HN(CO)CA NMR experiment C) HNCO NMR experiment and D) HN(CA)CO NMR experiment.

These different experiments described throughout this section are going to be used as preliminary data to help determine backbone assignments of amino acids in both TCR α WT (Chapter 2) and γ-Sarcoglycan (Chapter 3). The assignment of the backbone amino acids and the determination of the secondary structure of the proteins can lead to the measurement of structural changes. Future studies will involve using N-glycosyltransferase for the attachment of a sugar moiety to different membrane proteins. These glycoproteins will then be studied for structural changes that are a result of the attachment of the sugar and see if they play any role in the disease states in human beings.

1.1.5 N-linked Glycosylation

Post-translational modifications (PTMs) are generally covalent, enzymatic modifications of protein after their translation. These modifications are important to study in membrane proteins because these modifications affect the function of proteins upon their attachment. These attachments make the human proteome more complex for this reason.[39] Therefore, in studying different membrane protein involved in diseases, it is important to study different PTMs and the effects they have on these membrane proteins.

The specific PTM studied and talked about throughout this dissertation is N-linked glycosylation. N-linked glycosylation is the attachment of a sugar to an asparagine residue within the consensus sequence of Asn-X-Ser/Thr (N-X-S/T), where X is as any amino acid besides proline. This reaction can be catalyzed in eukaryotes by the membrane bound enzyme oligosaccharyltransferase (OST) which uses lipid-linked oligosaccharides as the sugar donor.[40] Another enzyme that has been shown to perform glycosylation is N-glycosyltransferase (NGT). It does so in a similar manner, at the same consensus sequence.[41] These asparagine-linked modifications are involved in nearly every cell process including protein folding, transduction, secretion, and cellular recognition.[42-44] This dissertation is focused on the use of NGT for *in vitro* glycosylation of the membrane protein γ-Sarcoglycan in its naturally folded state as a model for other membrane proteins.

This chemoenzymatic method of glycosylation with NGT has been shown to be effective with soluble proteins. However, there has been no shown success of this glycosylation with membrane proteins. Studying these membrane proteins with glycosylation is essential as they make up 20 to 30% of the proteome of multiple organisms and a majority of these are glycosylated proteins better known as glycoproteins.[45] Since membrane proteins have very hydrophobic regions, increased flexibility, and a lack of stability, it makes it difficult to solubilize the protein and study them.[3] Detergents and lipids are needed to solubilize the protein. The problem faced then is believed that the activity of the enzyme NGT will not be active to attach a sugar moiety onto the proteins.

Therefore, a model stepwise method for the *in vitro* glycosylation of the membrane protein γ-Sarcoglycan was determined in this paper. This information will all be described in Chapter 4 of this dissertation and give further insights on the troubles of glycosylating membrane proteins. These methods will also include the process used for quantifying the glycosylation of the membrane protein through LC-MS/MS. This stepwise method of *in vitro* glycosylation is vital, and once is optimized for glycosylation and purification of the glycoprotein, it can lead to future studies dealing with structural and functional changes in a multitude of membrane proteins that have yet to be studied.

1.1.6 Summary and Outlook

Currently, it is estimated that 60% of drug targets are membrane proteins. However, despite this number, only about 3% of membrane proteins structures have been determined. This data shows the importance and demand for further studies and expression of these membrane proteins for structural and functional determination. The ability to determine the structure-activity-relationship (SAR) of these proteins can give insight into ways of treating different diseases dealing with membrane proteins. Eventually the goal will be to find ways to produce pharmaceutical drugs that combat these diseases.

These chapters will discuss two specific membrane proteins TCR α (Chapter 2) and γ-Sarcoglycan (Chapter 3). Chapter 2 and 3 will discuss the process of protein expression and purification to start and finish with NMR analysis of the proteins in membrane mimetics for structural studies. Chapter 4 will explore a novel *in vitro* technique for the glycosylation of the membrane protein γ-Sarcoglycan. Not only will it show the glycosylation of the hydrophobic membrane protein, but also it will show the quantity of protein glycosylated through LC-MS/MS studies.

Abbreviations: Post-translational Modification (PTM); T-Cell Receptor (TCR); Nuclear Magnetic Resonance (NMR); Tobacco Etch Virus (TEV); Membrane Scaffold Protein (MSP); Critical Micelle Concentration (CMC); Sodium Dodecyl Sulfate (SDS); Decyl β-D-maltopyranoside (DM); n-Dodecyl β-D-maltoside (DDM); 1,2 diheptanoyl-sn-glycero-3-phosphocholine (DHPC); n-Dodecylphosphocholine (Fos-choline-12 or DPC); n-Hexadecylphosphocholine (Fos-choline-16); 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS); Heteronuclear Single Quantum Coherence (HSQC); Oligosaccharyltransferase (OST); N-glycosyltransferase (NGT); Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS); Structure-Activity-Relationship (SAR)

CHAPTER II

EXPRESSION AND PURIFICATION FROM THE T CELL RECEPTOR COMPLEX OF TCR α

2.1 Introduction

2.1.1 T-Cell Activation

T-cells, also known as T-lymphocytes, are a type of white blood cell that acts as immune response to foreign antigens in the human body. They adapt to the antigen that can range in form from microbes to cancer, and they denature it, destroying the antigen acting as a defense system. These T-cells are activated by a multisubunit complex consisting of T-cell receptors (TCR) subunits, and clusters of differentiation 3 (CD3) which all contain immunoreceptor tyrosine-based activation motifs (ITAMs). Throughout this process the TCR heterodimer has two functions consisting of antigen recognition and signal transduction. For this signaling to occur, it is essential that phosphorylation of the ITAMs occurs.[46] However for recognition of antigens, the TCR heterodimer binds peptide fragments of major histocompatibility complex (MHC) molecules.[47] There are two classes of MHC molecules. The first class of molecules are expressed on nearly every nucleated cell and present antigens to CD8, which is a transmembrane glycoprotein expressed on cytotoxic T-cells.[48] The second class of molecules are expressed on macrophages, B cells, and dendritic cells which deal with antigens being processed by T-cells with CD4. All of this occurs in a space known as immunological synapse and helps stabilize the

MHC/TCR interaction which leads to an autoimmune response mediated by antibodies **(Figure 2.1)**.[49, 50]

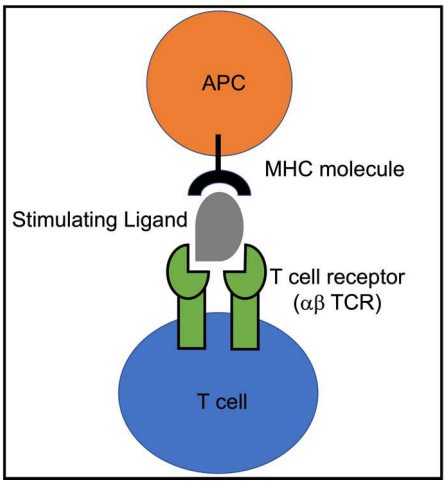

Figure 2.1- Antigen recognition by TCR requires a three-molecule interaction between TCR and a stimulating ligand which is typically a peptide and interacting with an MHC on an antigen presenting cell (APC). This leads to other interactions with the CD3 and zeta complex for signal transduction.[51]

2.1.2 T-Cell Receptors

T-cell receptors are expressed as 4 polypeptide chains which are named alpha (α), beta (β), gamma (γ), and delta (δ). However, with all 4 of these being expressed in humans, the heterodimer TCR αβ make up 95% of the TCR population while another subset of T cells express the other two peptides, γ and δ, making up only 5% of the population.[52] The T cell receptors consist of a small cytosolic tail, a transmembrane domain, and an extracellular domain which contains both a variable and constant domain within it. (**Figure 2.2)** In the constant domain there is a site for N-linked glycosylation and between the αβ dimers there is a disulfide bond linkage. The most important feature of this dimer is its variable domain which is the antigen recognition site of the dimer. Upon antigen binding, signal transduction can occur, but this does not come from the T cell receptor as the cytosolic tail portion does not induce intracellular signaling.[53] Therefore, the CD3 complex is needed for intracellular signaling which then leads to T cells releasing cytotoxins to kill the foreign antigen.[54] Therefore, the mediation of the T cell receptors

with the CD3 complex through its transmembrane is crucial for the transduction of signals upon antigen recognition by TCR.

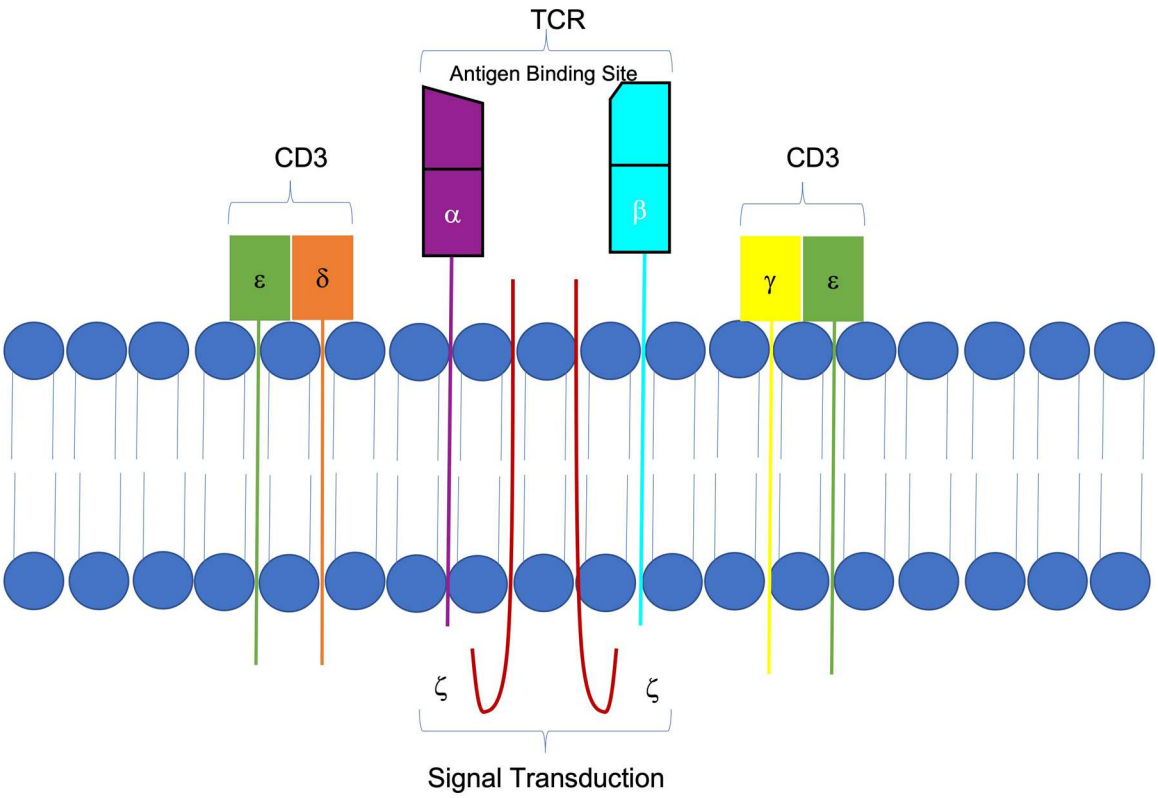

Figure 2.2- Structure of the common dimer form TCR αβ. The structure contains a cytosolic intracellular portion, transmembrane domain, and an extracellular domain. Important features in the extracellular domain are the variable domain where antigen binding occurs, a disulfide bond linkage, and an N-linked glycosylation site in the constant domains. This figure is adapted from Karki et al.[55] and Franco et. al.[55, 56]

2.1.3 Literature Review

The TCR-CD3 complex has been extensively studied in the past few years and structural determination for parts of the protein.[57] TCR is an innate part of the auto-immune response in human beings. This is due to the variable region of TCR that can bind specific antigens presented by MHCs and MHC-like molecules.[58] Even more impressive is the low concentrations of these complexes needed to bind to TCR to activate the T cells. It has shown that even weak binding of these ligands around 1-200uM is sufficient for this activation.[59, 60] Even though TCR can recognize these foreign antigens, they cannot signal the activation of the T cells. Therefore, they

must work in concert with different CD3 proteins inside this TCR-CD3 complex shown in **Figure 2.2**. Specifically, non-covalent interactions are the driving force between TCR and dimer forms of CD3 including CD3εγ, CD3εδ, and CD3ζζ.

The interactions as shown in Figure 2.2 between CD3 and TCR are through the transmembrane domain through charged residues.[61-63] A deeper look into the structure of CD3, the protein that leads to signal transduction shows that the ε, γ, and δ all have a single Immunoglobulin domain in the extracellular region and a single immunoreceptor tyrosine-based activation motif (ITAM).[64, 65] However, the ζ subunit of CD3 is different in which it has a very short extracellular domain and has 3 ITAMs.[66] Upon phosphorylation of the intracellular ITAMs and the interactions with the non-catalytic region of tyrosine kinase (Nck) signaling occurs and leads to T cell activation, proliferation, and survival.[67, 68] All these underlying effects that lead to T cell activation play important roles, but none of this is possible without the TCR recognizing these foreign antigens leading to this signaling cascade. Determining how TCR binds different ligands can lead to more information on how this whole process begins. Therefore, studies in the structure of the TCR can help us understand which ligands it binds to and the reasons behind its binding.

2.1.4 Summary and Outlook

TCR α is a 31 amino acid (3.5 kDa) protein that play a crucial role in antigen recognition in dimer form with TCR β. Upon this recognition it interacts with the CD3 complex which transduces signals leading to a cascade of signals for T cells to attack and kill foreign antigen cells. The purpose of this study is to express and purify the TCR α protein for different studies and to eventually elucidate the structure through NMR and electron paramagnetic resonance (EPR) studies. NMR studies will be done in different membrane mimetics using the wild-type protein shown in **Figure 2.3**. Throughout this chapter, TCR α was recombinantly expressed for the first time and a good membrane mimetic was determined for future studies of TCR α WT for

NMR studies. However, for EPR studies, 4 different mutants **(Figure 2.3)** of the TCR α protein in which a cysteine is mutated in at the specific sites L14C, K16C, A18C, and M24C will be used. These mutations to a cysteine residue allow for a chemical reaction to occur and the attachment of S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl mehanesulfonothioate (MTSL) onto cysteine for site-directed spin labelling (**Figure 2.4)**.[69] MTSL has a nitroxide spin label with an unpaired electron which enables structural studies on it to be done through EPR. These different studies will be crucial in determination of the secondary structure and orientation of TCR α in the in bilayers. The specific mutations were chosen to study the orientation of the transmembrane domain in the lipid bilayer.

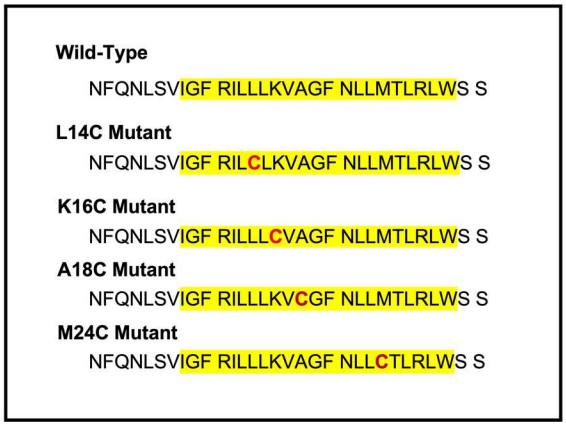

Figure 2.3- Amino acid sequence of TCR α WT and the four mutants L14C, K16C, A18C, and M24C used for structural studies. The red letters are the specific amino acids that were mutated in each variant of TCR α, and all are mutated to a cysteine.

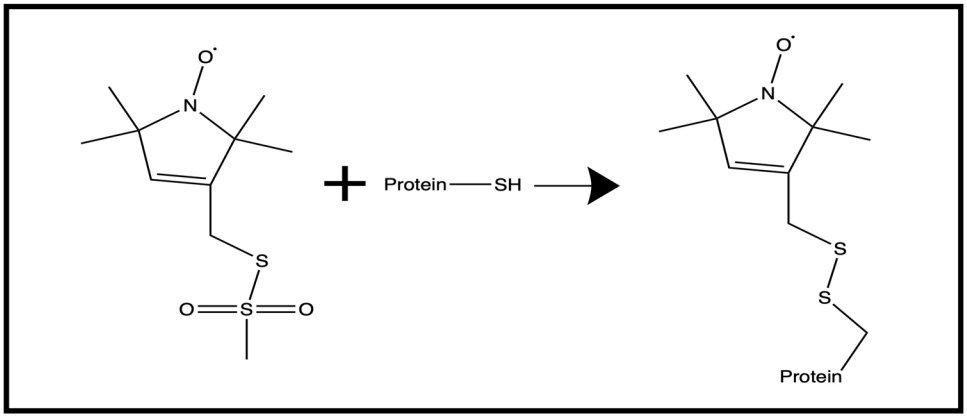

Figure 2.4- Attachment of MTSL with a spin-labeled nitroxide group to the amino acid cysteine, so the protein can be studied through EPR.

2.2 Materials and Methods

2.2.1 Expression of Unlabeled TCR α WT and Mutants

A pUC-19 plasmid from Genewicz (www.genwicz.com) with carbenicillin resistance and the correct DNA sequence for each the wild type and four mutants L14C, K16C, A18C, and M24C were transformed into BL21 cells for growth. Transformation cells were plated on LB Agar media plates supplemented with carbenicillin. This was left overnight in the 37°C incubator for cell colonies to grow. A cell stock solution was then made by inoculating 10 mL of LB/carbenicillin media with a single cell colony from the plate and was incubated on a shaker in a 37°C room. The optical density at 600 nm (OD_{600}) was checked until it reached 0.4. Sterilized glycerol at 70% was then added to the remaining solution until it had a concentration of 17%. Upon addition of the glycerol, 50 µL of cells were aliquoted into 1.5 mL eppendorf tubes and were placed in the -80°C freezer for storage, and to be used for expression growths.

For the expression of the wild type and mutant proteins, a starter culture was prepared by inoculating 5 mL of LB media with 5 µL of carbenicillin at a concentration of 100 mg/mL and 10 µL of the TCR wild type or mutant cell stock that was being grown. This starter culture was then placed on a shaker at 220 rpm 37°C for 2 h. After 2 h the sample was removed from the shaker and 200 µL were transferred to each of the two 250 mL overnight growth flasks containing 100 mL of LB media. 100 µL of carbenicillin at 100 mg/mL was also added to the overnight growth solutions. These overnight growth flasks were then placed back in the 37°C room and left shaking at 220 rpm for approximately 14 h. The next morning 50 mL of the overnight growths were then transferred to four different 2 L growth flasks containing 950 mL of LB media so each large growth flask will have 1 L of media. 1 mL of carbenicillin at 100 mg/mL was then added to the media and the large growth flasks were placed once again on the shaker in the 37°C room until the OD_{600} reached 0.6. This takes approximately 3 hours. Then growth was

then induced with 1 mL of Isopropyl β-D-thiogalactopyranoside (IPTG) (GoldBioTechnology, www.goldbio.com) at 120 mg/mL to each 1 L growth. After induction with IPTG the sample was placed back on the shaker in the 37°C room at 220 rpm for 4 h. The cells were harvested by transferring them to a 1 L polycarbonate centrifuge bottles and centrifuged at 5450 xg for 25 min at 4°C (Thermo scientific, SORVALL LINX 4000 Centrifuge). The cell pellets were then saved and stored in a -80°C freezer until they were to be purified.

2.2.2 Expression of ^{15}N-labeled TCR α WT

When expressing proteins that are in ^{15}N-labeled like TCR α WT, a 10 μL cell stock was inoculated in 5 mL LB media with 5 μL of carbenicillin at 100 mg/mL and grown in 37°C for 2 h exactly as described in the expression of unlabeled TCR α WT and mutants. 200 μL of the starter culture was then added to 100 mL of sterile M9 minimal media (7.5 mM of ^{15}N ammonium sulfate, 39 mM glucose, 40 mM sodium phosphate dibasic, 20 mM sodium phosphate monobasic, 9 mM sodium chloride, 1 mM magnesium sulfate, 0.1 mM calcium chloride, and 1% LB media). 100 μL of carbenicillin was added to the 100 mL M9 minimal media solution and placed in the 37°C room and left shaking at 220 rpm for 16 h. After the overnight growth of 16 h, 50 mL was transferred to back to 950 mL of the M9 solution, so it had a total volume of 1 L. Upon this addition, 1 mL of 100 mg/mL of carbenicillin was added and the four large growth flasks were placed on the shaker in the 37°C room until the OD_{600} reached 0.6. This took approximately 4 h and then the sample was induced with 1 mL of IPTG at a concentration of 120 mg/mL and was let grow for 4 h while shaking at 220 rpm in a 37°C room. The cells were then harvested according to the expression of unlabeled TCR α WT and mutants.

2.2.3 Purification

The cell pellets from the expression of the TCR α WT protein were then resuspended in 30 mL of Resuspension Buffer I (50 mM Tris-HCl, 1mM EDTA, 1 mM NaN_3, and 205 mM glycerol). Upon resuspension of the pellet, the sample was sonicated (Fisher Scientific Sonic Dismembrator, www.fishersci.com) on ice for a total of 4 minutes at intervals of 2 seconds on then 8 seconds off. The total elapsed time for the sample in the sonicator was 20 minutes. The samples were then transferred to a 35 mL centrifuge tube and counterbalanced with other tubes and spun down in the centrifuge at 35000 xg for 25 min at 4°C (Thermo scientific, SORVALL LINX 4000 Centrifuge). After centrifugation, the supernatant was discarded, and the cell pellet containing the membrane fraction and inclusion bodies were saved.

The cell pellet is then resuspended once more in Resuspension Buffer II (50 mM Tris, 1 mM EDTA, 1 mL NaN_3, 25.5 mM Deoxycholic acid, 17.5 mM IGEPAL). Vortexing and a spatula was used to break up the cell pellet to help resuspend the pellet. The solution was then sonicated on ice for a total of 4 minutes at intervals of 2 seconds on and 8 seconds off so that a total of 20 minutes elapsed during sonication. Samples were once again transferred to a 35 mL centrifuge tube and counterbalanced with other samples. The tubes are then spun down at 35000 xg for 25 min at 4° C. Following centrifugation, the supernatant is once again discarded.

The pellet, which contained the isolated inclusion bodies, was broken up in 30 ml of Binding Buffer (6 M guanidine hydrochloride, 0.5 M sodium chloride, and 20 mM Tris pH 8.0) and stored at 25°C overnight. The partially dissolved inclusion bodies were then sonicated on ice as previously described with the two resuspension buffers. After the sonication, the samples were spun down in the centrifuge at 18500 xg for 10 min at 4°C to remove any particulates that remained. The supernatant was then saved and dialyzed against 5 L of ddH_2O in 3.5 kDa dialysis tubing (Spectrum Laboratories, Inc., spectrumlabs.com). The water was changed at 30-minutes

and 90-minutes, then left overnight for dialysis. After the removal of guanidine is completed, the protein solution is frozen and dried on the lyophilizer overnight.

Next the fusion partner, Trp Leader, was cleaved off by treating the protein with cyanogen bromide (CNBr) (Alfa Aesar, www.alfa.com). When running this cleavage reaction, we used 70% formic acid (Fisher Chemical, www.fishersci.com) and cyanogen bromide. For every 10 mg of the target protein, we added 20 mg of CNBr and 400 µL of formic acid. Upon addition of the formic acid and CNBr, the sample was placed on a rotator in the absence of light for 3 hours to push the cleavage forward. After 3 h the sample was removed from the rotator and 1 Vol. of 1 M NaOH was added to the cleavage mixture and placed back on the rotator for 10 min. A white precipitate formed. 1 Vol. of the original solution of DI H_2O is then added to the sample and then it is transferred to a pre-wet 3.5 kDa dialysis bag. The reaction tube is then washed once more with 1 Vol. of DI H_2O and then transferred to the same dialysis bag.

This sample was then dialyzed against 4 L of DI H_2O to neutralize the sample and to remove any remaining CNBr. The water was switched after 30-minutes and 90-minutes and then left. The wastewater was neutralized to a pH of above 7 and placed in a waste container to be disposed by waste management. The sample was then transferred to a 50 mL centrifuge tubes and frozen in the -80°C freezer. Once the sample was frozen it was placed on the lyophilizer to dry the sample. The sample was taken off the lyophilizer once dry and prepared for FPLC Size Exclusion Chromatography (SEC).

For FPLC SEC runs, 7 mg of TCR α WT was dissolved in 125 µL of 10% SDS with vortexing. The sample was then placed in the sonication bath for 10 minutes until the protein was dissolved in solution. Once it was dissolved, 575 µL of FPLC Buffer (20 mM Sodium Phosphate, 4 mM SDS, 1mM EDTA, and 1mM NaN_3) at pH 9.6 with 100 mM NaCl was added to the sample. The sample was then centrifuged for 5 minutes and only the supernatant was kept for the FPLC run.

The sample was loaded through a filter and into a two-column system and the first column was a Superdex™ 200 increase 10/300GL and the second column being a Superdex™ 75 increase 10/300GL (GE Healthcare, www.gelifesciences.com). The sample was run on a Bio-Rad NGC Quest 10 Plus (www.bio-rad.com) at a flow rater of 0.5 ml/min. The elution of the protein was monitored using absorbance at 280 nm and the sample peaks were collected and analyzed using an SDS-Page gel. The fractions containing the pure protein were then pooled together and dialyzed against 5 L of DI H_2O in a 3.5 kDa dialysis bag. This was done to remove the detergent. Precipitated protein was then centrifuged, frozen, and dried on the lyophilizer. This protein could be used for mass spectrometry and NMR studies.

2.2.4 Mass Spectrometry

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) spectrometry was performed on a DE Pro mass spectrometer (AB SCIEX). The purified unlabeled TCR α WT protein was prepared in two ways. For the first, the protein was dissolved in 98% acetonitrile and (Alfa Aesar, www.alfa.com) 2% trifluoroacetic acid (Acros Organics, www.acros.com) and was mixed with a matrix solution of sinapinic acid (Asta Tech, www.astatechnic.com) dissolved in 50% acetonitrile 50 % water, and 0.1% trifluoroacetic acid. For the second way, the pure protein was dissolved in 98% trifluoroethanol (Acros Organics, www.acros.com) and 2% trifluoroacetic acid solution, which was then mixed with the same matrix solution as the first procedure.

2.2.5 NMR Sample Preparation and Spectrometry

Three different detergents were used for solution-state NMR experiments. These detergents were DHPC, Triton X-100, and Fos-12. Each sample was prepared for a total of 500 µL with the final sample having a 90% H_2O/10% D_2O. The DHPC (Avanti Polar Lipids Inc., www.avantilipids.com) samples are made by using 156.25 µL of 400 mM DHPC to dissolve ~1.0 mg of the ^{15}N TCR α WT or mutant protein. The pH was then lowered to 4.5 by the addition of 4

μL of 1 M HCl to help with solubility, and 293.75 μL of DI H_2O and 50 μL of D_2O (Acros Organics, www.acros.com). The mixtures were each placed in a sonication bath. The final concentration of the sample was then 125 mM DHPC. The Triton X-100 (Amresco, www.amresco-inc.com) sample was made and prepared as an NMR sample by starting with a stock solution of 4% w/v Triton X-100. 125 μL of this solution was added to 1.3 mg of the pure γ-Sarcoglycan protein. 4 μL of 1 M HCl was added to the sample and the sample was placed in the sonication bath until the sample was fully dissolved. Upon the solubilization of the protein 325 μL of H_2O and 50 μL of D_2O was added to the solution and the pH was checked and determined to be 4.0. Fos-12 NMR solution was prepared for a final concentration of 150 mM. 250 μL of a 300 mM stock solution was added to 1.0 mg of TCR α protein and 4 μL of 1M HCl was added as well. The sample was then placed in a sonication bath for 1 h to help with solubilization of the protein into Fos-12. Once dissolved 200 μL of DI H_2O and 50 μL of D_2O was added to the sample. The final sample had a concentration of 150 mM Fos-12 and had a pH of 4.0.

All 3 samples were then centrifuged to get rid of any particulates and the supernatant transferred separately to a standard 5 mm tube to run using an HSQC experiment with a correlation between ^{15}N and ^{1}H atoms for preliminary studies dealing with secondary structure of the protein. The NMR experiments were performed on a Bruker DMX 500 MHz Spectrometer (www.brukerbiospin.com). A triple resonance $^{1}H/^{13}C/^{15}N$ TXI high resolution NMR probe was used. The pulse sequence, hsqcetgpsi2 (www.brukerbiospin.com), was calibrated and the water peak was referenced to 4.7 ppm. The experiments were performed at 25°C. The ^{1}H-^{15}N HSQC experiment was run with 4092 t2 points and 256 t1 points. The data was processed in nmrPipe and visualized in Sparky (T.D Goddard and D.G Kneller, SPARKY, University of California, San Francisco).

