EXPRESSION, CHARACTERIZATION, AND *IN VITRO* GLYCOSYLATION OF ECTODOMAINS FROM THE MEMBRANE PROTEIN SYNDECAN-1

By

AUSTIN RAY ANDERSON

Bachelor of Science in Chemistry
Northwestern Oklahoma State University
Alva, Oklahoma
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
December, 2022

EXPRESSION, CHARACTERIZATION, AND *IN VITRO* GLYCOSYLATION OF ECTODOMAINS FROM THE MEMBRANE PROTEIN SYNDECAN-1

Dissertation Approved:

Dr. Gabriel A. Cook

Dissertation Adviser

Dr. Barry K. Lavine

Dr. Sadagopan Krishnan

Dr. Ziad El Rassi

Dr. Donghua Zhou

ACKNOWLEDGEMENTS

I would to take a moment to acknowledge and thank the people whom have helped me succeed in all my efforts over the past decade of my collegiate education. First I have to give thanks to both my graduate and undergraduate advisors Dr. Gabriel Cook and Dr. Jason Wickham. These two men have been a massive influence in my life. In my undergraduate studies, I was very fortunate to have Dr. Jason Wickham for my advisor. He made an excellent mentor and, without him, I likely would have never applied for graduate studies. After convincing me to continue in graduate studies, Dr. Wickham told me something that really stuck. He said, "it is easy to find an interest in any field of research, the most important thing is to find a good mentor that will take you under their wing." I took this advice to heart and could not be happier with having Dr. Gabriel Cook as my graduate advisor and mentor. Dr. Cook was happy to welcome me into his lab although I had zero prior knowledge or experience with his field of research. He has always been willing and ready to help with anything that has come up inside or outside the lab during my time here at Oklahoma State. I also need to thank Dr. Steven Hartson and Janet Rogers with the Oklahoma State Biochemistry and Molecular Biology Recombinant DNA/Protein Core Facility. Steve and Janet have been essential to the research efforts of our lab and have always been eager to help with any problems I may bring to their tables.

I would like to thank all of the past and present research members in the Cook Laboratory both graduate and undergraduate for making the past 5 years productive and enjoyable in my growth as a research scientist. I would specifically like to thank Dr. Leshani Liyanage for her help and expertise in the lab. In her years here she was proved herself time and again to be an invaluable member of our team.

Lastly, I'd like to thank my parents, grandparents, and the rest of my amazing family and friends; without whom I would not be where I am today.

Name: AUSTIN RAY ANDERSON

Date of Degree: DECEMBER, 2022

Title of Study: EXPRESSION, CHARACTERIZATION, AND IN VITRO GLYCOSYLATION OF ECTODOMAINS FROM THE MEMBRANE PROTEIN SYNDECAN-1

Major Field: Chemistry

Abstract: Membrane proteins are a key component of numerous biological processes such as signal transduction, transportation, and various cell-cell interactions. Nearly one-third of all genes encode for membrane proteins and around one-half of all drug targets are membrane proteins. Despite the significance of these proteins, a relatively small amount of membrane protein structures have been determined. Membrane proteins are difficult to express in high quantities and contain highly hydrophobic domains that generally require solubilization strategies using detergents or lipids. Post-translational modifications further diversify these proteins. Many membrane proteins are glycoproteins with one or more sugar attachment sites. Glycosylation, like other types of post-translational protein modifications, can have a profound effect on the structure and function of these proteins. Syndecan-1 (SDC-1), an integral membrane glycoprotein, is involved in the body's innate immune response system. When cells are under stress due to infection or trauma, the ectodomains of SDC-1 are cleaved and released into the bloodstream as a way to induce formation of a chemokine gradient that subsequently recruits leukocytes to the affected area. This protein is normally highly glycosylated in nature with one N-linked glycosylation site and five O-linked glycosylation sites. While many important biological functions of SDC-1 have been identified, its structure and specific interactions have not been elucidated. The research efforts in this study have developed a method to reliably express and isolate the recombinant core ectodomain of human syndecan-1 from *E. coli* cells. We have also laid the foundation for future studies on the structure and dynamics of these ectodomains using 2D and 3D NMR spectroscopy. Using novel methods of *in vitro* glycosylation, we were successful in attaching a glycan to the syndecan-1 core protein. Solution NMR spectroscopy requires highly homogenous sample conditions that are impossible to attain from glycoproteins that have been isolated from mammalian cells. This work with *in vitro* glycosylation will provide the highly homogenous sample conditions required to study specific effects of glycosylation on proteins with intact glycans attached to the core protein.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

RESEARCH OVERVIEW

This chapter provides an abridged overview of the methods background and research objectives of this dissertation. The research objectives **(Figure 1.1)** were designed to provide a pathway for gaining new insight into the effects of post-translational modifications (PTMs) on protein structure and dynamics. Specifically, we are interested in looking into the effects of glycosylation on the ectodomain of Syndecan-1 (SDC-1). The first section is dedicated to the creation of cell lines that are equipped to overexpress our two recombinant SDC-1 proteins. The second section describes the process to express and isolate the two SDC-1 ectodomains. The final section reviews the current work with *in vitro* glycosylation and NMR spectroscopic studies. This dissertation demonstrates the first successfully developed method of expressing and isolating the core ectodomain of Syndecan-1 from recombinant prokaryotic cells. This allowed us to generate the first resolved solution NMR spectra of the Syndecan-1 ectodomain. It also demonstrates the first instance of using *in vitro* glycosylation to attach an N-linked glycan to the core SDC-1 protein.