2.2.6 Labeling of TCR α Mutants with a Nitroxide Label

Labeling was done only with TCR α mutant proteins as the cysteine residues would be the amino acid that would react with the label. When labeling, the multiple protein FPLC fractions (~10 mL) were combined and used, and they were not dialyzed. The sample was then concentrated down in a 3.0 kDa Centrifugal Filter Units for concentration (Merck Millipore Ltd. www.millipore.com). The sample's pH was then adjusted to just below 7 with 1M HCl, as the label isn't reactive in basic conditions. 30 µL of acetonitrile (ACN) was used to dissolve the label which *S*-(1-oxyl-2,2,3,5,5-pentamethylimidazolidin-4-ylmethyl) ester (IMTSL). This IMTSL was provided by the Smirnova group at North Carolina State University and added to the protein solution at a 60-fold concentration compared to the 1 mg of the protein used in solution. The sample was then left on the rotator at room temperature for 48 h. The next morning the sample was transferred to a 3.5 kDa molecular weight cutoff slide-a-lyzer dialysis cassette (thermoscientific.com/protein) and was dialyzed against 20 mM phosphate buffer at a pH of 6.8 for 6 h.

After 6 h, 4.1 mg of DDM was added to the sample so the concentration of DDM was 16 mM. The sample was placed back into the slide-a-lyzer cassette and dialyzed for 6 h in 20 mM phosphate buffer at pH 6.8. Meanwhile a bicelle sample was made by preparing a 1200 µL sample of 8 mM DDM and adding 77.2 mg of DMPG. Ice bath, heating, and vortex was used to dissolve the sample. Once dialysis ended the two samples were combined for a sample with 1700 µL including the protein, detergent, and lipids. The sample was placed back in 20 mM phosphate buffer at pH of 6.8 with buffer changes every 6 h until the sample turned opalescent. Once the sample turned opalescent, the proteoliposome is formed and the sample is ready to be centrifuged. The sample was centrifuged in a Beckman Coulter ultracentrifuge (Optima L-90K) at

118,000 xg and 15°C for 2 h. The sample was then frozen in the -80°C freezer and shipped to North Carolina State University for electron paramagnetic resonance (EPR) studies.

2.3 Results

The first thing step for preliminary studies of the secondary structure is secondary structure prediction of TCR α followed by recombinant expression and purification of the full-length protein. Therefore, the protein was recombinantly expressed in *Escherichia coli (E. Coli)*. This will express the intracellular, extracellular, and transmembrane domains. E Coli is the most common and abundantly used bacterial host when expressing recombinant proteins. However, overexpression of proteins in cell membranes can be detrimental to bacteria, so the protein is directed into inclusion bodies to help with this overexpression. This is done by using TrpΔLE and TCR α to form a fusion protein. TrpΔLE is an appropriate protein to aid the expression of TCR α as it directs the protein to inclusion bodies for their expression. This allows for the expression of milligram quantities of the TCR α for NMR studies which otherwise would not be possible.

TrpΔLE was also constructed with a methionine residue at the end of its sequence and was expressed as a fusion at the C-terminus of TCR α protein. This was done so that TrpΔLE could be cleaved off from the protein by using cyanogen bromide cleavage which targets the methionine residue. To remove the chance of the protein cleaving, the single methionine residue at the 24th position was replaced with serine. The plasmid was transformed into BL21 (DE3) competent cells for growth and expression of the protein. The expression of the TCR α protein attached to the TrpΔLE were done in both minimal media and LB media. The pre- and post-induction samples were then run on an SDS-PAGE gel **(Figure 2.5)**. The band on the expression gel in question is at ~18-19 kDa as TCR is 3.5 kDa and the fusion protein is 15 kDa for a total of 18.5 kDa in mass. At induction there is no prominent band at 18 kDa and post-induction we see a prominent band showing expression of the TrpΔLE/TCR α fusion protein.

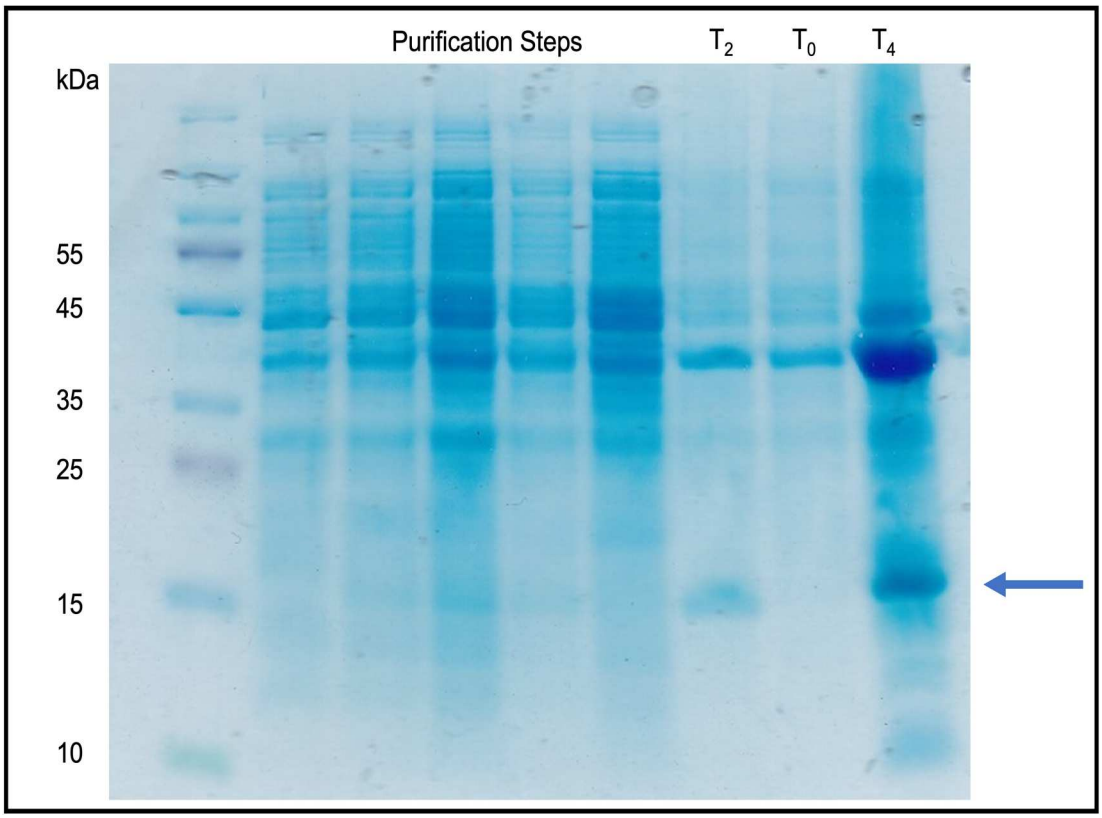

Figure 2.5- SDS-PAGE gel showing the TCR α protein expression with the fusion protein TrpΔLE. The arrow indicates the bands corresponding to the protein. T_0 is at the time of induction and T_2 and T_4 are 2 and 4 hours after induction respectively.

The purification of the protein was then started by separating soluble proteins in the cell lysate using Resuspension I Buffer containing both a Tris and glycerol component. Next, Resuspension Buffer II was used to solubilize the membrane fraction of the protein during sonication. The inclusion bodies containing the TCR α fusion protein were then pelleted by centrifugation.[70] Extraction of the protein from the inclusion bodies requires them to be solubilized in a denaturant with a refolding process following afterwards, so the protein returns to its natural folding. The inclusion bodies were dissolved in a binding buffer containing 6 M guanidine and sonication was performed to fully solubilize the protein. The sample was dialysed against DI H_2O to remove the guanidine and lead to precipitation of the protein. The protein was lyophilized to remove any liquids from the protein.

Following lyophilization, the fusion protein was dissolved in a formic acid solution and cleaved using cyanogen bromide which cleaves at the single methionine residue and separates TCR α from TrpΔLE. After cleavage, the sample was neutralized using 1 M NaOH to precipitate out the protein and exhaustive dialysis against DI H_2O was done. The sample was then once again frozen and lyophilized to obtain the protein mixture in a solid protein powder. The proteins remaining after cleavage were then purified using FPLC-Size Exclusion chromatography by running the samples in a sodium phosphate buffer with 4 mM SDS containing 0.1 M NaCl at pH 9.6. Purifying TCR α was possible due to the size difference between it and TrpΔLE and this was monitored by using absorbance at 280 nm.

The normal elution profile of the TCR α protein contained 4 prominent peaks as shown in **Figure 2.6**. These peaks were eluting at approximately 21.1 mL (p1), 24.2 mL (p2), 27 mL (p3), and 28.1 mL (p4). All four of these peaks were observed at 280 nm wavelength and 250 μL fractions were collected on either side of the maximum of each of the four peaks. The samples were then placed under vacuum and concentrated down to 100 μL. These samples were then used to run an SDS-PAGE gel (**Figure 2.7**) to analyze the peaks and determine which peak contained the TCR α protein.

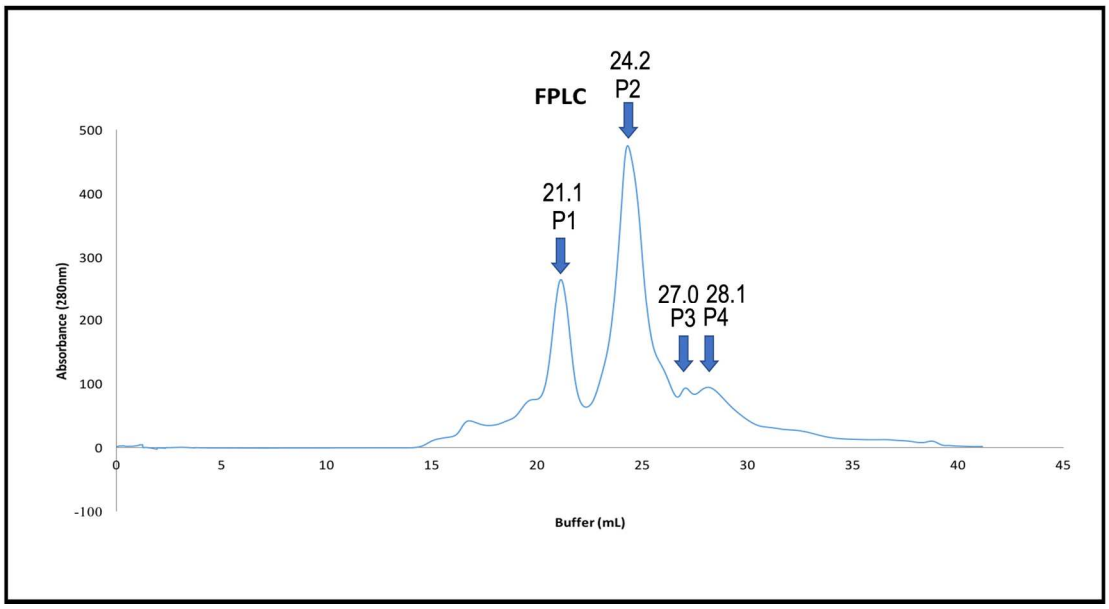

Figure 2.6- The elution profile of the cleavage products from the fusion protein cleavage. The four prominent peaks studied are p1) uncleaved protein at 21.1 mL, p2) TrpΔLE at 24.2 mL, p3) dimer of TCR α at 27.0 mL and p4) monomer of TCR α at 28.1mL according to the gel in **Figure 2.7**.

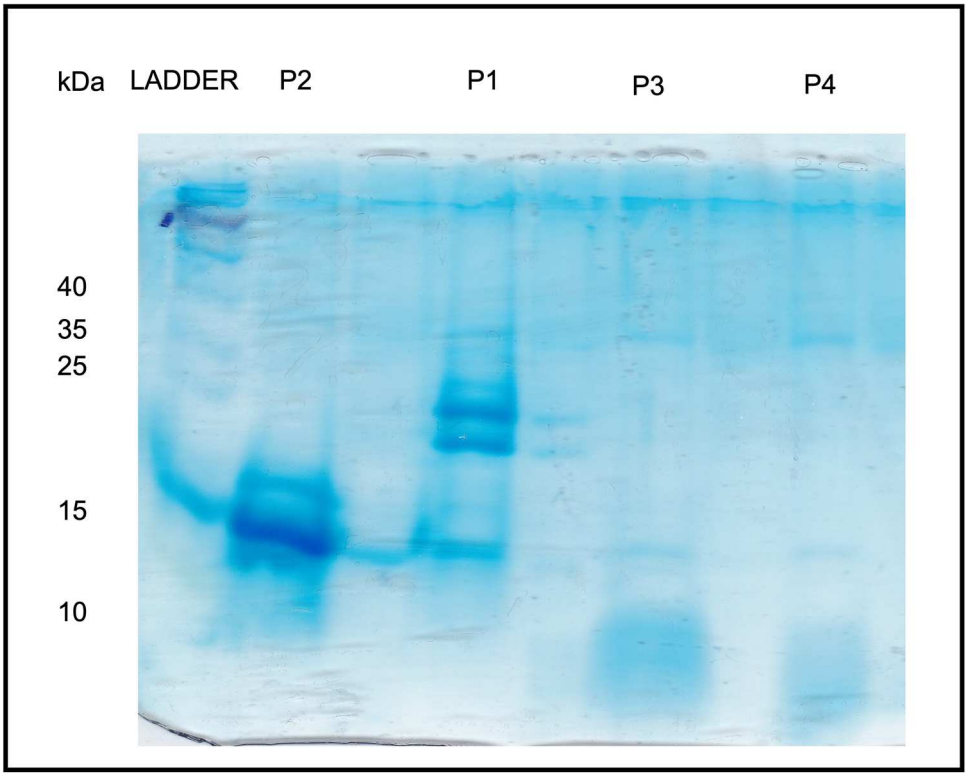

Figure 2.7- SDS-PAGE showing the approximate mass size of each peak from the FPLC-SEC of TCR α. P1 shows to peaks approximately at 21 and 18 kDa. P2 shows a main peak at 15 kDa corresponding to TrpΔLE. P3 shows a faint band at 7 kDa corresponding to the dimer of TCR α which is believed to be from disulfide linkage. P4 has a band around 3kDa which corresponds to the monomer mass of TCR α.

The fourth peak labeled, p4, which was determined to be TCR α was then collected and dialyzed against DI H_2O. The protein precipitated after SDS was reduced to a concentration below its critical micelle concentration. The protein was then frozen and dried on the lyophilizer so a protein powder could be obtained. This dry powder of TCR α was then dissolved in solution and MALDI-TOF mass spectrometry was performed on the sample as shown in **Figure 2.8**. This is done to show to confirm the purity of the identified protein. The theoretical mass of the protein was expected to be 3521.22 Da, and the major peak from the mass spectrometry sample showed to be 3510.96 Da proving that TCR α is present in the fourth peak of the FPLC chromatogram.

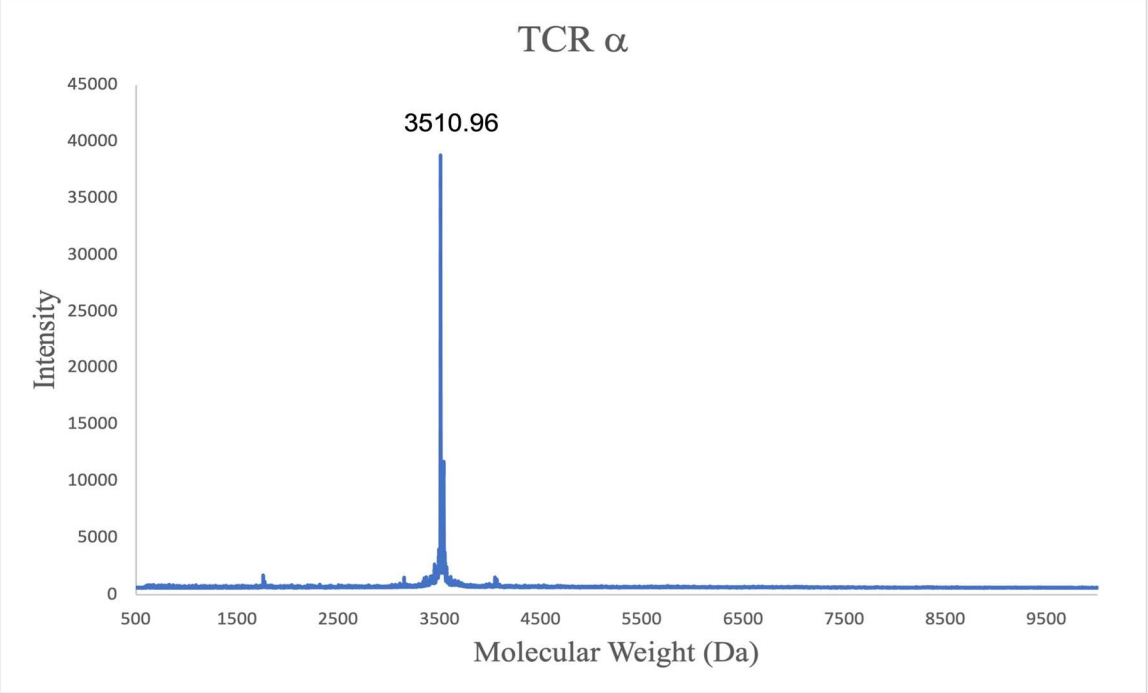

Figure 2.8- Mass spectrum of pure TCR α run in 98% ACN and 0.1%TFA. The spectrum shows a peak at 3510.96 Da corresponding to theoretical mass 3521.22 Da of TCR α.

Upon purification of unlabeled TCR α, the next step was following the same procedure, but incorporating in labeled protein. The labeled proteins have ^{15}N atoms incorporated into the protein backbone and side chains wherever a nitrogen atom is present. Once the labeled protein is purified, the next step requires the protein in a membrane mimetic that the membrane protein can be incorporated in. The membrane mimetic should be chosen to where it doesn't affect the protein

activity, but it also must solubilize the protein as membrane proteins are difficult to solubilize. Studies and literature reviews have shown that micelles are a good model as a membrane mimetic for NMR studies.[25] However, testing multiple different detergents and running NMR experiments can help determine an appropriate detergent for membrane proteins. This is important, as membrane proteins can interact differently with different detergents, but also at different pH, so analysis of different conditions can be vital in finding a suitable detergent for NMR studies.

Confirmation that the protein is not misfolded or aggregated can be done by analyzing line shapes and linewidths. By counting the resonances, it can be used as a check that you have the correct sequence as well through NMR.[71] Structural characterization of a protein can be done using solution-state NMR with a 2-D ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) experiment. In this type of experiment, magnetization is transferred between a proton and a labeled nitrogen which gives a single resonance for each backbone amide. The chemical shifts for each backbone amide is sensitive and affected easily. This can be seen in chemical shifts in spectra with different detergents. Some backbone amide resonances can be seen shifted in the HSQC spectra. Different detergents can have dramatic effects in the quality of spectral resolution. Therefore, it is important to screen multiple detergents through ^{1}H-^{15}N HSQC experiments to find a detergent that is suitable for TCR α.[72, 73]

Therefore, different detergents were screened for TCR α which included DHPC, Triton X-100, and Fos-12. When testing there was no signal from the Fos-12 detergent samples, so it was determined it was not a suitable detergent for further NMR studies. However, there was appreciable signal from the sample in the detergents DHPC and Triton X-100. The better-quality spectra came consistently from samples containing DHPC micelles, as shown in **Figure 2.9 and 2.10**. You can see better resolution and higher number of peaks in the HSQC experiments of the DHPC as compared to that of Triton X-100.

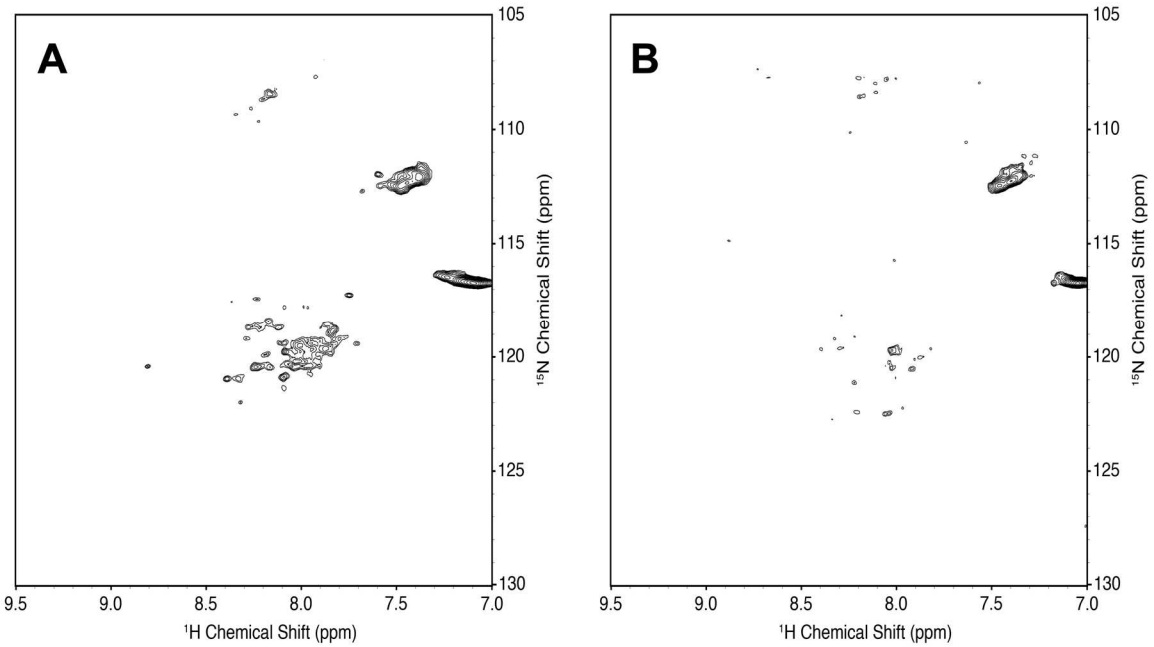

Figure 2.9- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N labeled TCR α. **A)** TCR α in 125 mM DHPC micelles. **B)** TCR α in 1% w/v Triton X-100 micelles.

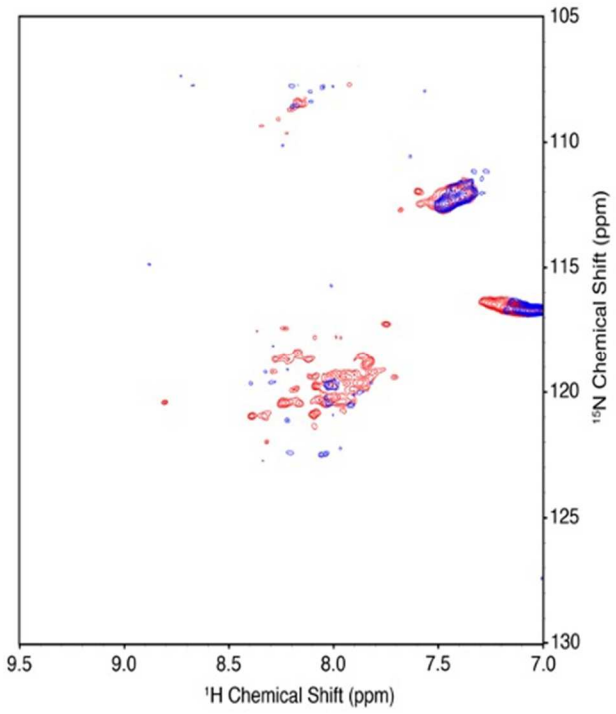

Figure 2.10- Overlay of ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N labeled TCR in 125 mM DHPC micelles (red) and 1% w/v Triton X-100 micelles (blue).

In analysis of the spectra in Figure 2.9 and 2.10, there is better signal in the DHPC micelle sample compared to the Triton X-100 sample. There is also better resolution of the peaks in these

samples. The overlay shows the significance of the detergent screening for membrane mimetics. It is clear that detergents affect the chemical shift of certain resonances of the backbone amide. In Figure 2.10 the chemical shifts are more noticeable in the nitrogen dimension certain peaks in the Triton X-100 sample are seen shifted downfield compared to the DHPC micelle sample.

2.4 Discussion

TCR α is an important part of the human auto-immune response that plays a crucial role in antigen recognition in dimer form with TCR β. Recent studies have shown the structure of different TCR portions and the full-length protein has been solved.[74, 75] A major cause of this is the difficulty of working with them due to their hydrophobic surfaces, lack of flexibility, lack of stability, and low yields in expression. In this chapter the focus was on the expression, purification, and initial structural studies of TCR α and membrane mimetic screening for NMR studies.

In this chapter, we were able to express the TCR α membrane protein, and then purify and isolate it through different purification steps. Protein identification was proven through MALDI-TOF mass spectrometry when we observed the mass of the 31 amino acid protein to be 3510.96 Da corresponding to theoretical mass 3521.22 Da of TCR α protein. Initial structural studies through NMR were then used and we found an appropriate detergent suitable (DHPC) for future NMR studies to dive deeper into structural determination. These preliminary studies are just the beginning as different membrane mimetics can be used to study this protein and help elucidate the structure of the protein.

Abbreviations: T-cell receptor (TCR); Cluster of Differentiation 3 (CD3); immunoreceptor tyrosine-based activation motifs (ITAM's); major histocompatibility complex (MHC); antigen presenting cell (APC); non-catalytic region of tyrosine kinase (Nck); nuclear magnetic resonance (NMR); electroparamagnetic resonance (EPR); Optical Density (OD); Isopropyl β-D-thiogalactopyranoside (IPTG); ethylenediaminetetraacetic acid (EDTA); Fast Protein Liquid Chromatography Size Exclusion (FPLC-SEC); *S*-(1-oxyl-2,2,3,5,5-pentamethylimidazolidin-4-ylmethyl (IMTSL): sodium dodecyl sulfate (SDS); n-Dodecyl β-D-maltoside (DDM); 1,2 diheptanoyl-sn-glycero-3-phosphocholine (DHPC); n-Dodecylphosphocholine (Fos-choline-12 or DPC); 1,2-dimyristoyl-sn-glycero-3-phosphoglycerol (DMPG); Matrix Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF); sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE); heteronuclear single quantum coherence (HSQC)

CHAPTER III

EXPRESSION, PURIFICATION, AND STRUCTURAL STUDIES OF THE MEMBRANE PROTEIN γ-SARCOGLYCAN

3.1 Introduction

3.1.1 Dystrophin-Glycoprotein Complex

The dystrophin-glycoprotein complex includes multiple transmembrane proteins and has been linked to different types of muscular dystrophy. Integral transmembrane proteins in this complex are the Sarcoglycan complex, dystroglycan complex, and sarcospan. Other important parts of this complex include dystrophin, actin, dystrobrevin, actin, syntrophins, and laminin. The rest of these proteins are all found inside the cellular membrane besides laminin as shown in **Figure 3.1**.

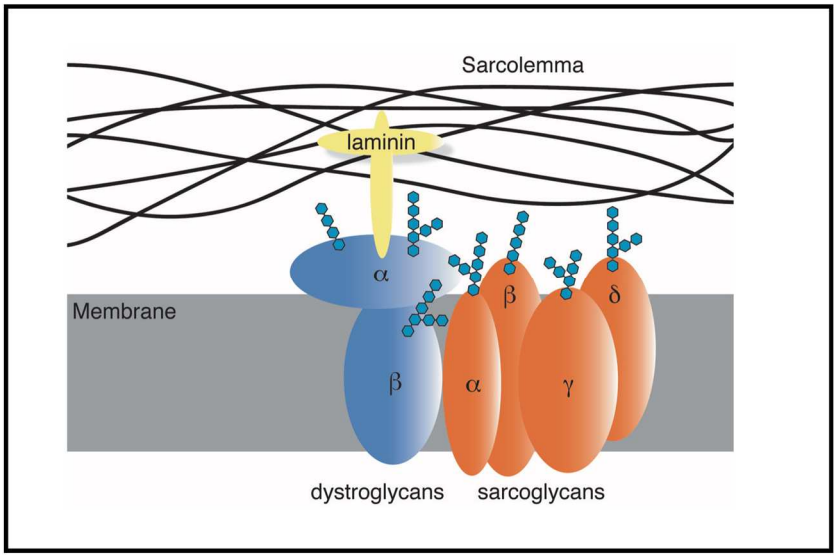

Figure 3.1- Schematic of the membrane and extracellular portions of the Sarcoglycan protein complex, dystroglycan complex, laminin, and sarcolemma. Adapted from Sweeney et. al.[76]

3.1.2 Duchenne and Limb-Girdle Muscular Dystrophies

The importance of studying this complex is because it is closely associated with muscular dystrophy. The two most common types of muscular dystrophy in humans have been linked to this complex are specifically Duchenne muscular dystrophy and limb-girdle muscular dystrophy.[77] Duchenne muscular dystrophy is a progressive muscle loss which leads to those affected to lose the ability to walk at the young age of 10 and usually pass away in their late teens or early 20's due to cardiac or respiratory complications.[78-80] This is caused due to a loss of dystrophin in the body. When this dystrophin is lost it causes a destabilization of the sarcolemma which can no longer deal with the stress of normal muscle contractions and results in muscle fiber death.[81]

Limb-girdle muscular dystrophy is different from Duchenne muscular dystrophy in which there are two types and neither lack the dystrophin protein. The first type better known as type 1 LGMD, is less severe but is usually recognized in patients due to weak proximal limbs and trunk muscles and starts in gluteal and hip adductor muscles.[82, 83] These can be caused by mutations in different proteins in the dystrophin glycoprotein complex, but are not usually life threatening to people affected by it. However, the second type or type 2 LGMD is a lot more severe and can have similar affects to that of Duchenne muscular dystrophy. Most of these are caused by specific deficiencies in the Sarcoglycan protein complex and are known as Sarcoglycanopathies. The four Sarcoglycans that are involved with this are the subunits α, β, γ, and δ. It is unknown why these dystrophies occur, but it is believed to be due to the destabilization in the sarcolemma similar to that of Duchenne muscular dystrophy. It also has patterns of progression like that of the dystrophinopathies which leads to this belief.[84]

3.1.3 Literature Review

γ-Sarcoglycan is a transmembrane protein that is part of the Sarcoglycan complex that contains an α, β, γ, and δ subunit and is shown in **Figure 3.1**. Amnio acid mutations in these proteins can lead to severe cases of limb-gridle muscular dystrophy. γ-Sarcoglycan has a short intracellular domain residing in the cytoplasm, a hydrophobic transmembrane, and a large extracellular portion. The extracellular domain has two sites of disulfide bond formation and a single N-linked glycosylation site.[85] Specific mutations of γ-Sarcoglycan have been studied and found to cause limb-girdle muscular dystrophy in humans including G69R, R116H, C283Y, and N287S.[86-88] These mutations are believed to lead to a loss of structure/function of the γ-Sarcoglycan which leads to the deterioration of muscle mass throughout the human body. Therefore, first the determination of the secondary structure of the wildtype γ-Sarcoglycan is needed to further study how the mutations affect the protein and can cause limb-girdle muscular dystrophy.

3.1.4 Summary and Outlook

γ-Sarcoglycan is a 291 amino acid (31.9 kDa) transmembrane protein, that has been proven to be essential for proper human muscle function. The protein being studied has had all the cysteines mutated to serine as it is a similar amino acid, and avoids non-specific disulfide bonds formation. The amino acid sequence is shown in **Figure 3.2**. Determining a stepwise approach for expressing and purifying the γ-Sarcoglycan is the first step in determining the secondary structure of the protein. The second step will consist of expressing the protein labeled with ^{15}N and ^{13}C for NMR studies. After expression and purification, the protein is incorporated into different detergent and lipid membrane mimetics so the protein can be studied by NMR including ^{15}N HSQC, ^{13}C HSQC, HNCA, HNCOCA experiments. Finding a membrane mimetic that is suitable to obtain good experimental NMR data is of vital importance and was determined throughout the work described in this chapter of the dissertation. This data can then be analyzed to determine the

secondary structure of the γ-Sarcoglycan protein. Future structural studies could then focus on what changes occur in the secondary structure after the protein undergoes N-linked glycosylation or how mutated variants vary. This could then help us understand how this protein is involved in the development of limb-girdle muscular dystrophy and help find a way to treat people who are struggling with this disease.

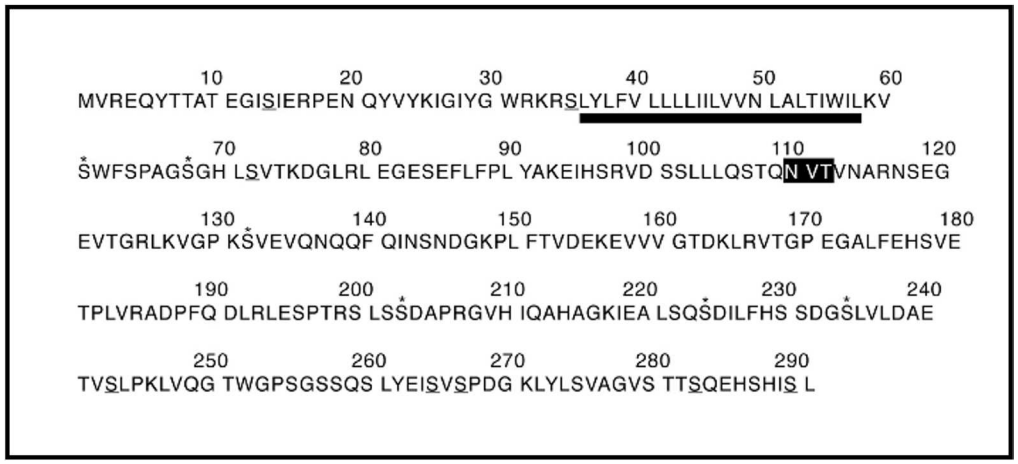

Figure 3.2- Amino acid sequence of the protein γ-Sarcoglycan. The underlined portion is the transmembrane domain of the protein while the S* demarcates methionines that were mutated to serine. The underlined S demarcates the cysteines that were mutated to serine and the portion highlighted in black is the N-linked glycosylation site.

3.2 Materials and Methods

3.2.1 Expression of Unlabeled γ-Sarcoglycan

A pUC-19 plasmid from Genewicz (www.genwicz.com) with carbenicillin resistance and the correct DNA sequence for human γ-Sarcoglycan were transformed into BL21 cells for growth. Transformation cells were plated on LB Agar media plates supplemented with Carbenicillin. The plates waere left overnight in the 37°C incubator for cell colonies to grow. A cell stock solution was then made by inoculating 10 mL of LB/carbenicillin media with a single cell colony from the plate and was left to grow by shaking in a 37°C room. The optical density at 600 nm (OD_{600}) was checked until it reached 0.4. Sterilized glycerol at 70% was then added to the remaining solution until it had a concentration of 17%. Upon addition of the glycerol, 50µL of cells were aliquoted

into 1.5 mL eppendorf tubes and were placed in the -80°C freezer for storage, and to be used for expression growths.

For expression of γ-Sarcoglycan, a starter culture was prepared by inoculating 5 mL of LB media with 5 µL of carbenicillin at a concentration of 100 mg/mL and 10 µL of the γ-Sarcoglycan cell stock. This starter culture was then placed in a 37°C room and let grow for 2 h while shaking at 220 rpm. After 2 h the sample was taken off the shaker and 200 µL were transferred to each of the two 250 mL overnight growth flasks containing 100 mL of LB media. 100 µL of carbenicillin at 100 mg/mL was also added to the overnight growth solutions. These overnight growth flasks were then placed back in the 37°C room and left shaking at 220 rpm for approximately 14 h. The next morning 50 mL of the overnight growths were then transferred back to four different 2 L growth flasks containing 950 mL of LB media so each large growth flask will have 1 L of solution media. 1 mL of carbenicillin at 100 mg/mL was then added to the solution and the large growth flasks were placed once again in the 37°C room and left shaking at 220 rpm until the OD_{600} reached 0.6, approximately 3 hours. At this point the growth was induced with 1 mL of Isopropyl β-D-thiogalactopyranoside (IPTG) (GoldBioTechnology, www.goldbio.com) at 120 mg/mL to each 1 L growth. After induction with IPTG the sample was placed back on the shaker in the 37°C room at 220 rpm for 4 h. The cells were harvested by transferring them to 1 L polycarbonate centrifuge bottles and centrifuged at 5450 xg for 25 min at 4°C (Thermo scientific, SORVALL LINX 4000 Centrifuge). The cell pellets were then saved and stored in a -80°C freezer until they were to be purified.

3.2.2 Expression of ^{15}N-labeled and ^{13}C-^{15}N labeled γ-Sarcoglycan

When expressing proteins that are in ^{15}N-labeled like γ-Sarcoglycan, a 10 µL cell stock was inoculated in 5 mL LB media with 5 µL of carbenicillin at 100 mg/mL and grown in 37°C for 2 h exactly as described in the expression of unlabeled γ-Sarcoglycan. 200 µL of the starter culture

was then added to 100 mL of sterile M9 minimal media (7.5 mM of ^{15}N ammonium sulfate, 39 mM glucose, 40 mM sodium phosphate dibasic, 20 mM sodium phosphate monobasic, 9 mM sodium chloride, 1 mM magnesium sulfate, 0.1 mM calcium chloride, and 1% LB media). 100 µL of carbenicillin was added to the 100 mL M9 minimal media solution and placed in the 37°C room and left shaking at 220 rpm for 16 h. After the overnight growth of 16 h, 50 mL was transferred back to 950 mL of the M9 solution, so it had a total volume of 1 L. Upon this addition, 1 mL of 100 mg/mL of carbenicillin was added and the four large growth flasks were placed in the 37°C room until the OD_{600} reached 0.6. This took approximately 4 h and once it reached this point the sample was induced with 1 mL of IPTG at a concentration of 120 mg/mL and was grown for 4 h while shaking at 220 rpm in a 37°C room. The cells were then harvested according to the expression of unlabeled γ-Sarcoglycan.

The expression of ^{13}C-^{15}N labeled γ-Sarcoglycan is very similar to that of single labeled ^{15}N γ-Sarcoglycan. The difference is that the glucose used includes a ^{13}C isotope, instead of the naturally abundant ^{12}C glucose. The concentration of ^{13}C glucose (Cambridge Isotope Laboratories Inc., https://www.isotope.com/) is 5.4 mM instead of the normal 39 mM glucose due to the cost associated with using isotopically labeled glucose. The ^{15}N-labeled ammonium sulfate is the exact same concentration of 7.5 mM to ensure the protein is labeled with both spin active atoms.

3.2.3 Purification

The cell pellets from the expression of γ-Sarcoglycan was then resuspended in 30 mL of Resuspension Buffer I (50 mM Tris-HCl, 1mM EDTA, 1 mM NaN_3, and 205 mM glycerol). Upon resuspension of the pellet, the sample was sonicated (Fisher Scientific Sonic Dismembrator, www.fishersci.com) on ice for a total of 4 minutes at intervals of 2 seconds on then 8 seconds off. The total elapsed time for the sample in the sonicator is 20 minutes. The samples were transferred

to a 35 mL centrifuge tube and spun down in the centrifuge at 35000 xg for 25 min at 4°C (Thermo scientific, SORVALL LINX 4000 Centrifuge). After the sample was done running, the supernatant is discarded, and cell pellet containing the membrane fraction and inclusion bodies were saved.

The cell pellet is then resuspended once more, but this time in Resuspension Buffer II (50 mM Tris, 1 mM EDTA, 1 mL NaN_3, 25.5 mM Deoxycholic acid, 17.5 mM IGEPAL). Vortexing and a spatula was used to break up the cell pellet to help resuspend the pellet. The solution was then sonicated on ice for a total of 4 minutes at intervals of 2 seconds on and 8 seconds off so that a total of 20 minutes elapsed during sonication. Samples are once again transferred to a 35 mL centrifuge tube and counterbalanced with other samples. The tubes are then spun down at 35000 xg for 25 min at 4° C. Once finished, the supernatant was discarded.

The pellet, which contained the isolated inclusion bodies, was broken up in 30 mL of Binding Buffer (6 M guanidine hydrochloride, 0.5 M sodium chloride, and 20 mM Tris pH 8.0) and stored at 25°C overnight. The partially dissolved inclusion bodies were then sonicated on ice as previously described with the two resuspension buffers. After the sonication, the samples were spun down in the centrifuge at 18500 xg for 10 min at 4°C to remove any particulates that remained. The supernatant was then saved and dialyzed against 5 L of ddH_2O in 10.0 kDa dialysis bag (Spectrum Laboratories, Inc., spectrumlabs.com). The water was changed at 30-minutes and 90-minutes, then left overnight for dialysis. After the removal of guanidine, the protein solution was frozen and dried on the lyophilizer overnight.

Next the fusion partner, Trp Leader, was cleaved off by treating the protein with cyanogen bromide (CNBr) (Alfa Aesar, www.alfa.com). When running this cleavage reaction, we used 70% formic acid (Fisher Chemical, www.fishersci.com) and cyanogen bromide. For every 10 mg of the target protein, we added 20 mg of CNBr and 400 µL of formic acid. Upon addition of the

formic acid and CNBr, the sample was placed on a rotator in the absence of light for 3 hours to push the cleavage forward. After 3 h the sample was removed from the rotator and 1 Vol. the total solution of 1 M NaOH was added and the sample was placed back on the rotator for 10 min. A white precipitate formed. 1 Vol. of the original solution of DI H_2O was then added to the sample and it was transferred to a pre-wetted 10.0 kDa dialysis bag. The reaction tube is then washed once more with 1 Vol. of DI H_2O and then transferred to the same dialysis bag.

The sample was dialyzed against 4 L of DI H_2O to neutralize the sample and to remove any remnants of CNBr. The water was switched after 30-minutes and 90-minutes then left overnight. The wastewater was neutralized above a pH of above 7 and placed in a waste container to be disposed of by waste management. The sample was then transferred to a 50 mL conical centrifuge tube and frozen in the -80°C freezer. Once the sample was frozen it was placed on the lyophilizer to freeze dry the sample. The sample was taken off the lyophilizer once dry and prepared for FPLC Size Exclusion Chromatography (SEC).

FPLC-SEC was performed by transferring 20 mg of dry γ-Sarcoglycan to a 2.0 mL eppendorf tube and the protein was dissolved by adding 2.0 mL of 10% SDS. The sample was vortexed and placed in the sonication bath for 10 minutes to help solubilize the protein. The sample was centrifuged for 5 minutes and only the supernatant was injected into the size-exclusion FPLC system on a Bio-Rad NGC Quest 10 Plus (www.bio-rad.com). The FPLC running buffer (20 mM sodium phosphate, 4 mM SDS, 1 mM EDTA, 1 mM sodium azide, and pH 8.2) was run through two columns attached in series at a flow rate of 1.5 mL/min for a total of 600 mL. The two columns used were both HiPrep™ 26/60 Sephacryl™ S-200 HR (GE Healthcare, www.gelifesciences.com). The elution of the protein was monitored using absorbance at 280 nm and the sample peaks were collected and analyzed using SDS-PAGE. The fractions containing the pure protein were then pooled together and dialyzed against 5 L of DI H_2O in a 10.0 kDa dialysis

bag. Precipitated protein was then centrifuged, frozen, and dried on the lyophilizer, so it could be used for mass spectrometry and NMR studies.

3.2.4 Mass Spectrometry

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) spectrometry was performed on a DE Pro mass spectrometer (AB SCIEX). The purified unlabeled γ-Sarcoglycan protein was prepared in two ways. In the first, the protein was dissolved in 98% acetonitrile and (Alfa Aesar, www.alfa.com), 2% trifluoroacetic acid (Acros Organics, www.acros.com) and was mixed with a matrix solution of sinapinic acid (Asta Tech, www.astatechnic.com) dissolved in 50 acetonitrile and 0.1% trifluoroacetic acid. In the second, pure protein was dissolved in 98% trifluoroethanol (Acros Organics, www.acros.com) and 2% trifluoroacetic acid solution, which was then mixed with the same matrix solution as the first procedure.

3.2.5 NMR Detergent Sample Preparation and Spectrometry

There were 6 different detergents used for making micelles for studies dealing with γ-Sarcoglycan structural studies through NMR. The detergents used were SDS, DHPC, Fos-16, DDM, DM, and N-Lauroylsarcosine (Sarkosyl). The samples were prepared for 500 µL and made each time with a 90% H_2O/10% D_2O ratio for deuterium lock of the NMR. The first detergent, SDS (Fisher BioReagents, www.fishersci.com), had a final concentration of 500 mM. This sample was prepared by dissolving 72.2 mg of SDS in 450 µL of water and adding that to 1.4 mg of γ-Sarcoglycan protein. The sample was vortexed and placed in the sonication bath for 10 minutes. The pH was adjusted to 4.0 and 50 µL of D_2O was added to get it at the final 500 µL with the correct ratio of H_2O to D_2O.

The DHPC (Avanti Polar Lipids Inc., www.avantilipids.com) samples are made by using 156.25 µL of 400 mM DHPC to dissolve approximately 1.0 mg of the ^{15}N TCR α WT or mutant

protein. The pH was then lowered to 4.5 by the addition of 4 μL of 1 M HCl to help with solubility. 293.75 μL of DI H_2O and 50 μL of D_2O (Acros Organics, www.acros.com) were added to the sample and it was placed in a sonication bath. The final concentration of DHPC in the sample was 125 mM. The sample was then centrifuged to remove particulates and the supernatant was used for NMR studies.

The DDM sample was made with a final concentration of 8 mM DDM. The sample was prepared using 50 μL of 80 mM DDM(Anatrace) to dissolve 1.7 and 3.6 mg of the protein in separate samples. 4 μL of 1 M HCl was added to lower the pH to 4 and the sample was placed in a sonication bath for 1 h to help solubilization. After solubilized, 400 μL of DI H_2O and 50 μL of D_2O were added for the final preparations of the sample. The sample was centrifuged to remove particulates and the supernatant was used for NMR studies.

The Fos-Choline 16 (Anatrace) sample was made at a final concentration of 1 mM. This was done by using a stock solution of 10 mM Fos-16 and adding 50 μL to the protein. 4 μL of 1 M HCl was added to the sample, and it was placed in a sonication bath for 1 h. After 1 h the sample was removed and 400 μL of H_2O and 50 μL of D_2O were added. The sample was then centrifuged to get remove any particulates and the supernatant was used for NMR studies.

The Sarkosyl (Sigma Chemical Co.) sample was prepared from a stock solution of 5% w/v of Sarkosyl. The final concentration of the sample was 1% Sarkosyl, so 100 μL of the 5% was added to 1.3 mg of the γ-Sarcoglycan protein. 4 μL of 1 M HCl was added to the sample, and it was placed in the sonication bath for 2 h to help solubilize the protein. After 2 h 350 μL of DI H_2O and 50 μL D_2O were added to the sample. The sample was then centrifuged to remove any particulates and supernatant was used for NMR studies.

The DM (Anatrace) sample was prepared with a final concentration of 60 mM DM solution by adding 14.5 mg of DM to 400 µL of DI H_2O for an initial concentration of 75 mM DM. The pH was then lowered to approximately 4.0 with 1 M HCl and checked using pH strips. The solution was then added to 1.3 mg of protein and placed in the sonication bath for 1 h. After an hour, 50 µL of H_2O and 50 µL of D_2O were added to the sample. The sample was then centrifuged to remove any particulates and the supernatant was used for NMR studies.

All samples were made with a detergent at a concentration above their respective critical micelle concentrations. After centrifugation, the supernatants of the samples were transferred separately to a standard 5 mm tube to run a ^{1}H-^{15}N HSQC experiment, on a Bruker DMX 500 MHz Spectrometer (www.brukerbiospin.com). A triple resonance $^{1}H/^{13}C/^{15}N$ TXI high resolution NMR probe was used. The pulse sequence, hsqcetgpsi2 (www.brukerbiospin.com), was calibrated and the water peak was referenced to 4.7 ppm. The experiments were performed at 25°C. The ^{1}H-^{15}N HSQC experiment was run with 4092 t2 points and 256 t1 points. The data was processed in nmrPipe and visualized in Sparky (T.D Goddard and D.G Kneller, SPARKY, University of California, San Francisco). Additionally, a ^{1}H-^{13}C HSQC an HNCA, HNCOCA, HNCO, and HNCACO experiments were run on a Bruker 800MHz spectrometer. A triple resonance $^{1}H/^{13}C/^{15}N$ CP-TCI high resolution NMR probe was used. The data was processed in nmrPipe and visualized in Sparky (T.D Goddard and D.G Kneller, SPARKY, University of California, San Francisco).

3.2.6 NMR Nanodisc Sample Preparation and Spectrometry

The first step in preparing purified nanodiscs for NMR was to express and purify both unlabeled Membrane Scaffold Protein (MSP) and Tobacco Etch Virus (TEV) protease. The plasmids for the expression of these proteins were received as a generous gift from the Opella Lab (University of California-San Diego). These plasmids were transformed int0 BL21 (DE3) E. Coli. using a

standard heat-shock protocol. Cell stocks for each was then used for the expression and purification of both MSP and TEV respectively.

3.2.6.1 MSP Expression and Purification

MSP was expressed as a fusion protein with a His_6-Tag attached at the N-terminus with a TEV cleavage site to eventually separate MSP from the His_6-Tag. MSP was then grown and expressed in LB media. For the expression of unlabeled MSP, a starter culture was prepared by combining 5 mL of LB media, 5 µL of the antibiotic kanamycin sulfate at a concentration of 30 mg/mL and inoculating it with 10 µL of MSP cells in a 50 mL conical centrifuge tube. The tube was placed in 37°C room and left on the shaker at 220 rpm for 2 h. After 2 h the starter culture was taken off the rotator and 200 µL of the starter was transferred to two different 100 mL overnight solutions of LB media. The LB media also contained 100 µL kanamycin at a concentration at 30 mg/mL. This was once again placed back on the shaker in the 37°C room at 220 rpm overnight for approximately 14 h. The next morning 50 mL of the 200 mL overnight growths were transferred to four different 950 mL LB media growth flasks. Upon this 1 mL of kanamycin was added to each 1 L growth at a concentration of 30 mg/mL.

The 4-1 L growths were then allowed to grow until the OD_{600}=1.0. Once the OD_{600} reached 1.0 the sample was then induced with 1 mL of IPTG at a concentration of 120 mg/mL. This approximately took 2 hours then the sample was placed back on the shaker for 3 h to express the MSP protein. At the conclusion of the 3 h, the cells were harvested by transferring them to a 1 L polycarbonate centrifuge bottles and centrifuged at 4600 xg for 10 min at 4°C (Thermo scientific, SORVALL LINX 4000 Centrifuge). The cell pellets were then saved and stored in a -80°C freezer until they were to be purified.

Purification of MSP cell pellets began by the adding 30 mL of 20 mM phosphate buffer, 150 µL of 200 mM phenylmethylsulfonyl fluoride (PMSF, Amresco, www.amresco-inc.com) and 0.300 g

Triton X-100. The cell pellet/buffer solution was sonicated (Fisher Scientific Sonic Dismembrator, www.fishersci.com) on ice for 4 min at 2 seconds on then 8 seconds off for a total elapsed time of 20 minutes. The sample was then centrifuged at 35000 xg for 25 min at 4°C. The supernatant was saved for Ni^{2+} NTA column affinity purification.

After charging the column with $NiSO_4$, 10 mL of DI H_2O was used to remove any excess nickel not attached to the column. Next, 25 mL of equilibration buffer (40 mM sodium phosphate, pH 7.4) was run through the column. The MSP sample, which was the supernatant from the cell lysis portion, was run through the column. 25 mL of Wash Buffer I (40 mM Tris-HCl, 300mM NaCl, 1% Triton X-100, pH 8.0), 25 mL of Wash Buffer II (40mM Tris-HCl, 300mM NaCl, 50 mM Na-Cholate, 20mM imidazole, pH 8.0), and 35 mL of Wash Buffer III (40 mM Tris-HCl, 300 mM NaCl, 50 mM imidazole, pH 8.0) were then run in successive order through the column. Lastly, 25 mL of the elution buffer (40 mM Tris-HCl, 300 mM NaCl, 400 mM imidazole, pH 8.0) was run through the column and the 10 mL fraction corresponding to the elution volume of 2.5-12.5 mL was collected and saved. Samples were also saved from each wash buffer step, the flowthrough, and elution fraction to run on SDS-PAGE for confirmation of pure MSP.

Dialysis of the sample was then done to remove the leftover imidazole. The 10 mL sample that was collected was transferred to a 10.0 kDa dialysis bag and dialyzed against a 1 L solution of dialysis buffer (20 mM Tris-HCl, 100 mM NaCl, and 0.5 mM EDTA) for 6 h. After 6 h the sample was then transferred to a fresh batch of dialysis buffer and left overnight to dialyze in solution. The next morning the sample was taken out and transferred to a 15 mL conical centrifuge tube and the absorbance at 280 nm was measured to determine the concentration of MSP. Dialysis buffer was used as a blank for measurement purposes and Beer's law Extinction coefficient was used for concentration calculations. Samples were stored in the -20°C freezer after adding 0.01% w/v of NaN_3.

3.2.6.2 TEV Expression and Purification

For the expression of unlabeled TEV, a starter culture was prepared by using 5 mL of LB media, 5 µL of the antibiotic carbenicillin at a concentration of 100 mg/mL and inoculating it with 10 µL of TEV expression cells in a 50 mL conical centrifuge tube. The sample was placed in 37°C room and left on shaker at 220 rpm for 2 h. After 2 h the starter culture was taken off the rotator and 200 µL of the starter was transferred to two different 100 mL overnight solutions of LB media. The LB media also contained 100 µL of carbenicillin at a concentration at 100 mg/mL. This was once again placed back on the shaker at 220 rpm 37°C room for approximately 14 h. The next morning 50 mL of the 200 mL overnight growths were transferred to four different 950 mL LB media growth flasks. 1 mL of carbenicillin was added to each 1 L growth at a concentration of 100 mg/mL.

The 4-1 L cultures were grown until the OD_{600}=0.6. Once the OD_{600} reached 0.6 the sample was induced with 1 mL of IPTG at a concentration of 120 mg/mL. This typically takes approximately 1.5 hours. The sample was placed back on the shaker for 4 h to express the TEV protein. At the conclusion of the 4 h shaking time, the cells were harvested by transferring them to a 1 L polycarbonate centrifuge bottles and centrifuged at 5450 xg for 25 min at 4°C (Thermo scientific, SORVALL LINX 4000 Centrifuge). The cell pellets were then saved and stored in a -80°C freezer until they were to be purified.

The TEV cell pellets were then purified by cell lysis and Ni^{2+} NTA column affinity purification. For lysis, 36 mL of Resuspension Buffer (40 mM Tris-HCl, 500 mM NaCl), 75 mg Lysosome egg white, 5 mL Lysis Buffer (8 M Urea, 20 mM Tris-HCl), and 3.6 g solid Urea were added to the cell pellet in a 50 mL conical centrifuge tube. The sample was sonicated (Fisher Scientific Sonic Dismembrator, www.fishersci.com) on ice for 4 min at 2 seconds on then 8 seconds off for

a total elapsed time of 20 minutes. Afterwards, the sample was centrifuged for 20 minutes at 4°C at 30000 xg. The supernatant was saved for a nickel column purification process.

Prior to loading the protein onto the nickel column, 10 mL of DI H_2O is added to remove any excess nickel in the column that is unbound. 20 mL of resuspension buffer (40 mM Tris-HCl, 500mM NaCl) with 40 mM imidazole was then run through the column for equilibration. Following the resuspension buffer, 35 ml of the supernatant of the TEV lysis with 50 mM of imidazole was run through the column. Next, 25 mL of the wash buffer (20 mM Tris-HCl, 500 mM NaCl, 80 mM imidazole) was run through the column. Lastly, 25 mL of the elution buffer (20 mM Tris-HCl, 500 mM NaCl, and 500 mM imidazole) was run through the column and milliliters of the 2.5-10.0 mL elution were collected for a total of 7.5 mL. Samples from the flowthrough, wash buffer, and elution buffer were collected and saved to run on an SDS-PAGE gel to verify the purification of TEV.

3.2.6.3 Cleavage of MSP by TEV

The 10 mL sample of MSP was used for cleavage and 2.6 mg of dithiothreitol (DTT) was added to MSP to prevent the formation of disulfide bonds. MSP was then mixed with 7.5 mL of pure TEV in a 50 mL conical centrifuge tube. The sample was then transferred to a 10 kDa dialysis bag, placed in 5 L of cleavage buffer (50 mM Tris-HCl, 300 mM NaCl, 50 mM imidazole) and placed in the 4°C room for 16 h for the cleavage to proceed. After 16 h, the sample was transferred to a 15 mL conical centrifuge tube and run through a nickel column to separate the cleaved MSP from uncleaved MSP and TEV.

When running the nickel column to separate the MSP from TEV, the column was first rinsed with 10 mL of DI H_2O. The column was then equilibrated with 20 mL of Wash Buffer I (40 mM Tris-HCl, 300 mM NaCl, 50 mM imidazole). Following equilibration, the 17.5 mL sample was run through, and the flowthrough is collected for dialysis. The column was washed with 10 mL of

elution buffer (40 mM Tris-HCl, 300 mM NaCl, 400 mM imidazole) to remove any bound protein. The column was stripped with 50 mM ethylenediaminetetraacetic acid (EDTA) and regenerated.

The flowthrough that was collected is then dialyzed in 4 L of End Buffer (20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA). This was dialyzed for 6 h then it is switched to a fresh volume of end buffer. The next morning the sample was taken off dialysis and transferred to a 50 mL conical centrifuge tube. The concentration of the sample was then calculated using Absorbance measurements at 280 nm and Beer's law Extinction coefficient for concentration determination. Samples were stored in the -20°C freezer after adding 0.01% w/v of NaN_3.

3.2.6.4 NMR Sample Preparation

Sample preparation of the nanodisc sample were prepared by dissolving 32mg of DMPC in 1 mL of chloroform, which was dried under N_2 gas. The dried film of DMPC was placed in the desiccator overnight to remove any residual chloroform. The next day the DMPC lipid film was hydrated with 500 µL of 58.1 mM Na-Cholate solution. This was done by vortexing and placing the sample in a sonication bath. While that was hydrating, 2.3 mg of ^{15}N labeled γ-Sarcoglycan was dissolved in 250 µL of 80 mM DDM, and 8 µL of 1 M NaOH was added to help solubilize the protein. Once both samples were in solution, the samples were combined and placed on a rotator for 2 h. After 2 h, 1 mL of MSP with a concentration of 18 mg/mL was added to the sample for a total sample volume of 1.75 mL.

The sample was then dialyzed in a 10.0 kDa dialysis bag against 5 L of 20 mM phosphate buffer at a pH of 7.2 with 5 mM β-cyclodextrin to remove Na-Cholate. The buffer was changed every 6 h and following buffer concentrations remained the same for phosphate, but the β-cyclodextrin concentration was lowered to 1 mM. The buffer was changed 5 different times. After the five different buffer changes following dialysis, the sample was concentrated down to 450 µL in a 10

kDa regenerated cellulose cut-off centrifugal filter (Merck Millipore Ltd., www.millipore.com). Once the sample was concentrated down to 450 µL, 50 µL of D_2O was added and the sample pH was adjusted to 6.5. The sample was transferred to a standard 5 mm tube and a ^{1}H-^{15}N HSQC experiment was run on the sample.

These NMR experiments were performed on a Bruker DMX 500 MHz Spectrometer (www.brukerbiospin.com). A triple resonance $^{1}H/^{13}C/^{15}N$ TXI high resolution NMR probe was used. The pulse sequence, hsqcetgpsi2 (www.brukerbiospin.com), was calibrated and the water peak was referenced to 4.7 ppm. The experiments were performed at 25°C. The ^{1}H-^{15}N HSQC experiment was run with 4092 t2 points and 256 t1 points. The data was processed in nmrPipe and visualized in Sparky (T.D Goddard and D.G Kneller, SPARKY, University of California, San Francisco).

3.2.7 NMR Bicelle and D_2O Sample Studies Preparation and Spectrometry

For NMR bicelle titration studies a 1 mL sample was made containing 3.8 mg of γ-Sarcoglycan. The protein was dissolved in 312.5 µL of a 400 mM DHPC solution. 587.5 µL of DI H_2O was added and 8 µL of 1 M HCl was added to help solubilize the sample. The sample was then placed in the sonication bath for 1 h for solubilization purposes. Upon solubilization 100 µL D_2O was added to the sample. The sample contained 90% H_2O/10% D_2O at a final concentration of 125 mM DHPC. 8.5 mg of DMPC was then added to the sample and was dissolved by freezing the sample in liquid nitrogen and thawing the sample in a heat bath at 40°C. This was repeated 30 times to dissolve the DMPC. This yielded a sample with q=0.1 bicelles (DMPC/DHPC). When looking at bicelles, the q value is the ration of long-chain lipid compared to short-chain lipid. The sample was then centrifuged to remove any particulates and 500 µL of the supernatant was transferred to a standard 5 mm tube to perform NMR experiments.

Once the first ^{1}H-^{15}N HSQC was measured, the sample was taken out of the NMR and transferred back to the original microcentrifuge tube containing the original solution. DMPC was then once again added until the bicelle had a ratio of q=0.2. The sample was solubilized the same way as the previous sample using liquid nitrogen and a hot water bath. The running of these experiments was then repeated at different bicelle q values of 0.2, 0.5, 1.0, 1.5, 2.0, and 2.5. Each sample was run at 500 µL at the specific q value.

The next set of experiments were run on samples with different percentages of H_2O: D_2O. The samples were prepared using DHPC as the detergent for solubilization and NMR studies. The final concentration of the first sample was 125 mM and it was prepared by adding 156.25 µL of 400 mM DHPC, 293.75 µL, and 4 µL of 1 M HCl. This sample was placed in the sonication bath to solubilize the protein. Once solubilized, 50 µL D_2O was added to the sample and it was centrifuged to remove any particulates. The sample was transferred to a standard 5 mm tube and the ^{1}H-^{15}N HSQC experiment was run on the sample. The sample was then taken out of the tube and transferred to a microcentrifuge tube, frozen, and placed on the lyophilizer to dry. The sample took 4 hours to dry and then the next sample was made by increase the D_2O to 20% and decreasing the H_2O to 80%. This process was repeated, and new samples were tested each time consisting of D_2O at 10-100% in 10% increments.

These NMR experiments were performed on a Bruker DMX 500 MHz Spectrometer (www.brukerbiospin.com). A triple resonance $^{1}H/^{13}C/^{15}N$ TXI high resolution NMR probe was used. The pulse sequence, hsqcetgpsi2 (www.brukerbiospin.com), was calibrated and the water peak was referenced to 4.7 ppm. The experiments were performed at 25°C. The ^{1}H-^{15}N HSQC experiment was run with 4092 t2 points and 256 t1 points. The data was processed in nmrPipe and visualized in Sparky (T.D Goddard and D.G Kneller, SPARKY, University of California, San Francisco).

3.2.8 Dual-Labeled NMR Experiments and Sample Preparation

For the dual-labeled experiments including ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, HNCA, HN(CO)CA, HNCO, and HN(CA)CO, the sample was prepared using 3.7 mg of γ-Sarcoglycan that was dissolved in DDM at a final concentration of 8 mM. The sample consisted of 90% H_2O and 10% D_2O and was run on the NEO800 NMR magnet. The sample was calibrated to the water peak and water suppression was performed to reduce the solvent signal. The data was processed in nmrPipe and visualized in Sparky (T.D Goddard and D.G Kneller, SPARKY, University of California, San Francisco)

3.3 Results

For structural studies of γ-Sarcoglycan, milligram amounts of the protein are needed for NMR experiments. Therefore, the full-length protein was recombinantly expressed in *Escherichia coli* which is the most common used bacterial host for recombinantly expressed membrane proteins. However, as stated earlier, high expression of proteins in cells can be lethal to the bacteria, so these proteins need to be directed to inclusion bodies to make high-level expression of proteins possible. γ-Sarcoglycan was expressed as a fusion protein with TrpΔLE helping to direct the protein into inclusion bodies for suitable expression at higher quantities. This would allow for the milligram quantities needed for different NMR studies. The TrpΔLE fusion was added at the N-terminus of γ-Sarcoglycan, and it was engineered with a methionine for the purpose of cleaving off the TrpΔLE protein post-expression.

The plasmid for the expression of the protein was transformed into BL21 competent cells. The expression of the γ-Sarcoglycan fusion protein was done in both minimal media and media. The post-induction samples were then run on an SDS-PAGE gel**(Figure 3.3)**. The band on the expression gel at ~47 kDa corresponds to the fusion protein as γ-Sarcoglycan is 31.9 kDa and

TrpΔLE is 15 kDa for a total of 46.9 kDa in mass. The gel below shows the increase in protein expression from initial induction to 4 hours after induction.

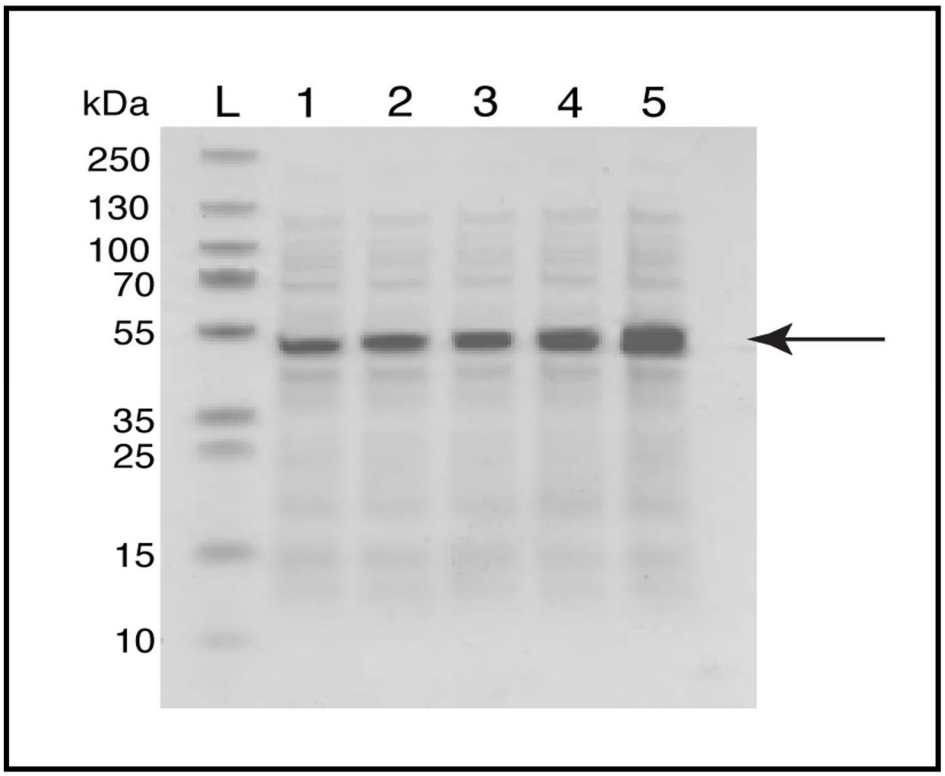

Figure 3.3- SDS-PAGE gel showing the increase of γ-Sarcoglycan protein expression with the fusion protein TrpΔLE. The arrow indicates the bands in question corresponding to the protein's size at 47 kDa. Each lane corresponds to the hourly time after induction of the protein by IPTG.

Once expressed, γ-Sarcoglycan needed to be purified. The first step after expression is to remove soluble proteins in the cell lysate by separating them using a resuspension buffer containing Tris and glycerol. Following this, a second resuspension buffer was used to solubilize the membrane fraction of the protein during sonication. The inclusion bodies containing the γ-Sarcoglycan fusion protein were then pelleted by centrifugation.[70] Extraction of the protein from the inclusion bodies requires them to be solubilized in a denaturant with a refolding process following afterwards, so the protein returns to its natural folded state. To achieve this, the inclusion bodies were dissolved in a binding buffer which included 6 M guanidine and sonication was performed to fully solubilize the sample. The sample then underwent refolding by placing the

sample in dialysis against DI H_2O to remove the guanidine and lead to precipitation of the protein. The protein then underwent a freeze dry process known as lyophilization to remove any moisture from the protein.

The fusion protein was then solubilized using formic acid and cyanogen bromide was added to cleave the fusion protein into separate TrpΔLE and γ-Sarcolgycan proteins. Upon cleavage, the proteins were precipitated, using 1 M NaOH then dialyzed against DI H_2O to neutralize the solution and remove any cyanogen bromide remaining in the cleavage mixture. The sample was then frozen and dried on the lyophilizer to obtain a dried powder. The remaining proteins after cleavage were then purified using FPLC-Size Exclusion chromatography by running the samples in a sodium phosphate buffer with 4 mM SDS at a pH 9.6. Purifying γ-Sarcolgycan was possible due to the size difference between it and TrpΔLE. This was monitored by using absorbance at 280 nm.

The normal elution profile of the γ-Sarcolgycan protein contained 3 prominent peaks as shown in **Figure 3.4**. These three peaks elute at 200 mL, 220 mL, and 250 mL. All three of these peaks were observed at 280 nm wavelength and 15 mL fractions were collected from each of the three peaks. 100 μL samples from each of these 3 peak fractions and the tail end of the first and second peak were taken as samples. These 5 samples were then used to run an SDS-PAGE gel (**Figure 3.5**) to analyze the peaks and determine which peak contained γ-Sarcolgycan and determine its purity.

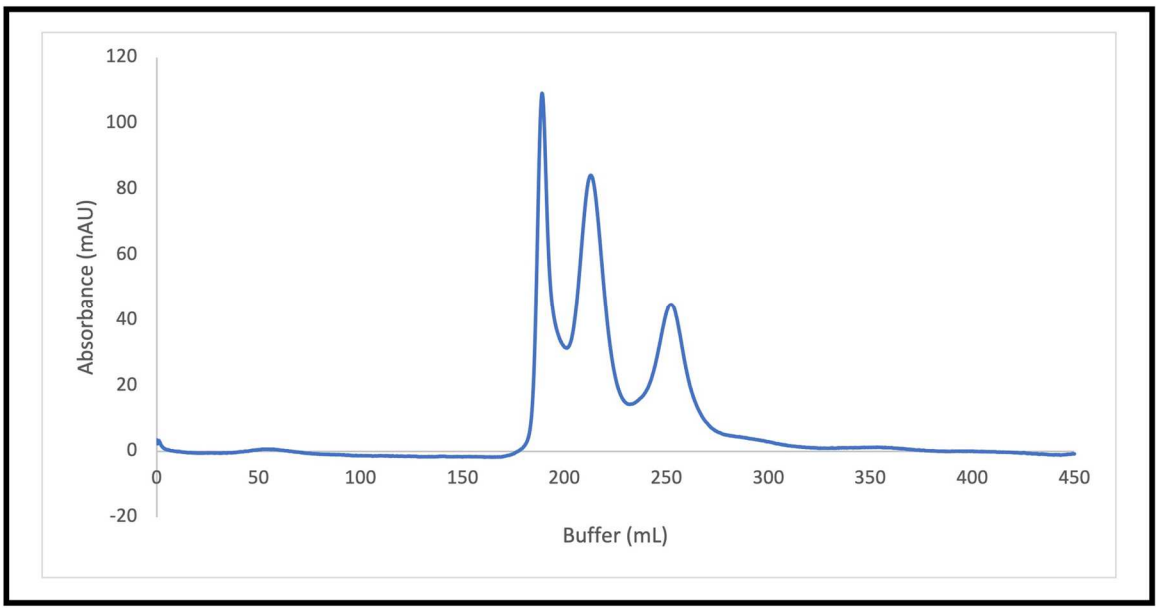

Figure 3.4- FPLC elution profile of the purification of γ-Sarcoglycan. The first peak at 200 mL corresponds to uncleaved fusion protein while the second peak at 220 mL and third peak at 250 mL correspond to γ-Sarcoglycan and TrpΔLE respectively. This is proven by the gel in Figure 3.5.

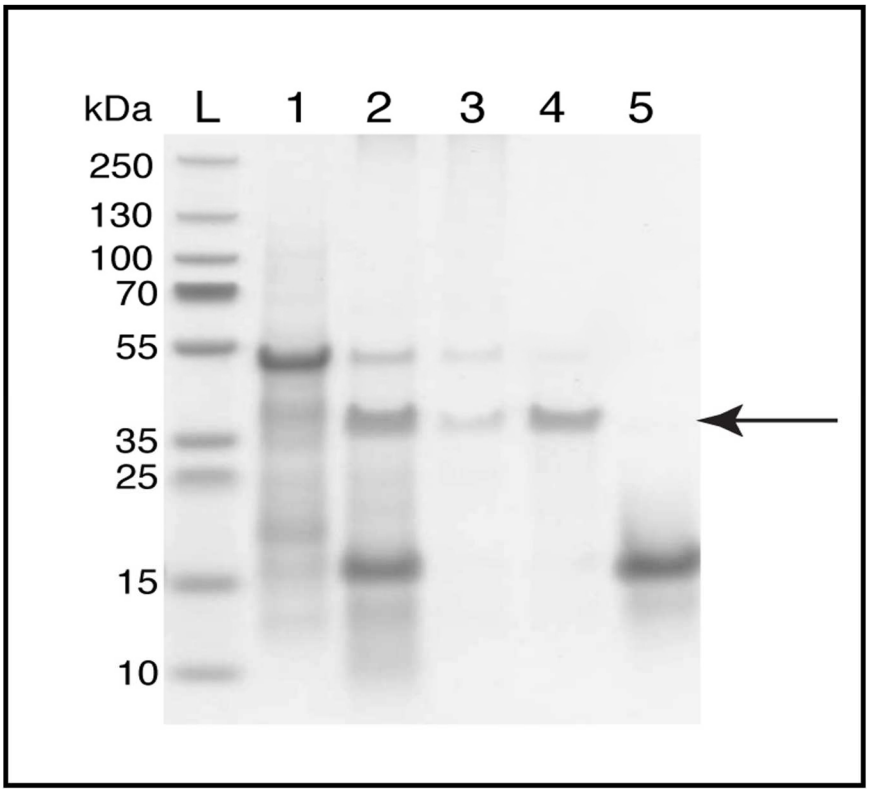

Figure 3.5- The gel above contains a ladder and four samples from the FPLC elution profile from Figure 3.4, while Lane 2 is cleaved material not run through FPLC. The lane of interest is 4 as it is pure Sarcoglycan without any other bands showing no impurities.

After obtaining a gel showing only a single band for γ-Sarcoglycan in lane 4 and determining its elution fraction is at 220 mL, the mass of the protein needed to be confirmed using MALDI-TOF mass spectrometry. Therefore, further FPLC runs were done, and the 220 mL elution samples were then dialyzed. The precipitated protein was then frozen and dried on the lyophilizer to obtain a pure protein powder. This dry powder of γ-Sarcoglycan was dissolved in organic solvents and MALDI-TOF mass spectrometry was run on the sample as shown in **Figure 3.6**. This is done to confirm the identity of the protein. The theoretical mass of the protein was expected to be 31959.16 Da, and the major peak from mass spectrometry sample was 31970.15 Da proving that the protein isolated by FPLC was γ-Sarcoglycan.

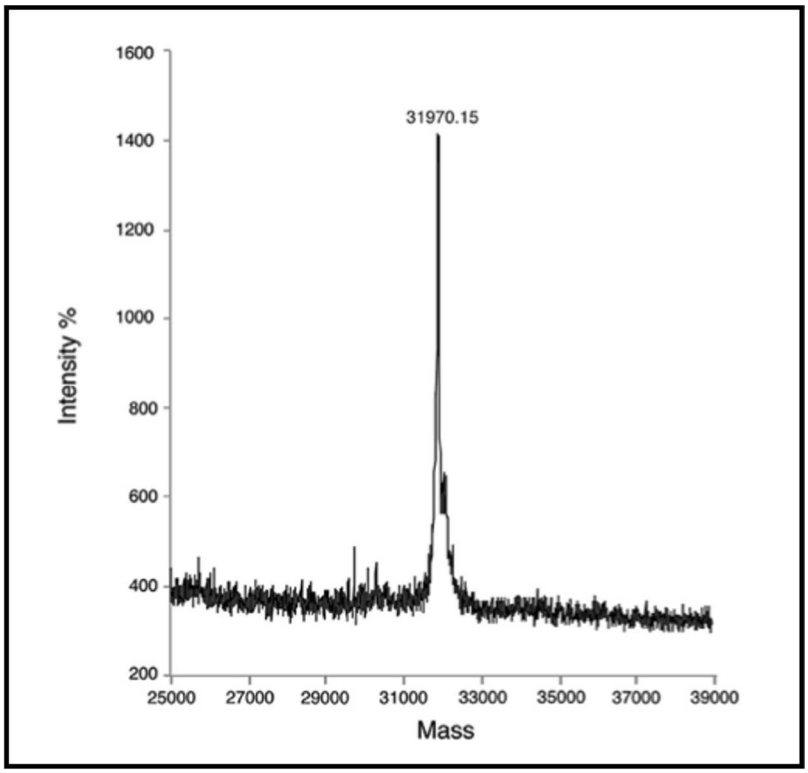

Figure 3.6- Mass spectrum of pure γ-Sarcoglycan in 98% ACN and 0.1%TFA. The spectrum shows a peak at 31970.15 Da corresponding to theoretical mass 31959.16 Da of γ-Sarcoglycan.

Before running NMR experiments on this protein, two more pure proteins were needed. These two proteins were Membrane Scaffold Protein (MSP) and Tobacco Etch Virus (TEV). They are

needed for making nanodisc samples of the protein. MSP and TEV were both expressed in E. Coli in BL21 competent cells and purification was confirmed by SDS-PAGE.

MSP was expressed as a fusion protein with a His_6-Tag attached at the N-terminus with a TEV cleavage site to separate MSP from the His_6-Tag. MSP was then grown and expressed in LB media. The protein was mechanically lysed by sonication in phosphate buffer containing phenylmethylsulfonyl fluoride (PMSF), a protease inhibitor, and Triton X-100. After sonicating and centrifuging, the supernatant was run through a Ni^{2+} column with different wash and elution buffers. It was determined that MSP eluted in a 10 mL fraction from 2.5-12.5 mL in the elution buffer. This sample was then dialyzed against phosphate buffer to remove imidazole and other impurities. The pure protein is 22 kDa in size and can be seen in the gel in **Figure 3.7** at that position in the gel**.** Upon purification, the concentration of the 10 mL fraction was determined using Beer's law equation $A=\varepsilon lc$, where $\varepsilon=21{,}430\ M^{-1}\ cm^{-1}$ $l=1$ cm, to calculate the concentration and determine the exact amount of MSP expressed and purified. This amount of MSP purified was consistently in the range of 20-40 mg per 10 mL of solution.

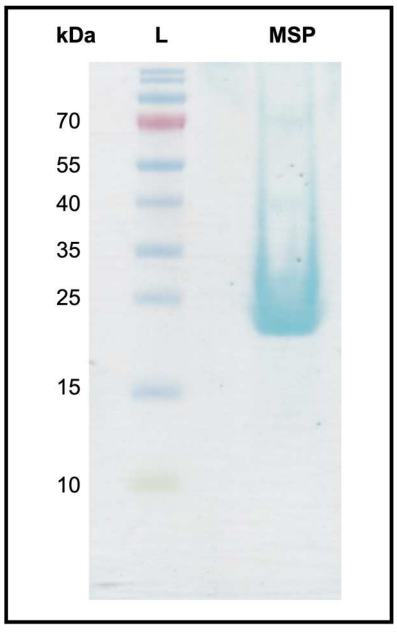

Figure 3.7- SDS-PAGE gel showing a sample band at ~22 kDa in size which corresponds to the size of pure MSP.

TEV was expressed in LB media and purified using Ni^{2+} NTA column to separate it from other impurities. It was mechanically lysed by sonication using a Tris/Urea buffer with Lysozome from egg white. After sonication, the sample was centrifuged, and the supernatant was saved for Ni^{2+} column purification. The different wash and elution buffers were run through the column, and it was determined that the TEV protease was in a 7.5 mL fraction from the 2.5-10 mL portion of the elution buffer. This was proved through SDS-PAGE **(Figure 3.8)** in which a clear band is present at 27 kDa which corresponds to the size of TEV. This pure TEV was then used for the cleavage of the His_6-Tag from the MSP fusion protein.

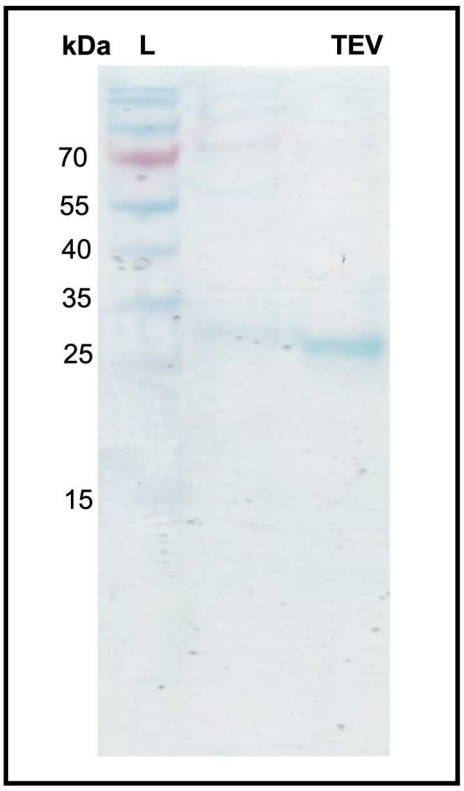

Figure 3.8- SDS-PAGE gel showing a sample band at ~27 kDa in size which corresponds to the size of pure TEV.

Upon purification of the unlabeled γ-Sarcoglycan protein, the protein was then expressed labeled with ^{15}N and dual labeled with ^{13}C-^{15}N label. The labeled protein was purified using the same techniques as unlabeled protein. The next step is characterization and structural determination through NMR, and the first thing to do is find a suitable membrane mimetic that gives good

signal with little noise and great resolution. As described in chapter 2 the membrane mimetic chosen should not affect protein activity, but also solubilize the highly hydrophobic membrane protein. Therefore, a multitude of membrane mimetics were tested to see what would give good resolution for further NMR studies.

The first experiment done to compare these different conditions was a ^{1}H-^{15}N HSQC. The initial testing of membrane mimetics started with micelles. The 6 different detergents used for these micelles were SDS, DHPC, Fos-16, DDM, DM, and Sarkosyl. These samples were run above their critical micelle concentration and the HSQC experiments were analyzed and processed to determine which detergent was optimal for γ-Sarcoglycan as shown in **Figure 3.9**.

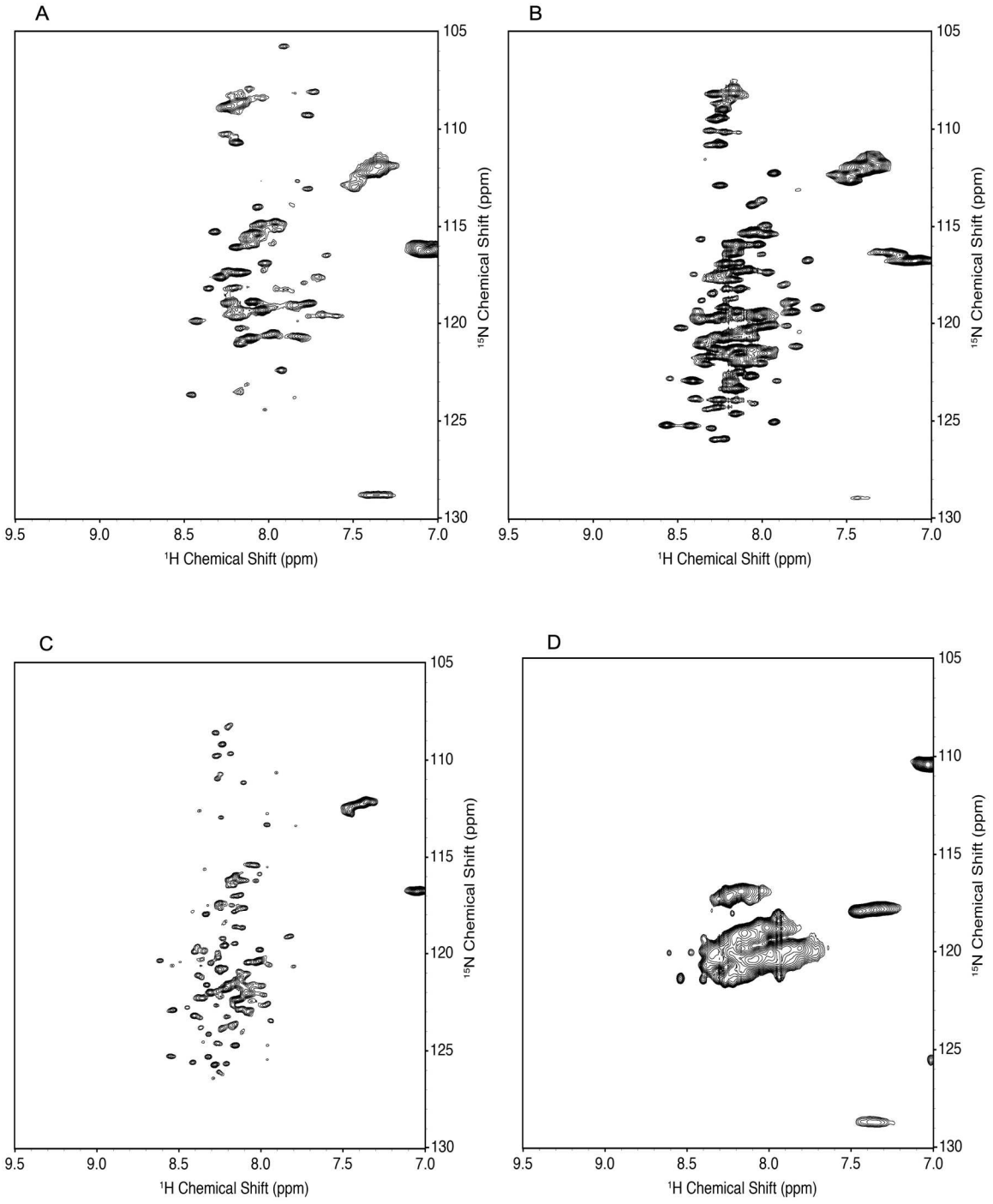

Figure 3.9- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N γ-Sarcoglycan. **A)** γ-Sarcoglycan in 500 mM SDS micelles, **B)** γ-Sarcoglycan in 125 mM DHPC micelles, **C)** γ-Sarcoglycan in 8 mM DDM micelles, and **D)** γ-Sarcoglycan in 60 mM DM micelles. Not pictured here are NMR experiments dealing with Sarkosyl and Fos-16 micelles as they did not have any relevant signal for the γ-Sarcoglycan protein.

Analysis of the spectra in different detergents show noticeable differences and why detergent screening is necessary. The first detergent tested was SDS and it does show some signal and

distinguishable peaks but is clearly not a good detergent for γ-Sarcoglycan when comparing it to DHPC and DDM as shown in an overlay in **Figure 3.10** of DDM and SDS. Closer analysis of the DHPC and DDM samples show that DDM has better resolution in the proton chemical shift range from 8.0-8.4 ppm. DDM clearly gives better resolved peaks and even has better signal as there are peaks seen that are not present in the other spectra. Therefore, for future NMR experiments with micelles, DDM was continually used for ^{15}N and ^{13}C-^{15}N labeled γ-Sarcoglycan samples. The other tested detergent spectra are not shown as they had no resolved signal.

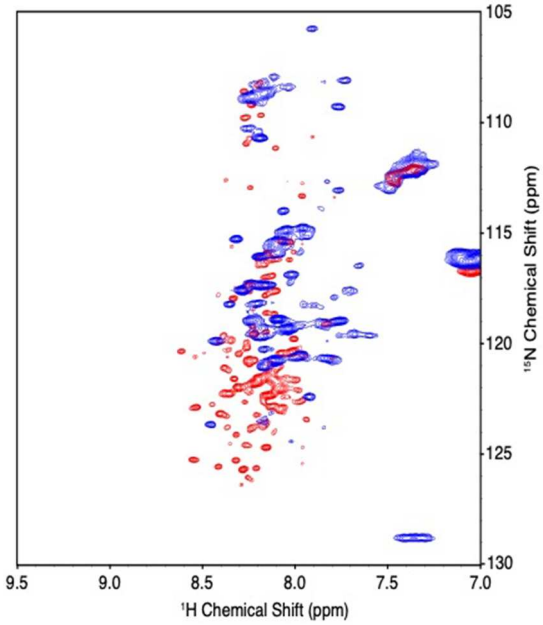

Figure 3.10- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N γ-Sarcoglycan in SDS (blue) and DDM micelles (red).

The next membrane mimetic tested with the γ-Sarcoglycan protein was nanodiscs from the membrane scaffold protein (MSP) **(Figure 3.11)**. Comparing the spectra of each and overlaying them show that there is still better resolution and more peaks in the DDM micelle sample as shown in **Figure 3.12.** The protein signal in the nanodisc samples was worse as the protein/nanodisc structure was on the edge of being too big for solution NMR. As a result, the tumbling is slowed down to a rate where the loss of some peaks is seen in the nanodiscs sample

compared to the DDM micelle sample. Further studies, going forward will then use DDM micelles when necessary and appropriate.

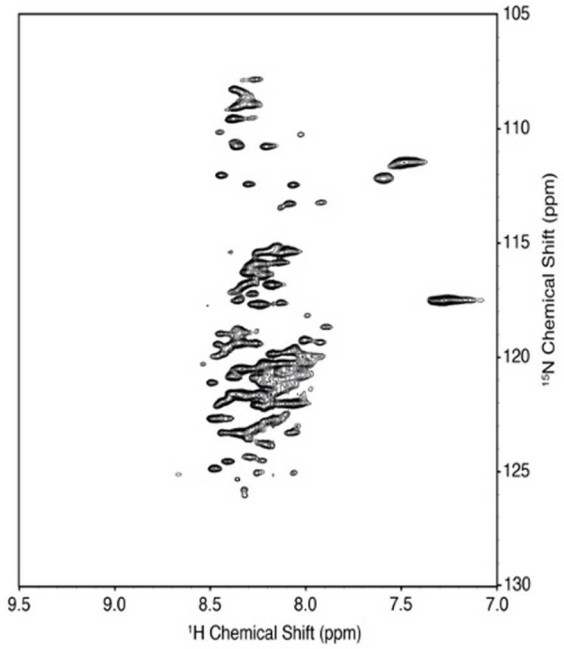

Figure 3.11- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N γ-Sarcoglycan incorporated into MSP nanodiscs.

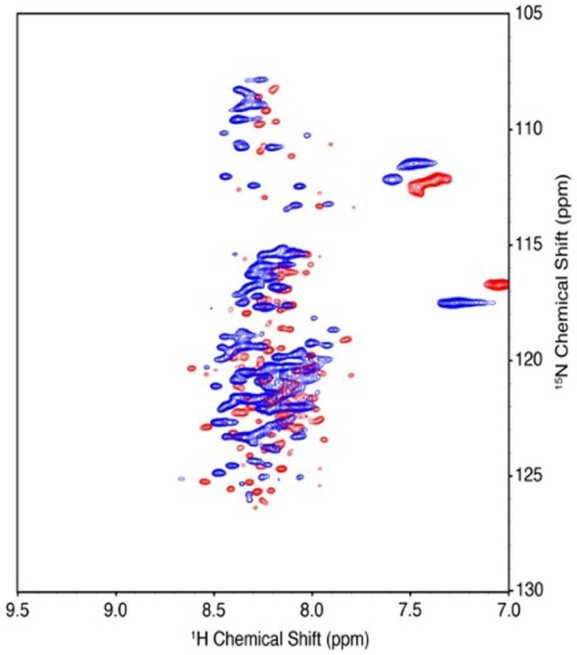

Figure 3.12- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N γ-Sarcoglycan incorporated into MSP nanodiscs (blue) and DDM micelles (red).

Since each resonance peak corresponds to a specific amino acid in the backbone, the next two experiments that were run deal with determining which resonances are in secondary structure and which are outside of the cell membrane. For determining which resonances are in secondary structure, D_2O titration experiment was performed. For determination of which are inside the lipid environment a bicelle titration experiment was performed.

The D_2O experiments were ^{1}H-^{15}N HSQC acquired on samples with increasing D_2O and decreasing H_2O. We started at 10% D_2O and increased it all the way up until the sample was 100% D_2O in 10% increments. The increase in D_2O leads to a loss of signal and the disappearance of resonance peaks as shown in the spectra of the D_2O titration experiments **(Figure 3.13).** The peaks still present in the later experiments are part of an ordered secondary structure like an alpha-helix or beta sheet as the backbone hydrogen in structures do not exchange with the deuterium as quickly as disordered regions. This is shown more clearly when looking at **Figure 3.14** where you can see the loss in certain resonances from the 10% to 50% then no distinguishable signal in the 80% D_2O samples. You can see the disappearance of peaks from the 10% D_2O sample to the 50% D_2O sample very clearly. The peaks that are still visible are believed to be in a secondary structure like an alpha-helix or beta sheet as stated earlier. This experiment will be valuable to use when looking at three-dimensional experiment for amino acid assignment of the different peaks.

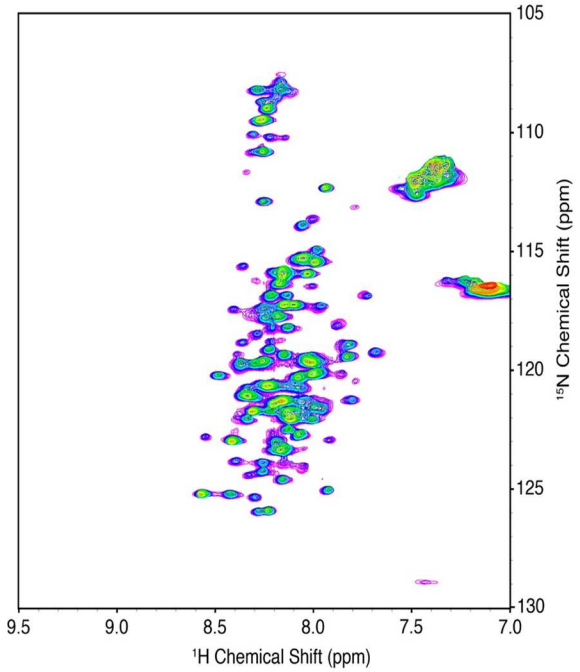

Figure 3.13- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N γ-Sarcoglycan incorporated into DDM micelles for D_2O titration experiments. The colors of the overlayed spectra correspond to the D_2O titration experiments as follows: 10% magenta, 20% purple, 30% blue, 40% cyan, 50% green, 60% bright green, 70% yellow, 80% orange, and 90% red.

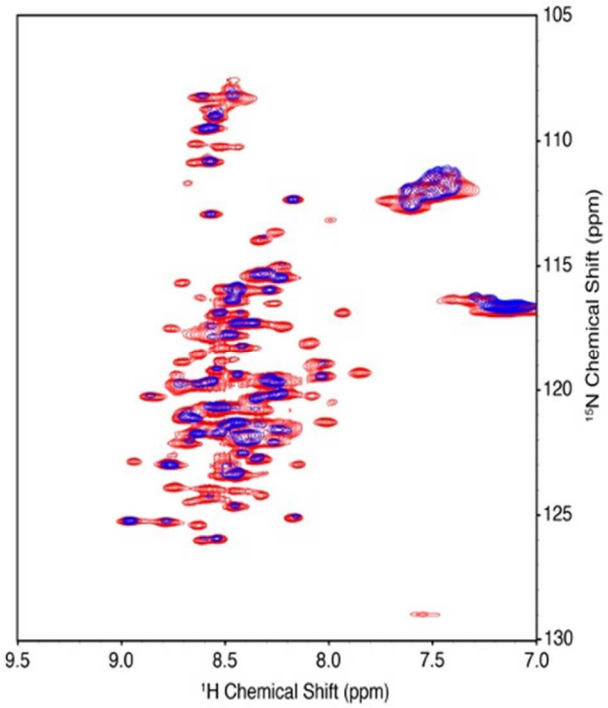

Figure 3.14- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N γ-Sarcoglycan incorporated into DDM micelles for D_2O titration experiments. The colors of the overlayed spectra correspond to the D_2O titration experiments as follows: 10% red, 50% blue, and 80% cyan.

The bicelle titration experiments started with a sample of γ-Sarcoglycan in just DHPC, a short-chain phospholipid, and adding DMPC, a long-chain phospholipid, at increasing concentration ratios to DHPC. The samples were run at a q, or ratio of long-chain phospholipid to short-chain phospholipid, of 0.1, 0.2, 0.5, 1.0. 1.5, 2.0 and 2.5. In this experiment, the larger bicelle slows down the tumbling motion of the protein inside the bicelle and leads to a loss of signal. This then can help distinguish certain resonances that are on the outside the bicelle as their signals still appear in the spectra while those inside the bicelle disappear. This is shown in **Figure 3.15 & 3.16** as we can see certain peaks that have disappeared in the overlayed spectra. The first figure showing all experiments ran and the second figure highlighting the changes between q=1.0 and q=2.5 where you can clearly see the effect the largerbicelle has on the tumbling motion of the protein.

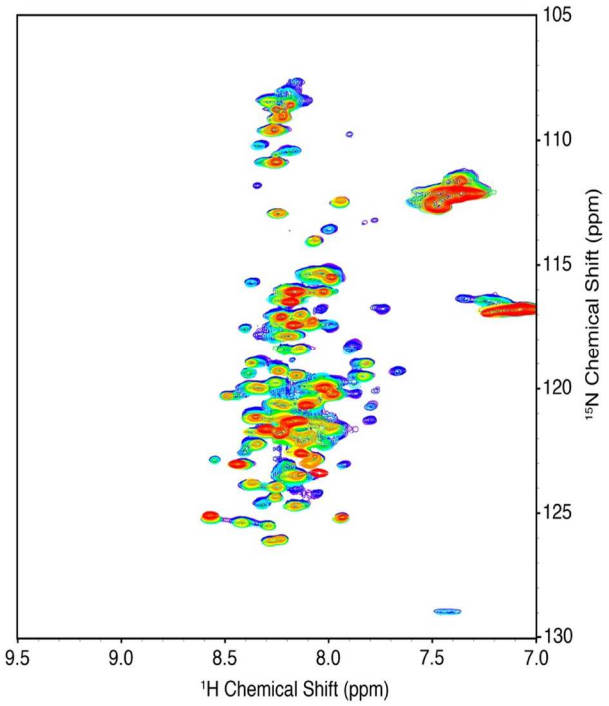

Figure 3.15- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N γ-Sarcoglycan incorporated into DHPC/DMPC bicelles for bicelle titration experiments. The colors of the overlayed spectra correspond to different q values as follows: q=0.1 purple, q=0.2 blue, q=0.5 cyan, q=1.0 green, q=1.5 yellow, q=2.0 orange, q=2.5 red.

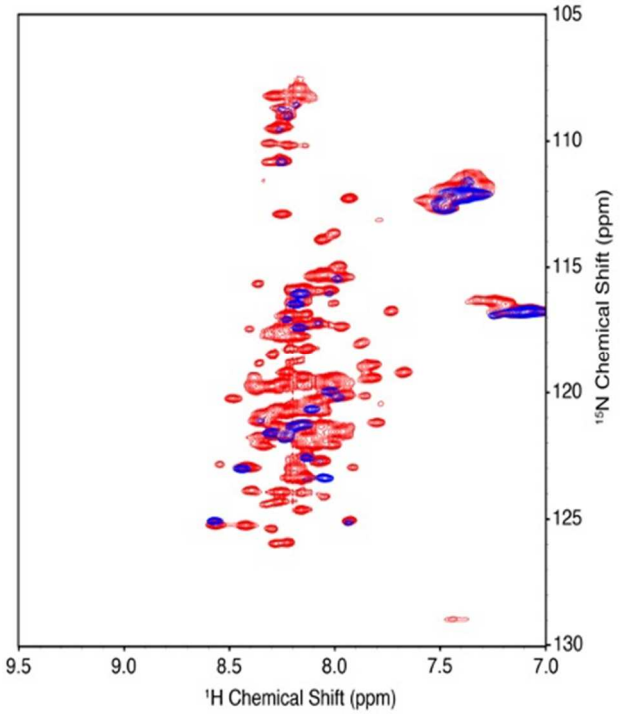

Figure 3.16- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N γ-Sarcoglycan incorporated into DHPC/DMPC bicelles comparing q=1.0 (red) and 2.5 (blue).

After running these experiments on ^{15}N labeled γ-Sarcoglycan, the next step forward was to see if we could replicate the ^{1}H-^{15}N HSQC spectrum in DDM with ^{15}N-^{13}C labeled γ-Sarcoglycan. Therefore, the protein was expressed and run on the 800 MHz NMR for better resolution and signal-to-noise so that the experiments could be run using fewer scans. A ^{1}H- ^{13}C HSQC was also run to make sure there was sufficient labeling from the ^{13}C glucose used in the expression of the protein. The ^{1}H-^{15}N HSQC experiment **(Figure 3.17)** looked the same as previous experiments, and the ^{1}H-^{13}C HSQC **(Figure 3.18)** showed good signal showing that we had sufficient ^{13}C labeling of the protein.

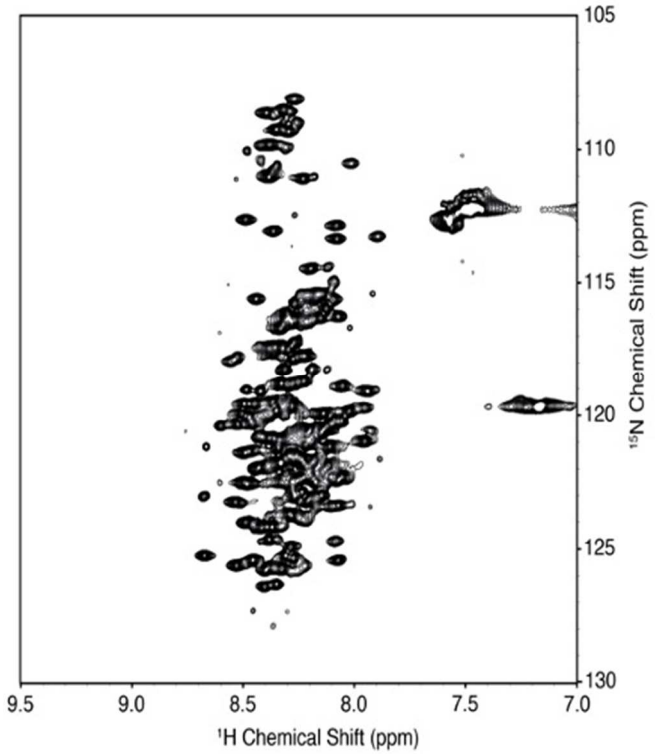

Figure 3.17- ^{1}H-^{15}N HSQC Spectra of uniformly labeled ^{15}N-^{13}C γ-Sarcoglycan in DDM micelles.

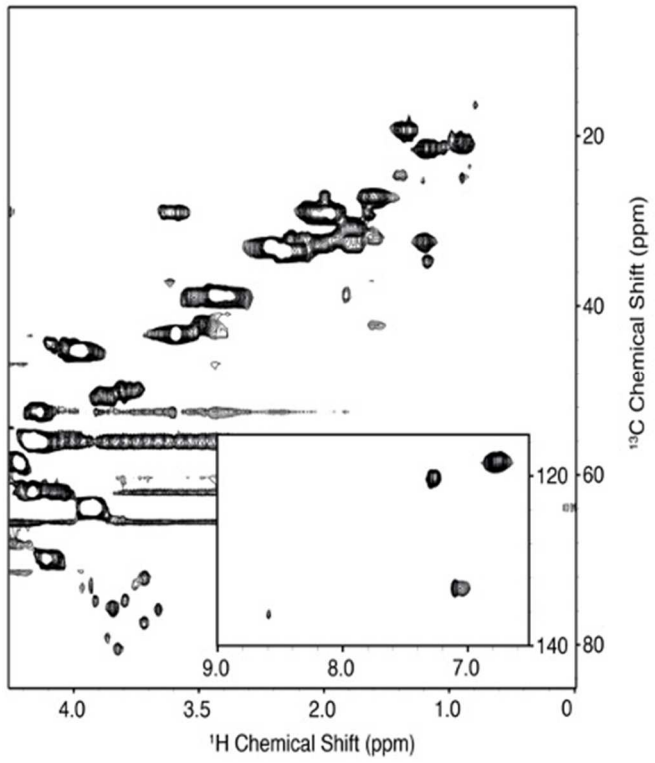

Figure 3.18- ^{1}H-^{13}C HSQC Spectra of uniformly labeled ^{15}N-^{13}C γ-Sarcoglycan in DDM micelles to show ^{13}C labeling.

The next two experiments run were HNCO & HNCACO in conjunction with each other. These two experiments allow the labelling of the ^{13}CO, ^{15}N, and ^{1}H atoms of the backbone that are connected to each other. The HNCO experiment is the primary spectrum used while the HNCACO is used to determine which carbonyl is part of the previous residue in the backbone. This is determined as the HNCO will have two peaks, one for the $^{13}CO_i$ and a peak which are connected in the backbone at a specific nitrogen and proton chemical shift. In the same nitrogen and proton chemical shift there will be another peak that lines up with the same ^{1}H and ^{15}N but at a different ^{13}C chemical shift and this as $^{13}CO_{i-1}$. This is confirmed by looking at the HNCACO and the peak that has all three the same chemical shifts is the $CO_{i-1,}$ and the other that does not have that peak corresponding the HNCACO is the CO_i peak. Using this you can show connectivity of the carbonyl group as seen in **Figure 3.19**. However, this will not allow us to go forward with the amino acid peak assignment. This will have to be done by using the HNCA and HNCOCA experiments in conjunction with each other.

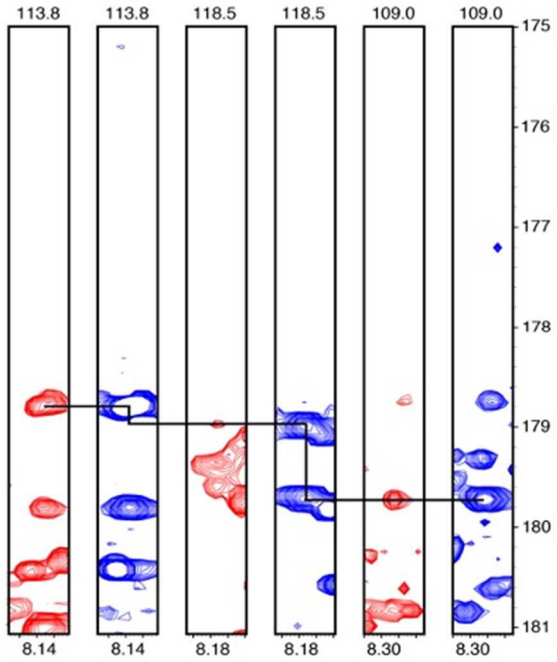

Figure 3.19- HNCO (blue) and HN(CA)CO (red) spectral slices showing backbone connectivity using lines for the attachment from each amino acid.

The HNCA and HNCOCA are examined in a similar way in which the primary spectrum is the HNCA and the HNCOCA is used to help determine the Cα corresponding to the previous residue. It is also similar in that the backbone residues that are connected have the same nitrogen and proton chemical shift, but they will have a different Cα chemical shift. To figure out which of the Cα is from the previous residue, $C\alpha_{i-1}$, one was the HNCOCA. The HNCOCA spectrum will show a single peak that has the exact same chemical shifts in all three dimensions. The peak that shows up in both spectra corresponds to the $C\alpha_{i-1}$ of the two peaks that are connected. This is shown in **Figure 3.20** and can be used for specific amino acid assignment of the backbone of the protein. This is done by using amino acids sequence in concert with the spectra and labeling from amino acids like glycine, threonine, and serine as they have distinct distinguishable chemical shifts from the other amino acids in the carbon dimension.

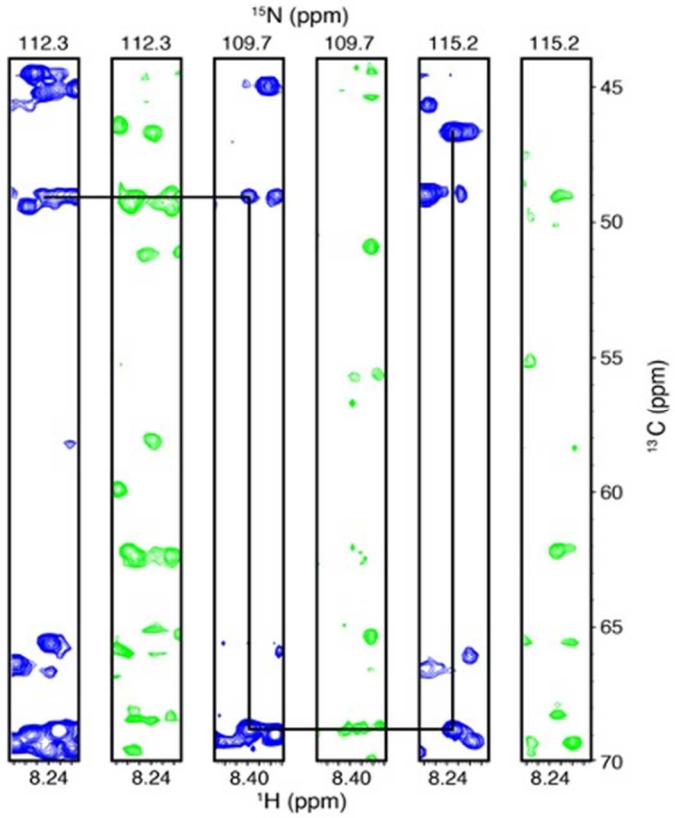

Figure 3.20- HNCA (blue) and HN(CO)CA (green) spectral slices showing backbone connectivity using lines for the attachment from each amino acid.

3.4 Discussion

γ-Sarcoglycan is a membrane protein that is part of the dystrophin complex. It is a vital portion of the complex and mutations in its amino acid sequence can lead to a loss of interaction with different proteins and ligands in the complex. If a loss of the interaction occurs, it can lead to a disease known as limb-girdle muscular dystrophy, a fatal disease for causing death to most people before they reach the age of twenty. Therefore, studying and understanding this protein is crucial in finding a way to treat and mitigate the effects of this disease that can be caused by mutations in this protein. The first step in this process is determining the secondary structure of the protein.

In this chapter, we showed the ability to express the γ-Sarcoglycan protein and then isolate it through purification steps. Upon purification and examination from a gel, the sample was then dissolved in solution and matrix and confirmed through MALDI-TOF mass spectrometry. The purification of the protein was confirmed when the experimental mass was determined to be 31970.15 Da corresponding to the theoretical mass 31959.16 Da of γ-Sarcoglycan. After purification, the steps for characterization of the protein started and a membrane mimetic was determined through examination of different detergents and a nanodisc sample. It was determined that DDM was the ideal detergent to use, as it gave better signal and resolution than the other detergents and nanodisc preparation. Once a detergent was chosen, initial experiments for amino acid peak assignment were run including: ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, HNCA, HNCO, HNCOCA, and HNCACO.

Abbreviations: cyanogen bromide (CNBR); 1,2 diheptanoyl-sn-glycero-3-phosphocholine (DHPC); n-Dodecyl β-D-maltoside (DDM); decylmaltoside (DM); dimyristoylphosphatidylcholine (DMPC); dithiothreitol (DTT); ethylenediaminetetraacetic acid (EDTA); Fast Protein Liquid Chromatography Size Exclusion (FPLC-SEC); n-hexadecylphosphocholine (Fos-choline 16); heteronuclear single quantum coherence (HSQC); Isopropyl β-D-thiogalactopyranoside (IPTG); *S*-(1-oxyl-2,2,3,5,5-pentamethylimidazolidin-4-ylmethyl (IMTSL); Limb-girdle muscular dystrophy (LGMD); Matrix Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF); membrane scaffold protein (MSP); nuclear magnetic resonance (NMR); Optical Density (OD); N-Luroylsarcosine (Sarkosyl); phenylmethylsulfonyl fluoride (PMSF); sodium dodecyl sulfate (SDS); sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE); tobacco etch virus (TEV);

CHAPTER IV

QUANTIFICATION OF *IN VITRO* GLYCOSYLATION OF THE MEMBRANE PROTEIN γ-SARCOGLYCAN USING N-GLYCOSYLTRANSFERASE

4.1 Introduction

4.1.1 Glycobiology

Glycobiology is the study of the structure, synthesis, biology, and evolution of saccharides that are also known as carbohydrates, sugar chains, or glycans.[89] This field of study encompasses the attachments of these carbohydrates onto proteins or lipids. When either mono-, oligo-, or polysaccharides are attached to either a protein or lipid a glycoprotein or glycolipid are formed respectively.[90] Studying these new glycoconjugates can lead to major contributions in the fields of biomedicine and biotechnology. For these reasons, studies in glycobiology are a rapidly increasing field of interest. The large diversity of structure and functions of these saccharides is shown in eukaryotes, and all cells in nature contain covalently linked sugars, monosaccharides, or sugar chains, oligosaccharides. Glycoproteins can contain multiple sugars or sugar chains and the process by which this attachment of sugars occurs is known as protein glycosylation. These new glycoproteins that are formed lead to a more diverse and complex proteome, with over half of all proteins expressed in a cell undergoing this modification.[41, 91] The different types of glycan structures attached to the same protein background, different oligosaccharides, and different sites of glycosylation lead to this diversity and complexity of the proteome.[92]

There are over several hundred distinct monosaccharides in nature. However, only a small minority of these are commonly found in vertebrates. Examples of these common monosaccharides include:[89]

- Pentoses: five-carbon neutral sugars, e.g., D-xylose (Xyl)
- Hexoses: six-carbon neutral sugars, e.g., D-glucose (Glc), D-galactose (Gal), and D-mannose (man)
- Hexosamines: hexoses with an amino group at the 2-position, which can be free or, more commonly, N-acetylated, e.g., N-acetyl-D-glucosamine (GlcNAc), and N-acetyl-D-galactosamine (GalNAc)
- Deoxyhexoses: neutral six-carbon sugar that lacks the hydroxyl group at the 6 position, e.g., L-fucose (Fuc)
- Uronic acids: hexoses with a negatively charged carboxylate group at the 6-position, e.g., D-glucuronic acid (GlcA) and L-iduronic acid (IdoA)
- Nonulosonic acid: family of nine-carbon acidic sugars, of which the most common in animals is the sialic acid N-acetylneuraminic acid (NANA)

Modifications of proteins by the attachment of glycans enhance the diversity and functions of proteins. The attachment of monosaccharides can change the diversity and structure of proteins, but more complex oligosaccharides or polysaccharides can affect them to an even greater extent. These attached glycans play a role in many different functions, and have been associated with cell-cell interactions, signal transduction, protein folding, secretion, inflammatory response, immune response, cancer development, and biological recognition processes.[41, 91]

4.1.2 N-linked and O-linked Glycosylation

The two most abundant forms of glycosylation in eukaryotes are N-linked glycosylation (attached at Asparagine) and O-linked glycosylation (attached at Serine or Threonine), named for how the sugar is attached to the protein. N-linked glycosylation is a fundamental and extensive post-translational modification in which an oligosaccharide is attached covalently to an asparagine residue in a polypeptide chain.[93] The attachment of the saccharide occurs at the amide nitrogen on the side chain and only when the asparagine residue is in the consensus sequence N-X-S/T, where X is any amino acid besides proline.[94] Dolichyl pyrophosphoryl oligosaccharide, containing the core $Man_3GlcNAc_2$ is the glycosyl donor for this reaction.[95] Oligosaccharyltrasnferase is the enzyme that catalyzes this reaction and transfers the glycan onto the asparagine forming a GlcNAc-(β-N)Asn bond. This acts as an initiating event to a trimming process where glycosidases hydrolytically remove numerous monosaccharide units, so that the pentasaccharide core is exposed.[95] The resulting glycoprotein then undergoes further modification by different glycosyltransferases and adds various sugars to the resulting pentasaccharide creating a diversity of oligosaccharide structures attached to the protein.[96, 97] This whole process occurs during translation meaning that it happens before the protein is in its final confirmation.[95] Oligosaccharyltransferase is a very intricate enzyme and it exists in low concentrations in human tissues and is extremely hard to isolate.[98, 99] This technique is hardly used since previous studies for glycosylation were unsuccessful. For this enzyme to be used for *in vitro* glycosylation unfolding of the protein is required.[100, 101] Therefore, due to the unavailability and sophistication of oligosaccharyltransferase it is impractical for use in the *in vitro* glycosylation of proteins for glycosylation studies.

O-linked glycosylation is different than that of N-linked glycosylation as a complex lipid-linked oligosaccharide is not needed and that the attachment site of the saccharide is at either a threonine or serine. The initiator of this type of glycosylation is the transfer of N-acetylgalactosamine

(GalNAc) from UDP-GalNAc to the oxygen of serine or threonine residue in the protein. This type of glycosylation occurs post-translationally and is catalyzed by the enzyme N-acetylgalactosaminyl transferase.[102] Upon the attachment of the first saccharide to the protein, subsequent monosaccharides can be added at the C3 and the C6 hydroxyl groups of GalNAc by different glycosyltransferases. These attachments lead to greater diversity and a wide range of glycoprotein structures.[95]

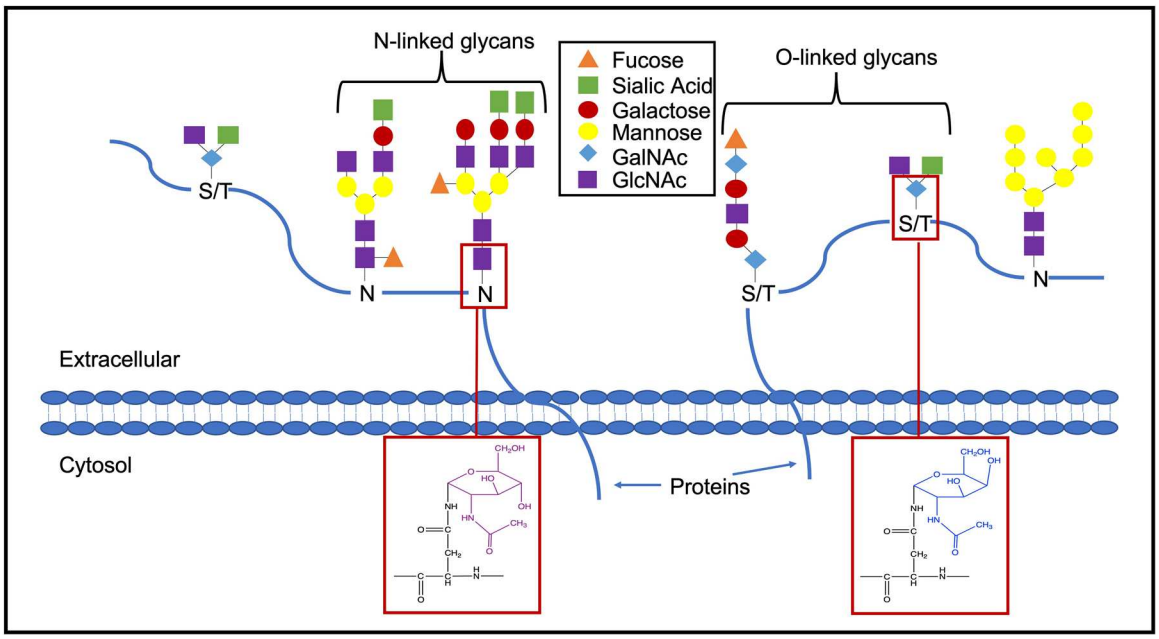

Figure 4.1- Two types of glycoconjugates are shown above. Proteins are glycosylated by the covalent linkage of the saccharide to the protein background, via N-linkage to asparagine or O-linkage to serine or threonine.

Glycoproteins play a pivotal role in many processes in the human body. However, the phenomenon of natural microheterogeneity makes studying protein glycosylation troublesome. Glycosylation can result in a range of different glycan structures, sites of their attachment, and variability in length and branching pattern, leading to the formation of different glycoforms. It is clear from a functional point that glycosylation is important, but the significance of microheterogeneity still remains unclear. This could further diversify the endogenous recognition function in cells.[89]

4.1.3 Diseases Associated with Glycosylation

Not only is glycosylation important for different functions of proteins such as cell-cell interactions and protein folding, but it can also be harmful and lead to disease if not maintained properly. Congenital disorders of glycosylation (CDG) are very common and can be classified as either type I or type II.[103] Type I CDG's are caused by the irregularities in the oligosaccharide structure on the glycolipid precursor before the attachment to the asparagine residue. Type II CDG's deal with defects in controlling N-linked branching structure of the new formed glycoprotein.[104, 105] The problems that can result from these CDG's can range from mental retardation to liver disfunction and intestinal disorders.[106]

Glycan attachment also plays major roles in inflammation and the immune system in humans. This was found to be true in interactions between endothelial cells and leukocytes.[107] They are crucial in recruitment of leukocytes to sites of tissue injury and are in turn regulated by cellular glycosylation. Glycosylation was then believed to be contributing to inflammatory vascular diseases when inflammatory cytokines induced changes in cell surface of N-linked glycosylation of endothelial cells.[107, 108]

In the adaptive immune system, glycosylation and glycans are vital to multiple roles in B cells and T cells. These functions are important in different cell surface and secreted proteins in the immune system (i.e., CD43, CD45, selectins, galectins, and siglecs). These different proteins upon glycosylation work in concert with T cells and B cells for cell-cell interactions and recognition of glycan-containing antigens.[109-111] Immunoglobulins are also important in the immune system for antibody-mediated immunity. When glycosylation patterns have been altered, and different isotypes have formed they have resulted in different diseases states such as rheumatoid arthritis and HIV infections.[112, 113]

Abnormal glycosylation has also been linked to cancer cell growth and their migration throughout the body. This is because glycosylation plays a role in different factors such as cellular division, the display of death signals, and immune surveillance. When different proteins are not glycosylated properly it allows for tumor growth as cells are prevented from entering apoptosis. An example is the abnormal growth factor signaling which occurs with specific irregular glycosylation motifs on relevant proteins causing cancer development.[114] For this exact reason glycosylation patterns were then studied and used as a biomarker to determine which cells were more likely to develop into cancer cells in both disease and healthy tissues.[115, 116]

Whether it be from a lack of glycosylation, over-glycosylation, or variance in glycan structures it is more important than ever to study these differences. Studying both N-linked and O-linked glycosylation can have an impact in the treatment of many disease states in humans and can even be used as a biomarker for cancer. The detection of these glycoconjugates and glycoproteins will only further help future studies and aid in determining the onset earlier in patients. Studies of these different potential biomarkers such as CA15-3, CA19-9, CA125, CEA, AFP, and PSA **(Table 4.1)**[117] can lead to applications for fighting cancer and possibly help as a model for determination of other cancer biomarkers. This would be the first step in improving early diagnosis, monitoring of the disease, and could be used to find potential therapeutic treatments for specific forms of cancers throughout the human body.

Table 4.1- Different serological biomarkers in applications of cancer.[117]

Serological Biomarker	Glycofrom or Glycoprotein	Cancer	Applications
CA15-3	MUC1	Breast	Monitoring
CA19-9	(Sle^a)	Colorectal, Pancreatic, and Gastric	Monitoring and recurrence
CA125	MUC16	Ovarian	Monitoring and recurrence
CEA	CEA	Colorectal	Monitoring and recurrence
PSA	PSA	Prostate	Diagnosis, monitoring, and recurrence

4.1.4 Literature Review

This large variation in oligosaccharide structure and their attachment to proteins have been the reason for the limited knowledge in determination of glycosylated protein structures and the mechanism by which they are glycosylated. This lack of understanding has stimulated research for designing methods for synthesizing these glycoproteins which are difficult to acquire from nature. Current studies from a multitude of research groups focus on formation of these glycopeptides and glycoproteins through chemical, enzymatic, and chemoenzymatic reactions.

One of the most common and straightforward approaches is through the attachment of a glycan to amino acids through a chemical reaction. The reaction uses an already reactive amino acid in the peptide or protein to conjugate a glycan.[95] This strategy has been used by multiple groups and the glycoproteins made are some of the most reported throughout literature.[97, 118, 119] An example of this, bovine serum albumin glycoconjugates, were formed using dextran and the Amadori

reaction.[120] Another important feature is that simple saccharides can be attached to albumin through commercial linkers.[121] A very similar example is the conjugation of a naturally derived mannopentose through reductive amination and its attachment to Ribonuclease A via coupling using an azide heterobifunctional reagent.[122] A different alternative was developed by Yuan and co-workers in which they ligate together two glycopeptides through a chemical method with a stable acyl donor.[123] However, this is difficult for many reasons including the complexity of the oligosaccharides and their chemical synthesis with glycoproteins require anhydrous conditions. They are usually incompatible with certain proteins, and there are many unfeasible protecting and deprotecting steps of both the glycans and peptide chains.[97]

Due to these problems, chemoenzymatic methods have become more of a focus by synthesizing an oligosaccharide precursor which can in turn be modified by a range of different glycosyltransferases. A chemoenzymatic method for site-specific ligation was developed by Wang and co-workers. This ligation entails an activated glycan oxazoline and a GlcNAc-protein to then create homogenous glycoproteins having asymmetrically branched N-linked glycans.[124, 125] The important factor for this chemoenzymatic method is the endoglycosidases that are used. These endoglycosidases first remove the N-linked glycans attached to the proteins and ligate on a larger intact oligosaccharide from a glycosyl donor, at the site in one step. These endoglycosidases work only with Asparagine linked donors for this transglycosylation and have been shown to create very complex glycopeptides, but yields are low.[126, 127] Depending on the desired oligosaccharide attachment, specific glycosyltransferases will be needed to initiate this process. Different enzymes transfer a specific sugar nucleotide donor substrate to a specific hydroxyl on the group of the accepting sugar.[94] Since the selectivity of the glycosyltransferases is unique, complex glycans are designed and synthesized in a predetermined manner specific for what is needed.[128, 129]

It is difficult to obtain specific protein modification and have the formation of a homogenous glycoprotein or glycopeptide if there is a single reactive amino acid that is targeted by the modifying reaction. To synthesize these homogenous glycoconjugates a few different strategies have been developed in the last decade, in which a specific glycan ligation site in a peptide chain is used for selective glycan attachment.[127, 129] A specific report of this is with Haemophilus influenzae HMW1 adehsin, in which glycosylation occurred in all but one case at the consensus sequence for N-glycosylation.[130] This research by Grass and co-workers showed the interesting feature that hexoses and dihexoses were attached at these sites instead of N-acetylated sugars which revealed the ability of a novel enzymatic attachment of hexose sugars to asparagine residues within a polypeptide chain.[130] More recent studies have been done by Aebi and co-workers have demonstrated the ability of an enzyme homolog of HMW1C from *Actinobacillus pleuroneumoniae* is a soluble N-glycosyltransferase (NGT). This enzyme has the capability of identifying proteins or peptides with the consensus sequence Asn-X-Ser/Thr, when X is any amino acid besides proline, and attach simple hexose monosaccharides from nucleotide activated sugar substrates.[41] This group was able to establish an assay for *in vitro* glycosylation of short peptides and have shown the ability of NGT to act as a novel enzyme in glycosylating soluble proteins.

Even though N-linked glycosylation occurs in both soluble and membrane bound proteins, most of the successful research has been with soluble proteins, and none had been shown with membrane proteins until recently. Liyanage et. al were able to use NGT and achieve N-glycosylation of two peptides and a membrane protein.[131] Membrane proteins make up 20-30% of the proteome and most of these are glycoproteins. Studying these membrane proteins after they have been glycosylated is important as they are relevant in many different diseases states and can be targeted for therapeutical treatments. However, studying these N-glycosylated membrane proteins can be very difficult due to the fact these membrane proteins have very hydrophobic

surfaces, are flexible, and lack stability.[3] These difficulties have not yet allowed for structure determination and structure activity relationships (SAR) of membrane glycoproteins. This states the urgency of the need to further investigate and determine ways to synthesize and quantify these newly formed glycoproteins and then study the structural and dynamic affects of glycosylation.

4.1.5 Summary and Outlook

The Cook group research has shown the ability to glycosylate certain membrane proteins through a chemoenzymatic method with the enzyme N-glycosyltransferase.[131] Therefore the main purpose of the experiments here was to accurately quantify the glycosylation of the membrane protein γ-Sarcoglycan. Studies were also done using different detergents for solubilization of the membrane protein, so it could interact with the soluble NGT for glycosylation. Upon glycosylation of the protein, it was quantified through trypsinolysis and proteomics using LC-MS/MS. For specific quantification studies normalization factors were needed for the peptide in question. This glycosylated peptide fragment from γ-Sarcoglycan is shown below in **Figure 4.2** and consists of amino acids 98-115.

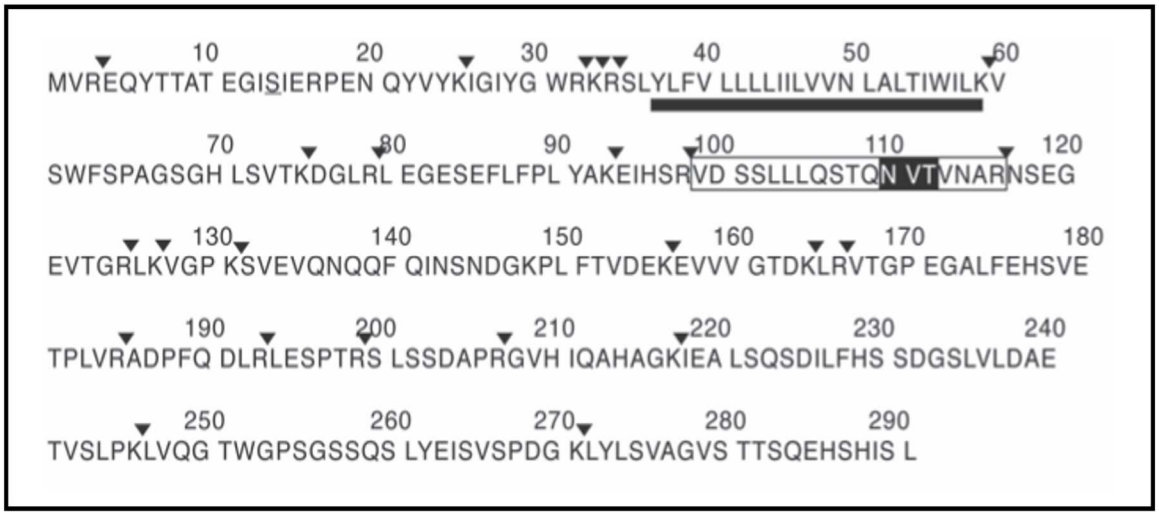

Figure 4.2- Amino acid sequence of the protein γ-Sarcoglycan with the transmembrane portion underlined, and the glycosylation site highlighted. Other important demarcations in the figure include the arrows showing where the protein is cleaved when interacted with trypsin, and the peptide of interest has a box around it as that is the cleaved peptide in which the glycosylation will be studies through LC-MS/MS.

4.2 Materials and Methods

4.2.1 Expression and Purification of Unlabeled γ-Sarcoglycan

For expression of γ-Sarcoglycan, a starter culture was prepared by inoculating 5 mL of LB media with 5 µL of carbenicillin at a concentration of 100 mg/mL and 10 µL of the γ-Sarcoglycan cell stock. This starter culture was then placed in a 37°C room and let grow for 2 h while shaking at 220 rpm. After 2 h the sample was taken off the shaker and 200 µL were transferred to each of the two 250 mL overnight growth flasks containing 100 mL of LB media. 100 µL of carbenicillin at 100 mg/mL was also added to the overnight growth solutions. These overnight growth flasks were then placed back in the 37°C room and left shaking at 220 rpm for approximately 14 h. The next morning 50 mL of the overnight growths were then transferred back to four different 2 L growth flasks containing 950 mL of LB media so each large growth flask will have 1 L of solution media. 1 mL of carbenicillin at 100 mg/mL was then added to the solution and the large growth flasks were placed once again in the 37°C room and left shaking at 220 rpm until the OD_{600} reached 0.6, approximately 3 hours. At this point the growth was induced with 1 mL of Isopropyl β-D-thiogalactopyranoside (IPTG) (GoldBioTechnology, www.goldbio.com) at 120 mg/mL to each 1 L growth. After induction with IPTG the sample was placed back on the shaker in the 37°C room at 220 rpm for 4 h. The cells were harvested by transferring them to 1 L polycarbonate centrifuge bottles and centrifuged at 5450 xg for 25 min at 4°C (Thermo scientific, SORVALL LINX 4000 Centrifuge). The cell pellets were then saved and stored in a -80°C freezer until they were to be purified.

The cell pellets from the expression of γ-Sarcoglycan was then resuspended in 30 mL of Resuspension Buffer I (50 mM Tris-HCl, 1mM EDTA, 1 mM NaN_3, and 205 mM glycerol). Upon resuspension of the pellet, the sample was sonicated (Fisher Scientific Sonic Dismembrator, www.fishersci.com) on ice for a total of 4 minutes at intervals of 2 seconds on then 8 seconds off.

The total elapsed time for the sample in the sonicator is 20 minutes. The samples were transferred to a 35 mL centrifuge tube and spun down in the centrifuge at 35000 xg for 25 min at 4°C (Thermo scientific, SORVALL LINX 4000 Centrifuge). After the sample was done running, the supernatant is discarded, and cell pellet containing the membrane fraction and inclusion bodies were saved.

The cell pellet is then resuspended once more, but this time in Resuspension Buffer II (50 mM Tris, 1 mM EDTA, 1 mL NaN_3, 25.5 mM Deoxycholic acid, 17.5 mM IGEPAL). Vortexing and a spatula was used to break up the cell pellet to help resuspend the pellet. The solution was then sonicated on ice for a total of 4 minutes at intervals of 2 seconds on and 8 seconds off so that a total of 20 minutes elapsed during sonication. Samples are once again transferred to a 35 mL centrifuge tube and counterbalanced with other samples. The tubes are then spun down at 35000 xg for 25 min at 4° C. Once finished, the supernatant was discarded.

The pellet, which contained the isolated inclusion bodies, was broken up in 30 mL of Binding Buffer (6 M guanidine hydrochloride, 0.5 M sodium chloride, and 20 mM Tris pH 8.0) and stored at 25°C overnight. The partially dissolved inclusion bodies were then sonicated on ice as previously described with the two resuspension buffers. After the sonication, the samples were spun down in the centrifuge at 18500 xg for 10 min at 4°C to remove any particulates that remained. The supernatant was then saved and dialyzed against 5 L of ddH_2O in 10.0 kDa dialysis bag (Spectrum Laboratories, Inc., spectrumlabs.com). The water was changed at 30-minutes and 90-minutes, then left overnight for dialysis. After the removal of guanidine, the protein solution was frozen and dried on the lyophilizer overnight.

Next the fusion partner, Trp Leader, was cleaved off by treating the protein with cyanogen bromide (CNBr) (Alfa Aesar, www.alfa.com). When running this cleavage reaction, we used 70% formic acid (Fisher Chemical, www.fishersci.com) and cyanogen bromide. For every 10 mg of

the target protein, we added 20 mg of CNBr and 400 μL of formic acid. Upon addition of the formic acid and CNBr, the sample was placed on a rotator in the absence of light for 3 hours to push the cleavage forward. After 3 h the sample was removed from the rotator and 1 Vol. the total solution of 1 M NaOH was added and the sample was placed back on the rotator for 10 min. A white precipitate formed. 1 Vol. of the original solution of DI H_2O was then added to the sample and it was transferred to a pre-wetted 10.0 kDa dialysis bag. The reaction tube is then washed once more with 1 Vol. of DI H_2O and then transferred to the same dialysis bag.

The sample was dialyzed against 4 L of DI H_2O to neutralize the sample and to remove any remnants of CNBr. The water was switched after 30-minutes and 90-minutes then left overnight. The wastewater was neutralized above a pH of above 7 and placed in a waste container to be disposed of by waste management. The sample was then transferred to a 50 mL conical centrifuge tube and frozen in the -80°C freezer. Once the sample was frozen it was placed on the lyophilizer to freeze dry the sample. The sample was taken off the lyophilizer once dry and prepared for FPLC Size Exclusion Chromatography (SEC).

FPLC-SEC was performed by transferring 20 mg of dry γ-Sarcoglycan to a 2.0 mL eppendorf tube and the protein was dissolved by adding 2.0 mL of 10% SDS. The sample was vortexed and placed in the sonication bath for 10 minutes to help solubilize the protein. The sample was centrifuged for 5 minutes and only the supernatant was injected into the size-exclusion FPLC system on a Bio-Rad NGC Quest 10 Plus (www.bio-rad.com). The FPLC running buffer (20 mM sodium phosphate, 4 mM SDS, 1 mM EDTA, 1 mM sodium azide, and pH 8.2) was run through two columns attached in series at a flow rate of 1.5 mL/min for a total of 600 mL. The two columns used were both HiPrep™ 26/60 Sephacryl™ S-200 HR (GE Healthcare, www.gelifesciences.com). The elution of the protein was monitored using absorbance at 280 nm and the sample peaks were collected and analyzed using SDS-PAGE. The fractions containing the pure protein were then pooled together and dialyzed against 5 L of DI H_2O in a 10.0 kDa dialysis

bag. Precipitated protein was then centrifuged, frozen, and dried on the lyophilizer, so it could be used for glycosylation studies.

4.2.2 Sample Preparation and Glycosylation of γ-Sarcoglycan

First, 100μg of γ-Sarcoglycan was solubilized in 500 μL of Trifluoroethanol (TFE). Stock solutions of DHPC were made as well, at concentrations of 450 mM and 900 mM, and these were also both solubilized in 500 μL of TFE. Each was placed in a sonication bath, and once they were fully solubilized the samples were combined so that the final DHPC concentrations were 225 mM and 450 mM with a total of 1 mL of solution for each respectively. Following this step, the solutions were dried under N_2 gas until there was a protein-lipid film. Following this, the samples were then left in the desiccator overnight to remove any residual moisture in the protein-lipid film.

The next day the sample was hydrated with 250 μL glycosylation buffer (25 mM Tris at pH 8, 150 mM NaCl) and placed in the sonication bath to help solubilize the sample. When running glycosylation experiments for quantification purposes, there was a reaction mixture and a control mixture. The reaction mixture consisted of the hydrated sample, 1.4 mg of UDP-glucose, and 250 μL of 25 μM NGT. The control samples were also prepared the same way, except that either NGT or glucose was absent. This was done to ensure that neither were prepared differently in the final concentrations of DHPC. Both samples were incubated on a rotator for 16 h at 25°C. Next, the samples were dialyzed against DI H_2O in 10.0 kDa dialysis bag. Precipitated protein was then spun down in a 1.5 mL microcentrifuge tube, and the pellet was frozen and dried on the lyophilizer.

4.2.3 Trypsinolysis LC-MS/MS of Glycosylated γ-Sarcoglycan

Following glycosylation, the dried protein powder of γ-Sarcoglycan was prepared for running SDS-PAGE. Approximately 0.1 mg of the protein was dissolved in 80 µL of 1X LDS buffer (50 mM Tris-HCl, 50 mM Tris base, 2% SDS, 10% glycerol, 1% 2-mercaptoethanol, 12.5 mM ethylenediaminetetraacetic acid, 0.02% bromophenol blue). The sample was then mechanically suspended with a 1 mL syringe by pulling the sample through the 26-gauge needle 30 times and then back out of the needle. The sample was then boiled for 10 minutes at 100°C.

Twenty µL of the prepared sample was loaded onto an SDS-PAGE gel. 4 µL of Page Ruler pre-stained protein ladder was loaded with at least one empty lane between the ladder and the sample to avoid contamination from the ladder to the sample. The samples were run on the gel at 150 V and 400 mA for 40 minutes. The gel was then transferred to coomassie blue stain (50% methanol, 40% water, 10% acetic acid, 1 g Brilliant Blue G-250) and was microwaved for 45 seconds. The gel was left on a rocker for 1 h before the coomassie blue stain was removed and the gel was placed in destain (50% water, 40% methanol, 10% acetic acid). The gel was then microwaved for 45 seconds in the destain solution and placed back on the rocker overnight. The gel was then taken out of destain after 16 h and was stored in DI H_2O until the band was ready to be excised for trypsinolysis and LC-MS. The control samples of γ-Sarcoglycan were prepared the exact same for trypsinolysis as the glycosylated samples. These samples were in the exact same manner for trypsin digestion and analysis at the Core Facility.

The γ-Sarcoglycan protein bands were excised and digested with trypsin using standard methodologies (e.g., Voruganti et al.).[132] Trypsinolytic peptides
were injected onto a 75 µm × 50 cm nanocolumn (Acclaim PepMap, Thermo PN 164942) and separated using a water-acetonitrile gradient (3−30% acetonitrile in 120 min) containing 0.1% formic acid. Peptides were eluted through a stainless-steel emitter and ionized within a

NanoSpray Flex ion source (Thermo). Peptide ions were analyzed by a "high−high" "top-speed" data-dependent acquisition using a quadrupole-Orbitrap mass spectrometer (Thermo Fusion), wherein the parent peptide ions were analyzed in the Orbitrap sector at a nominal resolution of 120000. The quadrupole sector was used to select peptide ions for HCD dissociation, and the fragment ions were analyzed within the Orbitrap sector at a nominal resolution of 30,000.

Peptides were identified by using the Byonic software application (Protein Metrics) to search RAW instrument files against a database of 4306 E. coli protein sequences downloaded from Uniprot and supplemented with 9 N-glycosyltransferase sequences representing recombinant proteins studied in this and related projects. Search settings were optimized using the Byonic Preview module, after which the searches were repeated, now including N-linked glycosylation (+162.0528) as an additional variable modification.

For select reaction monitoring, the quadrupole was programmed for targeted MS/MS scans of ion m/z's 527.53, 703.03, and 1054.05, representing +4, +3, and +2 ions, respectively, of the parent peptide containing the glycosylation site consisting of amino acids 98−115 of γ-Sarcoglycan. Each selected target ion was fragmented by HCD, followed by the wide-band scanning of fragment ions using the Orbitrap sector at a nominal resolution of 60,000. Ion-specific chromatograms were extracted using the parent ion scan filter, in conjunction with secondary filtering using the specific fragment m/z values indicated in figure legends. The process that the methods follow is shown in **Figure 4.3**.

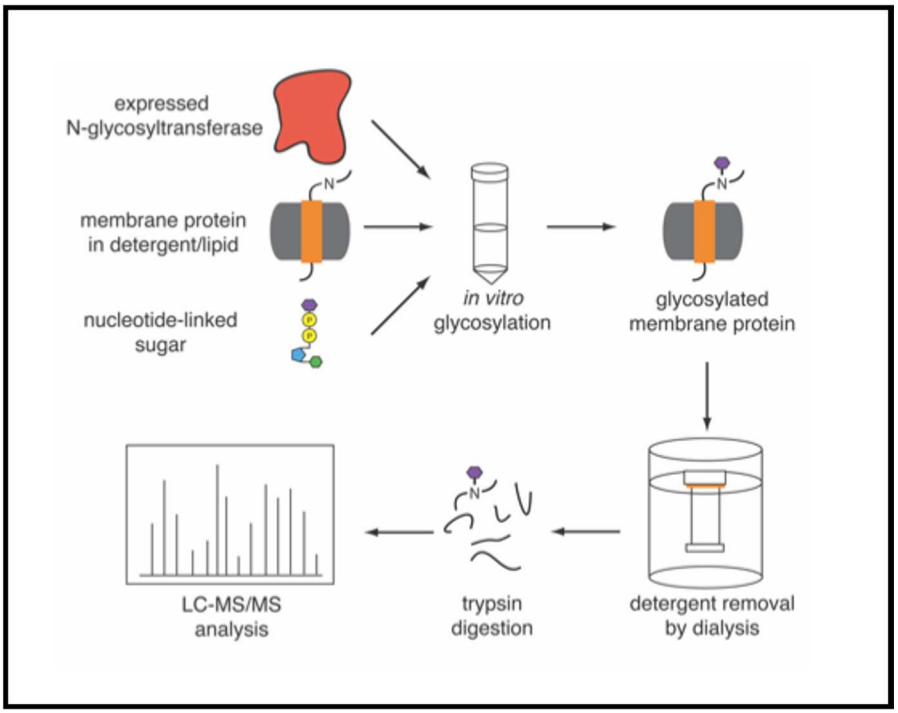

Figure 4.3- Diagram showing the stepwise process of glycosylation of γ-Sarcolgycan and the experimentation ran for percent glycosylation determination.

4.3 Results

This study was performed to glycosylate a membrane protein and quantify the percentage of glycosylation achieved through this chemoenzymatic reaction. Therefore, upon the purification of γ-Sarcoglycan, reaction and control samples were setup to determine if glycosylation occurs from the chemoenzymatic reaction. The reaction mixture consists of γ-Sarcoglycan in DHPC micelles, NGT, and UDP-glucose. The control lacks UDP-glucose so that no reaction occurs and can be used as a control for determination of glycosylation of the peptide and secondly for quantification of glycosylation.

After glycosylation, the sample was dialyzed against DI H_2O to remove the DHPC. This was done as detergents do not work well with the LC-MS/MS machines and can give inaccurate results. Therefore, the sample was dialyzed until the protein precipitated and the pelleted protein was frozen and dried on the lyophilizer. This dried samples were dissolved in 1X LDS buffer and

mechanically solubilized with a 1 mL syringe and ran on separate gels to avoid any possible contamination between the control and reaction samples. The gel **(Figure 4.4)** shown below was then sent to Dr. Hartson and Janet Rogers in the Recombinant DNA/Protein Core Facility (OSU) for trypsinolysis where the protein would be broken down into peptides shown in Figure 4.2. This sample was run through the LC-MS/MS as a proteomics study for glycosylation of the protein.

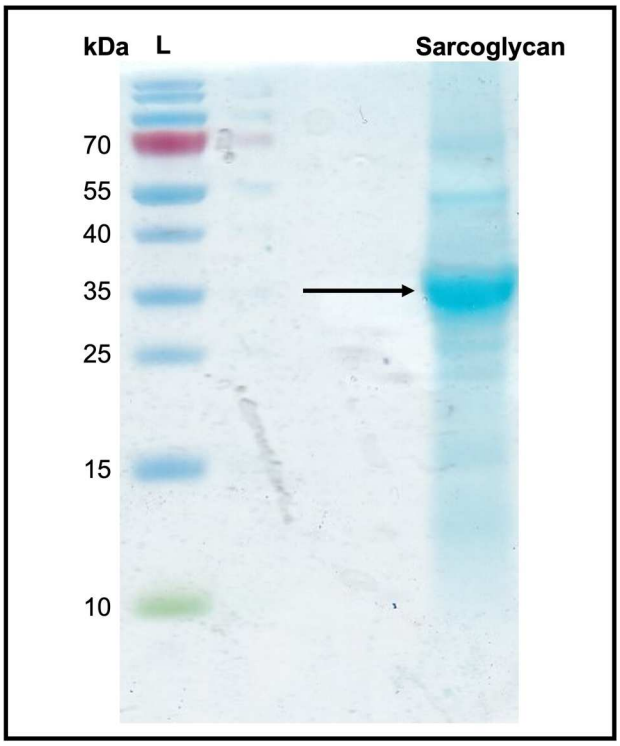

Figure 4.4- Picture of the gel sent down to Recombinant DNA/Protein Core Facility for Trypsinolysis and LC-MS/MS studies. The band at ~32 kDa depicted by an arrow is the band excised and studied.

The first study using LC-MS/MS was done to show that there was successful glycosylation in the reaction mixture and analyzed by Dr. Steve Hartson and myself. The mass at the fragment containing the modified asparagine was 162 Da higher than the control fragment. This mass was determined by using a peptide spectral match of the peptide through LC-MS/MS and the peaks were taken specifically looking at the y fragmention. The spectral peptide match for this has asparagine as the y7 amino acid and when not glycosylated the fragment should have a mass of 773 Da. The peptide spectral match **(Figure 4.5)** shows a peak at 935 Da which corresponds to

the attachment of the glucose onto the asparagine residue. This confirmed that we had achieved glycosylation of γ-Sarcoglycan.

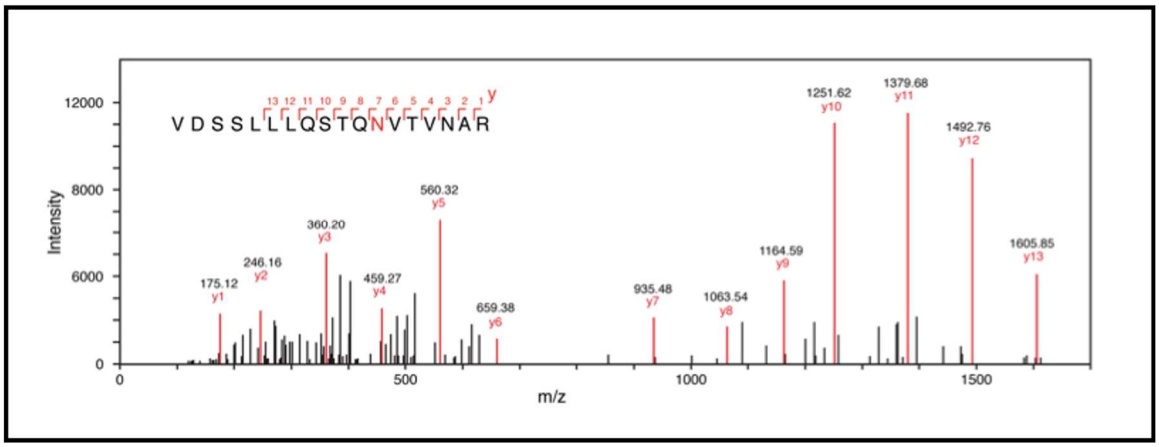

Figure 4.5- Peptide spectral match of the 98-115 peptide showing the attached glucose sugar on asparagine (y7) as the mass is 162 Da larger than expected if the peptide was not glycosylated.

After confirming glycosylation, the next step was to show that we have a control that shows no glycosylation. This was done by comparing peptide spectral match of the 98-115 peptide fragment and using select reaction monitoring (SRM) to show there is no meaningful elution at the time of elution of the glycosylated peptide from the reaction mixture. The peptide spectral match between the glycosylated sample and control sample is shown in **Figure 4.6** and the difference is clearly shown by 162 Da mass difference starting at the y7 fragment. The SRM data in **Figure 4.7** shows a small peak intensity for the control at 3.08 X 10^3 which is attributed to contaminants from the gel as the reaction mixture peak intensity was 27-fold larger at 8.40 X 10^4.

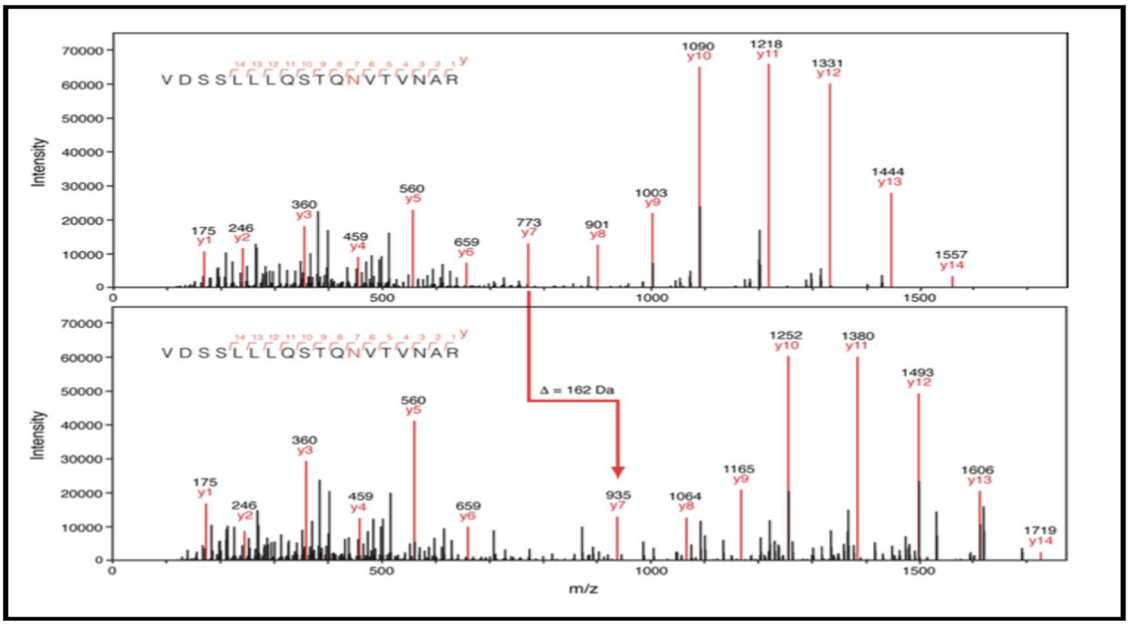

Figure 4.6- Peptide spectral match of the 98-115 peptide which shows the control was not glycosylated as we do not see a mass increase at the y7 asparagine peak (top), but the reaction mixture is glycosylated as there is a 162 Da mass increase at the y7 (bottom).

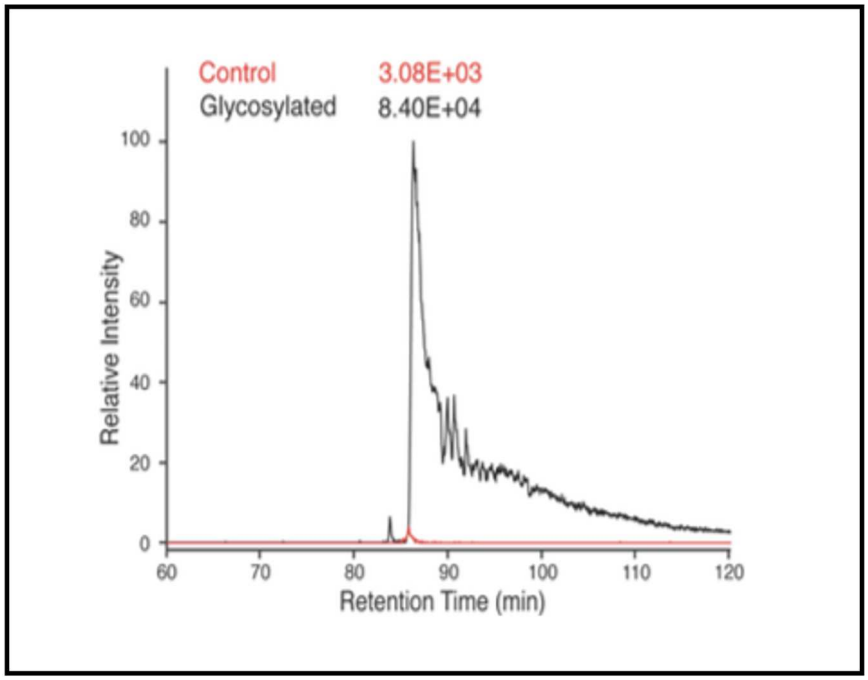

Figure 4.7- Select reaction monitoring (SRM) data showing the retention time of the 98-115 peptide and the intensity of the peptide. The retention time is around 88 minutes and shows that the control does not show any glycosylated peak and the small peak seen can be attributed to gel contamination.

After confirmation of glycosylation in only the reaction mixture, the next step was to run another set of samples and determine the percentage of glycosylation of γ-Sarcoglycan. This is done by

looking at the same 98-115 peptide that is not glycosylated and looking at the reduction in the intensity of this peak in the reaction mixture. The reduction of the peak in the extracted ion chromatogram after normalization allows us to quantify the amount of protein being glycosylated as the ratio corresponds to the amount of peptide that was being glycosylated. Our initial intensity of the peptide in question for the reaction and control were determined to be 3.89×10^8 and 7.48×10^8 respectively. However, working with quantities as small as 100 μg is difficult, so a normalization factor is required because it is impossible to ensure we have the same amount of protein in the control and reaction mixture. Therefore, a Max Quant data search was done to compare every other peptide from the γ-Sarcoglycan protein and quantify the amount of each one in each mixture. Peptides that were quantified in each mixture were then compared and outliers with intensities larger than 2 were removed from data analysis. The total list of peptide intensities found are shown in **Table 4.2** and includes 33 peptides. However, as stated before, if the intensity ratios were larger than 2, they were not used in data analysis, so only 24 peptides were used for normalization. The normalization factor of the 24 remaining peptides were determined to be 0.8441 and corresponds to the ratio of glycosylation mixture peptide intensities/control mixture peptide intensities. Correcting this, we then multiplied the control peak by the normalization factor to account for the differences in original protein amounts. The 98-115 peptide of the reaction mixture was still 3.89×10^8 and the control was now determined to be 6.31×10^8 **(Figure 4.8)**. This shows that we have a reduction of this peak in the reaction mixture of 38.5% indicating that 38.5% of the protein was glycosylated.

Table 4.2- Peptide intensities from LC-MS/MS data seen both in the control and reaction mixture. This data was used to determine the normalization factor between the control and reaction mixtures.

ID	Sequence	Intensity in control (x10^6)	Intensity in glycosylated sample (x10^6)
1	ADPFQDLR	14153.00	10272.00
2	ADPFQDLRLESPTR	57.32	86.44
3	DGLRLEGESEFLFPLYAK	794.13	645.21
4	EQYTTATEGISIER	82.92	71.33
5	EQYTTATEGISIERPENQYVYK	1507.90	8696.30
6	EVVVGTDK	8344.30	6796.50
7	EVVVGTDKLR	1465.60	1216.40
8	GVHIQAHAGK	8.67	0.00
9	IEALSQSDILFHSSDGSLVLDAETVSLPK	14997.00	13882.00
10	IGIYGWR	5248.30	3323.80
11	LEGESEFLFPLYAK	20436.00	16460.00
12	LEGESEFLFPLYAKEIHSR	8.80	0.00
13	LESPTRSLSSDAPR	4.54	18.81
14	LRVTGPEGALFEHSVETPLVR	61.96	57.96
15	LVQGTWGPSGSSQSLYEISVSPDGK	3743.00	5673.40
16	NSEGEVTGR	4963.70	6305.40
17	PENQYVYK	63.47	60.64
18	PLFTVDEK	143.78	101.74
19	PLFTVDEKEVVVGTDK	25.06	16.35
20	PLFTVDEKEVVVGTDKLR	5.42	0.00
21	SLSSDAPR	11119.00	9151.90
22	SLSSDAPRGVHIQAHAGK	16.62	2.34
23	SVEVQNQQFQINSNDGK	116.50	95.11
24	SVEVQNQQFQINSNDGKPLFTVDEK	9470.10	7781.20
25	SVEVQNQQFQINSNDGKPLFTVDEKEVVVGTDK	13.43	0.00
26	VDSSLLLQSTQNVTVNAR	4787.20	1770.10
27	VGPKSVEVQNQQFQINSNDGKPLFTVDEK	3.87	0.00
28	VREQYTTATEGISIER	24.07	15.13
29	VREQYTTATEGISIERPENQYVYK	8168.60	5991.20
30	VSWFSPAGSGHLSVTK	2634.20	2428.00
31	VSWFSPAGSGHLSVTKDGLR	19.31	8.07
32	VTGPEGALFEHSVETPLVR	26179.00	21224.00
33	VTGPEGALFEHSVETPLVRADPFQDLR	2.97	0.00

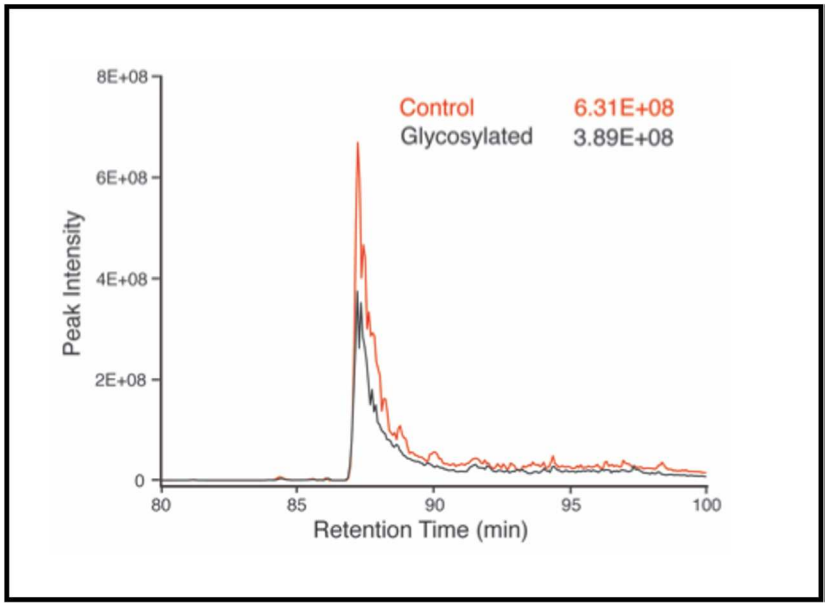

Figure 4.8- Extracted-Ion Chromatogram (XIC) of the 98-115 peptide that is not glycosylated and has been normalized using the factor determined from the peptides in Table 2. The disappearance in the glycosylation mixture correlates to the amount of glycosylated protein.

4.4 Discussion

Membrane proteins are difficult to study when they are not in their natural membrane environment. Therefore, when trying to attach glycans *in vitro* to these membrane proteins, special attention is needed to closely mimic the membrane environment. One way to do this is by dissolving the proteins in detergents that form micelles so the protein is in an environment that mimics the cell membrane. However, it was believed that these detergents could inhibit the ability of enzymes to attach glycans onto the proteins while in these conditions. This was proven to not be case as we were able to glycosylate the protein with UDP-glucose in these conditions. The protein was glycosylated using N-glycosyltransferase which attached a glucose sugar at the consensus NVT recognition site of the membrane protein γ-Sarcoglycan. The LC-MS/MS data showed that we were able to perform glycosylation indicated by the 162 Da shift in the data at the asparagine residue of the consensus sequence. This shows that the detergent at CMC levels did not inhibit the glycosylation and was even able to attach in a membrane mimetic environment.

Not only were we able to glycosylate the protein, but we were able to go from trace amounts of protein to quantifiable amounts of glycosylated protein. Through studies and determining the correct detergent and the specific concentrations needed for the reaction we were able to glycosylate up to 38.5% of the protein in the reaction mixture. This was a progressive step forward from the novel *in vitro* glycosylation by former graduate student Leshani Liyanage in the Cook Lab. This knowledge can be used as a model going forward for other membrane proteins in our lab and other labs throughout the world for the glycosylation of membrane proteins.

Abbreviations: Xylose (Xyl); Glucose (Glc); Galactose (Gal); Mannose (Man); N-acetyl-D-glucosamine (GlcNAc); N-acetyl-D-galactosamine (GalNAc); Fucose (Fuc); Glucuronic Acid (GlcA); Iduronic Acid (IdoA); N-acetylneuraminic acid (NANA); Uridine diphosphate-glucose (UDP-glc); Uridine diphosphate-N-acetyl-D-galactosamine (UDP-GalNAc); N-glycosyltransferase (NGT); Congenital Disorders of Glycosylation (CDG); Mucin (MUC); Sialyl Lewis Acid (Sle^a); carcinoembryonic antigen (CEA); prostate specific antigen (PSA); carcinoma antigen (CA); Structure activity relationship (SAR); Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS), Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Introduction

In this chapter, I will conclude my findings in the previous three chapters. I will also discuss future directions and goals for each research project.

5.2 Expression, Purification, and Preliminary Structural Studies of TCR α WT

Chapter 2 demonstrates the ability that we were the first group to recombinantly express and purify the TCR α protein. This is confirmed not only by SDS-PAGE gels, but through MALDI-TOF mass spectrometry in which we had a peak at 3510.96 Da corresponding to theoretical mass of 3521.22 Da for TCR α protein. After this we were able to do some preliminary structural studies of the protein through detergent screening using ^{1}H-^{15}N HSQC, and it was determined that DHPC was a suitable detergent as it had good signal to noise and resolution that was better than other detergents.

Future studies in this research will include structural determination of the TCR a protein, and determining how the attachment of an N-linked sugar effects the structure and function of the protein. This would consist of at least 3 different steps. The first step would be purifying ^{15}N and ^{13}C dual labelled protein to run more NMR studies including another ^{1}H-^{15}N HSQC, HNCA, HN(CO)CA, HNCO, H(NCA)CO, and ^{1}H-^{13}C HSQC experiments. The process following this

would be assigning amino acids to individual peaks and using the chemical shifts acquired by NMR as constraints for structure calculations. The second step would be the continual purification of the four TCR mutants L14C, K16C, A18C and M24C and attaching the IMTSL spin label. Labeled protein would be delivered to collaborators at NC State in the Smirnova lab for EPR studies. Lastly, TCR could be incorporated into different membrane mimetics including bicelles and nanodiscs for NMR studies. These membrane mimetics have yet to be tested and could yield spectra improved signal and resolution than was achieved with the DHPC micelles.

Another area of future interest with TCR α is in glycosylation of the protein. TCR α has an N-linked glycosylation site in its amino acid sequence at the amino acids 4-6 at the NLS consensus site. Upon structural determination, N-linked glycosylation experiments could be run to attach a sugar molecule onto the protein and these samples could be used to see how the attachment of a sugar effects the structure of TCR α through NMR ^{1}H-^{15}N HSQC experiments and other aforementioned experiments.

5.3 Expression, Purification, and Structural Studies of the Membrane Protein γ-Sarcoglycan

Chapter 3 shows the approach taken to express and purify the γ-Sarcoglycan protein. The purification of the protein was first confirmed through SDS-PAGE, and then through MALDI-TOF mass spectrometry. The peak for the spectrum found at 31970.15 Da corresponds to theoretical mass 31959.16 Da of γ-Sarcoglycan. This was the first time γ-Sarcolgycan protein has been recombinantly expressed and purified. After purification, a suitable membrane mimetic for the protein was determined to be DDM micelles. From here NMR studies were conducted including ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, HNCA, HNCO, HNCOCA, and HNCACO experiments to use for amino acid peak assignment.

Future studies for this research project will consist first of assigning the amino acid backbone to each peak in the different spectra obtained. Next, the chemical shifts for each amino acid will be

used as constraints for secondary structure determination. Once the structure is determined, the next step would be to use the glycosylation technique described in chapter 4 to glycosylate γ-Sarcoglycan in quantities high enough for NMR experiments. These experiments may tell us how glycosylation affects the structure. However, since not all of the protein is glycosylated in those experiments, a purification step would need to be performed to separate the glycosylated protein from the non-glycosylated protein.

5.4 Quantification of *In Vitro* Glycosylation of the Membrane Protein γ-Sarcoglycan Using N-glycosyltransferase

Chapter 4 demonstrates our ability to quantify the amount of protein that was glycosylated from our *in vitro* glycosylation reactions. This study was done by solubilizing a hydrophobic membrane protein in detergent acting as a membrane mimetic. After that, a chemoenzymatic reaction was carried out in which NGT attached a glucose onto the site-specific consensus sequence for N-linked glycosylation of γ-Sarcoglycan. This was then followed by trypsinolysis LC-MS/MS to quantify the amount of glycosylation by comparing it to a control using a normalization factor. We were able to determine that we were able to glycosylate up to 38.5% of γ-Sarcoglycan in the sample. This was a promising sign as it was the first membrane protein to be glycosylated through *in vitro* N-linked glycosylation above trace amounts. This should lead to eventual glycosylation of specific membrane proteins to study glycosylation effects and help us get closer to determining the cause of their protein disease states.

The plan is to expand the scope of this glycosylation study in three different steps. The next step forward in this research is trying to glycosylate this protein in other membrane mimetics besides detergents. The next specific membrane mimetic to try is nanodiscs using the membrane scaffold protein. During this step, optimization of the reaction would be ideal to see what reaction conditions work for protein nanodisc samples. The second step of this research is to see if we can

replace the already attached glucose sugar on the protein with a longer chain sugar. This process is done by purifying a larger sugar and using the Endoglycosidase A enzyme for transglycosylation of a larger sugar in place of the previously attached glucose. This would be confirmed by LC-MS/MS to prove that the larger sugar had been attached to the protein. The last step would then be using these two methods to attach the both the glucose and the larger sugars to the labelled ^{15}N and ^{15}N-^{13}C γ-Sarcoglycan protein. Since not all of the protein would be glycosylated, the sample would then need to be separated to obtain just glycosylated protein. This can be done by using a glycan affinity column in which the glycosylated protein would bind and be separated through glycan affinities. NMR experiments could then be run on the protein and compared to unglycosylated γ-Sarcoglycan to see how it effects the protein structure and dynamics.

REFERENCES

1. Almen, M. S.; Nordstrom, K. J.; Fredriksson, R.; Schioth, H. B., Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol* **2009,** *7*, 50.

2. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **2001,** *305* (3), 567-80.

3. Carpenter, E. P.; Beis, K.; Cameron, A. D.; Iwata, S., Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol* **2008,** *18* (5), 581-6.

4. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L., How many drug targets are there? *Nat Rev Drug Discov* **2006,** *5* (12), 993-6.

5. Page, R. C.; Moore, J. D.; Nguyen, H. B.; Sharma, M.; Chase, R.; Gao, F. P.; Mobley, C. K.; Sanders, C. R.; Ma, L.; Sonnichsen, F. D.; Lee, S.; Howell, S. C.; Opella, S. J.; Cross, T. A., Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins. *J Struct Funct Genomics* **2006,** *7* (1), 51-64.

6. Seddon, A. M.; Curnow, P.; Booth, P. J., Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta* **2004,** *1666* (1-2), 105-17.

7. Lin, S. H.; Guidotti, G., Purification of membrane proteins. *Methods Enzymol* **2009,** *463*, 619-29.

8. Yeagle, P. L., Chapter 10 - Membrane Proteins. In *The Membranes of Cells (Third Edition)*, Yeagle, P. L., Ed. Academic Press: Boston, 2016; pp 219-268.

9. Blobel, G., Intracellular protein topogenesis. *Proc Natl Acad Sci U S A* **1980,** *77* (3), 1496-500.

10. White, S. H.; Wimley, W. C., Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* **1999,** *28*, 319-65.

11. Killian, J. A.; von Heijne, G., How proteins adapt to a membrane-water interface. *Trends Biochem Sci* **2000,** *25* (9), 429-34.

12. Lodish H, B. A., Zipursky SL, et al., Membrane Proteins. In *Molecular Cell Biology*, 4th ed.; W.H. Freeman: New York, 2000.

13. Janeway CA Jr, T. P., Walport M, et al., General Principles of transmembrane signaling. In *Immunobiology: The Immune System in Health and Disease*, 5th ed.; Garland Science: New York, 2001.

14. Sezonov, G.; Joseleau-Petit, D.; D'Ari, R., Escherichia coli physiology in Luria-Bertani broth. *J Bacteriol* **2007,** *189* (23), 8746-9.

15. Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology* **2014,** *5* (172).

16. Pope, B.; Kent, H. M., High efficiency 5 min transformation of Escherichia coli. *Nucleic Acids Res* **1996,** *24* (3), 536-537.

17. Roy, A., Membrane preparation and solubilization. *Methods Enzymol* **2015,** *557*, 45-56.

18. Diefenderfer, C.; Lee, J.; Mlyanarski, S.; Guo, Y.; Glover, K. J., Reliable expression and purification of highly insoluble transmembrane domains. *Anal Biochem* **2009,** *384* (2), 274-8.

19. Klopfer, K.; Hagn, F., Beyond detergent micelles: The advantages and applications of non-micellar and lipid-based membrane mimetics for solution-state NMR. *Prog Nucl Magn Reson Spectrosc* **2019,** *114-115*, 271-283.

20. Antrace, Detergents and Their Uses in Membrane Protein Science. p. 12. https://www.anatrace.com/Technical-Documentation/Catalogs/Anatrace-Detergents-Booklet-FINAL#:~:text=While%20lipids%20also%20have%20the,concentration%20range%20required%20for%20aggregation%20.

21. Anandan, A.; Vrielink, A., Detergents in Membrane Protein Purification and Crystallisation. *Adv Exp Med Biol* **2016,** *922*, 13-28.

22. Warschawski, D. E.; Arnold, A. A.; Beaugrand, M.; Gravel, A.; Chartrand, E.; Marcotte, I., Choosing membrane mimetics for NMR structural studies of transmembrane proteins. *Biochim Biophys Acta* **2011,** *1808* (8), 1957-74.

23. le Maire, M.; Champeil, P.; Møller, J. V., Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2000,** *1508* (1), 86-111.

24. Lund, S.; Orlowski, S.; de Foresta, B.; Champeil, P.; le Maire, M.; Moller, J. V., Detergent structure and associated lipid as determinants in the stabilization of solubilized Ca2+-ATPase from sarcoplasmic reticulum. *J Biol Chem* **1989,** *264* (9), 4907-15.

25. Bhairi, S. M.; Mohan, C., Detergents; A guide to the properties and uses of detergents in biology and biochemsitry. EMD Biosciences: San Diego, CA, 2007. http://sevierlab.vet.cornell.edu/resources/Research-Links-Page/2-Molecular-Biology-and-Biochemisty-Resources/CALBIOCHEM-DetergentsIV.pdf.

26. G-Biosciences Detergents: Handbook and Slection Guide to Detergents & Detergent Removal. https://www.gbiosciences.com/image/pdfs/handbook/Detergent_Handbook.pdf.

27. Otzen, D. E., Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature. *Biophys J* **2002,** *83* (4), 2219-30.

28. Information, N. C. f. B. PubChem Compound Summary for CID 5590, Octooxinol. . https://pubchem.ncbi.nlm.nih.gov/compound/5590#section=Information-Sources (accessed August 12).

29. Information, N. C. f. B. PubChem Compound Summary for CID 3423265, Sodium Dodecyl Sulfate. https://pubchem.ncbi.nlm.nih.gov/compound/Sodium-dodecyl-sulfate (accessed August 12).

30. Lipids., A. P. Avanti Polar Lipids. https://avantilipids.com/?gclid=CjwKCAjwjdOIBhA_EiwAHz8xm1EsVvfj3TicC2HuhBL-L98uVvD8h3X1L37TxVEto4mJXgkPOIhV0xoCBNkQAvD_BwE (accessed August 12).

31. Puthenveetil, R.; Vinogradova, O., Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments. *J Biol Chem* **2019,** *294* (44), 15914-15931.

32. Marion, D., An introduction to biological NMR spectroscopy. *Mol Cell Proteomics* **2013,** *12* (11), 3006-25.

33. Zhou, Y.; Cierpicki, T.; Jimenez, R. H.; Lukasik, S. M.; Ellena, J. F.; Cafiso, D. S.; Kadokura, H.; Beckwith, J.; Bushweller, J. H., NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. *Mol Cell* **2008,** *31* (6), 896-908.

34. Liang, B.; Tamm, L. K., NMR as a tool to investigate the structure, dynamics and function of membrane proteins. *Nat Struct Mol Biol* **2016,** *23* (6), 468-74.

35. Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A., Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **1990,** *89* (3), 496-514.

36. Grzesiek, S.; Bax, A., Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969)* **1992,** *96* (2), 432-440.

37. Muhandiram, D. R.; Kay, L. E., Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B* **1994,** *103* (3), 203-216.

38. Clubb, R. T.; Thanabal, V.; Wagner, G., A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N□13C-labelled proteins. *Journal of Magnetic Resonance (1969)* **1992,** *97* (1), 213-217.

39. Wang, Y. C.; Peterson, S. E.; Loring, J. F., Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res* **2014,** *24* (2), 143-60.

40. Kelleher, D. J.; Gilmore, R., An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **2006,** *16* (4), 47R-62R.

41. Schwarz, F.; Fan, Y. Y.; Schubert, M.; Aebi, M., Cytoplasmic N-glycosyltransferase of Actinobacillus pleuropneumoniae is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *J Biol Chem* **2011,** *286* (40), 35267-74.

42. Jefferis, R., Glycosylation as a strategy to improve antibody-based therapeutics. *Nat Rev Drug Discov* **2009,** *8* (3), 226-34.

43. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** *126* (5), 855-67.

44. Marth, J. D.; Grewal, P. K., Mammalian glycosylation in immunity. *Nat Rev Immunol* **2008,** *8* (11), 874-87.

45. Wallin, E.; von Heijne, G., Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **1998,** *7* (4), 1029-38.

46. Weiss, A.; Littman, D. R., Signal transduction by lymphocyte antigen receptors. *Cell* **1994,** *76* (2), 263-74.

47. Samelson, L. E.; Harford, J. B.; Klausner, R. D., Identification of the components of the murine T cell antigen receptor complex. *Cell* **1985,** *43* (1), 223-31.

48. Garcia, K. C., Molecular interactions between extracellular components of the T-cell receptor signaling complex. *Immunol Rev* **1999,** *172*, 73-85.

49. Mak, T. W. S., M.E. , T Cell Activation. In *The Immune Response*, 1st ed.; 2005; pp 373-401.

50. Budd, R. C. F., K. A. , T Lymphocytes. In *Kelley and Firestein's Textbook of Rheumatology*, 10th ed.; Elseveier: 2017; pp 189-206.

51. Clambey, E. T.; Davenport, B.; Kappler, J. W.; Marrack, P.; Homann, D., Molecules in medicine mini review: the alphabeta T cell receptor. *J Mol Med (Berl)* **2014,** *92* (7), 735-41.

52. Rad., B. Mini-review: The T Cell Marker, CD3 Antigen and Antibodies. . https://www.bio-rad-antibodies.com/minireview-cd3-antibody.html (accessed Oct. 11).

53. Kirchgessner, H.; Dietrich, J.; Scherer, J.; Isomaki, P.; Korinek, V.; Hilgert, I.; Bruyns, E.; Leo, A.; Cope, A. P.; Schraven, B., The transmembrane adaptor protein TRIM regulates T cell receptor (TCR) expression and TCR-mediated signaling via an association with the TCR zeta chain. *J Exp Med* **2001,** *193* (11), 1269-84.

54. Alberts B, J. A., Lewis J, et al., Helper T Cells and Lymphocyte Activation. In *Molecular Biology of the Cell*, Garland Science: New York, 2002.

55. Karki, G. T-cell receptor (TCR): structure, role and TCR-CD3 complex. https://www.onlinebiologynotes.com/t-cell-receptor-tcr-structure-role-and-tcr-cd3-complex/.

56. Franco, R.; Martinez-Pinilla, E.; Lanciego, J. L.; Navarro, G., Basic Pharmacological and Structural Evidence for Class A G-Protein-Coupled Receptor Heteromerization. *Front Pharmacol* **2016,** *7*, 76.

57. Xu, X.; Li, H.; Xu, C., Structural understanding of T cell receptor triggering. *Cellular & Molecular Immunology* **2020,** *17* (3), 193-202.

58. Eckle, S. B.; Turner, S. J.; Rossjohn, J.; McCluskey, J., Predisposed alphabeta T cell antigen receptor recognition of MHC and MHC-I like molecules? *Curr Opin Immunol* **2013,** *25* (5), 653-9.

59. Irvine, D. J.; Purbhoo, M. A.; Krogsgaard, M.; Davis, M. M., Direct observation of ligand recognition by T cells. *Nature* **2002,** *419* (6909), 845-9.

60. Rudolph, M. G.; Stanfield, R. L.; Wilson, I. A., How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol* **2006,** *24*, 419-66.

61. Kuhns, M. S.; Badgandi, H. B., Piecing together the family portrait of TCR-CD3 complexes. *Immunol Rev* **2012,** *250* (1), 120-43.

62. Wucherpfennig, K. W.; Gagnon, E.; Call, M. J.; Huseby, E. S.; Call, M. E., Structural biology of the T-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling. *Cold Spring Harb Perspect Biol* **2010,** *2* (4), a005140.

63. Call, M. E.; Pyrdol, J.; Wiedmann, M.; Wucherpfennig, K. W., The organizing principle in the formation of the T cell receptor-CD3 complex. *Cell* **2002,** *111* (7), 967-79.

64. Sun, Z. J.; Kim, K. S.; Wagner, G.; Reinherz, E. L., Mechanisms contributing to T cell receptor signaling and assembly revealed by the solution structure of an ectodomain fragment of the CD3 epsilon gamma heterodimer. *Cell* **2001,** *105* (7), 913-23.

65. Arnett, K. L.; Harrison, S. C.; Wiley, D. C., Crystal structure of a human CD3-epsilon/delta dimer in complex with a UCHT1 single-chain antibody fragment. *Proc Natl Acad Sci U S A* **2004,** *101* (46), 16268-73.

66. Call, M. E.; Schnell, J. R.; Xu, C.; Lutz, R. A.; Chou, J. J.; Wucherpfennig, K. W., The structure of the zetazeta transmembrane dimer reveals features essential for its assembly with the T cell receptor. *Cell* **2006,** *127* (2), 355-68.

67. Samelson, L. E., Signal transduction mediated by the T cell antigen receptor: the role of adapter proteins. *Annu Rev Immunol* **2002,** *20*, 371-94.

68. Borroto, A.; Abia, D.; Alarcon, B., Crammed signaling motifs in the T-cell receptor. *Immunol Lett* **2014,** *161* (1), 113-7.

69. Altenbach, C.; Oh, K.-J.; Trabanino, R. J.; Hideg, K.; Hubbell, W. L., Estimation of Inter-Residue Distances in Spin Labeled Proteins at Physiological Temperatures: Experimental Strategies and Practical Limitations. *Biochemistry* **2001,** *40* (51), 15471-15482.

70. Singh, A.; Upadhyay, V.; Upadhyay, A. K.; Singh, S. M.; Panda, A. K., Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial Cell Factories* **2015,** *14* (1), 41.

71. Cook, G. A.; Stefer, S.; Opella, S. J., Expression and purification of the membrane protein p7 from hepatitis C virus. *Biopolymers* **2011,** *96* (1), 32-40.

72. Stangl, M.; Veerappan, A.; Kroeger, A.; Vogel, P.; Schneider, D., Detergent properties influence the stability of the glycophorin A transmembrane helix dimer in lysophosphatidylcholine micelles. *Biophys J* **2012,** *103* (12), 2455-64.

73. Laguerre, A.; Lohr, F.; Henrich, E.; Hoffmann, B.; Abdul-Manan, N.; Connolly, P. J.; Perozo, E.; Moore, J. M.; Bernhard, F.; Dotsch, V., From Nanodiscs to Isotropic Bicelles: A Procedure for Solution Nuclear Magnetic Resonance Studies of Detergent-Sensitive Integral Membrane Proteins. *Structure* **2016,** *24* (10), 1830-1841.

74. Berman, H.; Henrick, K.; Nakamura, H., Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **2003,** *10* (12), 980.

75. Dong; Zheng, L.; Lin, J.; Zhang, B.; Zhu, Y.; Li, N.; Xie, S.; Wang, Y.; Gao, N.; Huang, Z., Structural basis of assembly of the human T cell receptor-CD3 complex. *Nature* **2019,** *573* (7775), 546-552.

76. Sweeney, H. L.; Barton, E. R., The dystrophin-associated glycoprotein complex: What parts can you do without? *Proceedings of the National Academy of Sciences* **2000,** *97* (25), 13464-13466.

77. Durbeej, M.; Campbell, K. P., Muscular dystrophies involving the dystrophin-glycoprotein complex: an overview of current mouse models. *Curr Opin Genet Dev* **2002,** *12* (3), 349-61.

78. McDonald, C. M.; Abresch, R. T.; Carter, G. T.; Fowler, W. M., Jr.; Johnson, E. R.; Kilmer, D. D.; Sigford, B. J., Profiles of neuromuscular diseases. Duchenne muscular dystrophy. *Am J Phys Med Rehabil* **1995,** *74* (5 Suppl), S70-92.

79. Brooke, M. H.; Fenichel, G. M.; Griggs, R. C.; Mendell, J. R.; Moxley, R.; Florence, J.; King, W. M.; Pandya, S.; Robison, J.; Schierbecker, J.; et al., Duchenne muscular dystrophy: patterns of clinical progression and effects of supportive therapy. *Neurology* **1989,** *39* (4), 475-81.

80. Emery, A. E. H., The muscular dystrophies. *BMJ* **1998,** *317* (7164), 991-995.

81. Petrof, B. J.; Shrager, J. B.; Stedman, H. H.; Kelly, A. M.; Sweeney, H. L., Dystrophin protects the sarcolemma from stresses developed during muscle contraction. *Proc Natl Acad Sci U S A* **1993,** *90* (8), 3710-4.

82. Wagner, K. R., Genetic diseases of muscle. *Neurol Clin* **2002,** *20* (3), 645-78.

83. Fardeau, M.; Eymard, B.; Mignard, C.; Tome, F. M.; Richard, I.; Beckmann, J. S., Chromosome 15-linked limb-girdle muscular dystrophy: clinical phenotypes in Reunion Island and French metropolitan communities. *Neuromuscul Disord* **1996,** *6* (6), 447-53.

84. Angelini, C.; Fanin, M.; Freda, M. P.; Duggan, D. J.; Siciliano, G.; Hoffman, E. P., The clinical spectrum of sarcoglycanopathies. *Neurology* **1999,** *52* (1), 176-9.

85. UniProtKB - Q13326 (SGCG_HUMAN). https://www.uniprot.org/uniprot/Q13326 (accessed November 1st).

86. Nowak, K. J.; Walsh, P.; Jacob, R. L.; Johnsen, R. D.; Peverall, J.; McNally, E. M.; Wilton, S. D.; Kakulas, B. A.; Laing, N. G., Severe gamma-sarcoglycanopathy caused by a novel missense mutation and a large deletion. *Neuromuscul Disord* **2000,** *10* (2), 100-7.

87. McNally, E. M.; Passos-Bueno, M. R.; Bonnemann, C. G.; Vainzof, M.; de Sa Moreira, E.; Lidov, H. G.; Othmane, K. B.; Denton, P. H.; Vance, J. M.; Zatz, M.; Kunkel, L. M., Mild and severe muscular dystrophy caused by a single gamma-sarcoglycan mutation. *Am J Hum Genet* **1996,** *59* (5), 1040-7.

88. Piccolo, F.; Jeanpierre, M.; Leturcq, F.; Dodé, C.; Azibi, K.; Toutain, A.; Merlini, L.; Jarre, L.; Navarro, C.; Krishnamoorthy, R.; Tomé, F. M. S.; Urtizberea, J. A.; Beckmann, J. S.; Campbell, K. P.; Kaplan, J.-C., A Founder Mutation in the γ-Sarcoglycan Gene of Gypsies Possibly Predating Their Migration Out of India. *Human Molecular Genetics* **1996,** *5* (12), 2019-2022.

89. Varki, A.; Kornfeld, S., Historical Background and Overview. In *Essentials of Glycobiology*, rd; Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor (NY), 2015; pp 1-18.

90. Taylor, M. E.; Drickamer, K., *Introduction to Glycobiology*. OUP Oxford: 2011.

91. Slynko, V.; Schubert, M.; Numao, S.; Kowarik, M.; Aebi, M.; Allain, F. H., NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. *J Am Chem Soc* **2009,** *131* (3), 1274-81.

92. Gamblin, D. P.; Scanlan, E. M.; Davis, B. G., Glycoprotein synthesis: an update. *Chem Rev* **2009,** *109* (1), 131-63.

93. Schwarz, F.; Aebi, M., Mechanisms and principles of N-linked protein glycosylation. *Current Opinion in Structural Biology* **2011,** *21* (5), 576-582.

94. Kornfeld, R.; Kornfeld, S., Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* **1985,** *54*, 631-64.

95. Khmelnitsky, Y. L., Current strategies for in vitro protein glycosylation. *Journal of Molecular Catalysis B: Enzymatic* **2004,** *31* (4), 73-81.

96. Taylor, C. M., Glycopeptides and glycoproteins: Focus on the glycosidic linkage. *Tetrahedron* **1998,** *54* (38), 11317-11362.

97. Grogan, M. J.; Pratt, M. R.; Marcaurelle, L. A.; Bertozzi, C. R., Homogeneous glycopeptides and glycoproteins for biological investigation. *Annu Rev Biochem* **2002,** *71*, 593-634.

98. Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., In *Essentials of Glycobiology*, Cold Spring Harbor (NY), 2009.

99. Meynial-Salles, I.; Combes, D., In vitro glycosylation of proteins: an enzymatic approach. *J Biotechnol* **1996,** *46* (1), 1-14.

100. Liu, Y. L.; Hoops, G. C.; Coward, J. K., A comparison of proteins and peptides as substrates for microsomal and solubilized oligosaccharyltransferase. *Bioorg Med Chem* **1994,** *2* (11), 1133-41.

101. Pless, D. D.; Lennarz, W. J., Enzymatic conversion of proteins to glycoproteins. *Proc Natl Acad Sci U S A* **1977,** *74* (1), 134-8.

102. Elhammer, A. P.; Kezdy, F. J.; Kurosaka, A., The acceptor specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycoconj J* **1999,** *16* (2), 171-80.

103. Al Teneiji, A.; Bruun, T. U.; Sidky, S.; Cordeiro, D.; Cohn, R. D.; Mendoza-Londono, R.; Moharir, M.; Raiman, J.; Siriwardena, K.; Kyriakopoulou, L.; Mercimek-Mahmutoglu, S., Phenotypic and genotypic spectrum of congenital disorders of glycosylation type I and type II. *Mol Genet Metab* **2017,** *120* (3), 235-242.

104. Hennet, T.; Cabalzar, J., Congenital disorders of glycosylation: a concise chart of glycocalyx dysfunction. *Trends Biochem Sci* **2015,** *40* (7), 377-84.

105. Peanne, R.; de Lonlay, P.; Foulquier, F.; Kornak, U.; Lefeber, D. J.; Morava, E.; Perez, B.; Seta, N.; Thiel, C.; Van Schaftingen, E.; Matthijs, G.; Jaeken, J., Congenital disorders of glycosylation (CDG): Quo vadis? *Eur J Med Genet* **2018,** *61* (11), 643-663.

106. Freeze, H. H.; Westphal, V., Balancing N-linked glycosylation to avoid disease. *Biochimie* **2001,** *83* (8), 791-799.

107. Chacko, B. K.; Scott, D. W.; Chandler, R. T.; Patel, R. P., Endothelial surface N-glycans mediate monocyte adhesion and are targets for anti-inflammatory effects of peroxisome proliferator-activated receptor gamma ligands. *J Biol Chem* **2011,** *286* (44), 38738-38747.

108. Scott, D. W.; Vallejo, M. O.; Patel, R. P., Heterogenic endothelial responses to inflammation: role for differential N-glycosylation and vascular bed of origin. *J Am Heart Assoc* **2013,** *2* (4), e000263.

109. Wolfert, M. A.; Boons, G. J., Adaptive immune activation: glycosylation does matter. *Nat Chem Biol* **2013,** *9* (12), 776-84.

110. Schnaar, R. L., Glycobiology simplified: diverse roles of glycan recognition in inflammation. *J Leukoc Biol* **2016,** *99* (6), 825-38.

111. Giovannone, N.; Liang, J.; Antonopoulos, A.; Geddes Sweeney, J.; King, S. L.; Pochebit, S. M.; Bhattacharyya, N.; Lee, G. S.; Dell, A.; Widlund, H. R.; Haslam, S. M.; Dimitroff, C. J., Galectin-9 suppresses B cell receptor signaling and is regulated by I-branching of N-glycans. *Nat Commun* **2018,** *9* (1), 3287.

112. Moore, J. S.; Wu, X.; Kulhavy, R.; Tomana, M.; Novak, J.; Moldoveanu, Z.; Brown, R.; Goepfert, P. A.; Mestecky, J., Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals. *AIDS* **2005,** *19* (4), 381-9.

113. Tomana, M.; Schrohenloher, R. E.; Koopman, W. J.; Alarcon, G. S.; Paul, W. A., Abnormal glycosylation of serum IgG from patients with chronic inflammatory diseases. *Arthritis Rheum* **1988,** *31* (3), 333-8.

114. Ferreira, I. G.; Pucci, M.; Venturi, G.; Malagolini, N.; Chiricolo, M.; Dall'Olio, F., Glycosylation as a Main Regulator of Growth and Death Factor Receptors Signaling. *Int J Mol Sci* **2018,** *19* (2).

115. Nairn, A. V.; Aoki, K.; dela Rosa, M.; Porterfield, M.; Lim, J. M.; Kulik, M.; Pierce, J. M.; Wells, L.; Dalton, S.; Tiemeyer, M.; Moremen, K. W., Regulation of glycan structures in murine embryonic stem cells: combined transcript profiling of glycan-related genes and glycan structural analysis. *J Biol Chem* **2012,** *287* (45), 37835-56.

116. Schultz, M. J.; Holdbrooks, A. T.; Chakraborty, A.; Grizzle, W. E.; Landen, C. N.; Buchsbaum, D. J.; Conner, M. G.; Arend, R. C.; Yoon, K. J.; Klug, C. A.; Bullard, D. C.; Kesterson, R. A.; Oliver, P. G.; O'Connor, A. K.; Yoder, B. K.; Bellis, S. L., The Tumor-Associated Glycosyltransferase ST6Gal-I Regulates Stem Cell Transcription Factors and Confers a Cancer Stem Cell Phenotype. *Cancer Res* **2016,** *76* (13), 3978-88.

117. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer* **2015,** *15* (9), 540-555.

118. G. Davis, B., Recent developments in glycoconjugates. *Journal of the Chemical Society, Perkin Transactions 1* **1999,** (22), 3215-3237.

119. Seitz, O., Glycopeptide synthesis and the effects of glycosylation on protein structure and activity. *Chembiochem* **2000,** *1* (4), 214-46.

120. Boratynski, J.; Roy, R., High temperature conjugation of proteins with carbohydrates. *Glycoconj J* **1998,** *15* (2), 131-8.

121. Lampropoulou, M.; Misiakos, K.; Paravatou, M.; Mavridis, I. M.; Yannakopoulou, K., Synthesis of cyclodextrin derivatives with monosaccharides and their binding with ampicillin and selected lectins. *ARKIVOC* **2015,** *2015* (3), 232-243.

122. Yoshida, T., Glycamine formation via reductive amination of oligosaccharides with benzylamine. *Methods Enzymol* **1994,** *247*, 55-64.

123. Yuan, Y.; Chen, J.; Wan, Q.; Wilson, R. M.; Danishefsky, S. J., Toward fully synthetic, homogeneous glycoproteins: advances in chemical ligation. *Biopolymers* **2010,** *94* (4), 373-84.

124. Umekawa, M.; Huang, W.; Li, B.; Fujita, K.; Ashida, H.; Wang, L.-X.; Yamamoto, K., Mutants of Mucor hiemalis Endo-β-N-acetylglucosaminidase Show Enhanced Transglycosylation and Glycosynthase-like Activities*. *Journal of Biological Chemistry* **2008,** *283* (8), 4469-4479.

125. Wang, Z.; Chinoy, Z. S.; Ambre, S. G.; Peng, W.; McBride, R.; de Vries, R. P.; Glushka, J.; Paulson, J. C.; Boons, G. J., A general strategy for the chemoenzymatic synthesis of asymmetrically branched N-glycans. *Science* **2013,** *341* (6144), 379-83.

126. Koeller, K. M.; Wong, C.-H., Synthesis of Complex Carbohydrates and Glycoconjugates: Enzyme-Based and Programmable One-Pot Strategies. *Chemical Reviews* **2000,** *100* (12), 4465-4494.

127. Takegawa, K.; Tabuchi, M.; Yamaguchi, S.; Kondo, A.; Kato, I.; Iwahara, S., Synthesis of neoglycoproteins using oligosaccharide-transfer activity with endo-beta-N-acetylglucosaminidase. *J Biol Chem* **1995,** *270* (7), 3094-9.

128. Longo, M. A.; Combes, D., A novel chemoenzymatic glycosylation strategy: application to lysozyme modification. *FEBS Lett* **1995,** *375* (1-2), 63-6.

129. Witte, K.; Sears, P.; Martin, R.; Wong, C.-H., Enzymatic Glycoprotein Synthesis: Preparation of Ribonuclease Glycoforms via Enzymatic Glycopeptide Condensation and Glycosylation. *Journal of the American Chemical Society* **1997,** *119* (9), 2114-2118.

130. Grass, S.; Buscher, A. Z.; Swords, W. E.; Apicella, M. A.; Barenkamp, S. J.; Ozchlewski, N.; St Geme, J. W., 3rd, The Haemophilus influenzae HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol Microbiol* **2003,** *48* (3), 737-51.

131. Ahangama Liyanage, L.; Harris, M. S.; Cook, G. A., In Vitro Glycosylation of Membrane Proteins Using N-Glycosyltransferase. *ACS Omega* **2021,** *6* (18), 12133-12142.

132. Voruganti, S.; Kline, J. T.; Balch, M. J.; Rogers, J.; Matts, R. L.; Hartson, S. D., Proteomic Profiling of Hsp90 Inhibitors. *Methods Mol Biol* **2018,** *1709*, 139-162.

VITA

Michael Shaun Harris

Candidate for the Degree of

Doctor of Philosophy

Dissertation: EXPRESSION AND CHARACTERIZATION OF THE MEMBRANE PROTEINS TCR α WT AND γ-SARCOGLYCAN WITH THE QUANTIFICATION OF *IN VITRO* GLYCOSYLATED γ-SARCOGLYCAN

Major Field: Chemistry

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Chemistry at Oklahoma State University, Stillwater, Oklahoma in May, 2022.

Completed the requirements for the Bachelor of Science in Chemistry at Missouri State University, Springfield, Missouri in 2017.

Experience:

Graduate Assistant, Department of Chemistry, Oklahoma State University (2017-2022)

- Research fields- Protein Biochemistry, Molecular cloning, Recombinant expression and purification of membrane proteins, Solution-state Nuclear Magnetic Resonance Spectroscopy
- Teaching-Organic Chemistry Laboratories and General Chemistry I laboratories

Professional Memberships:

American Chemical Society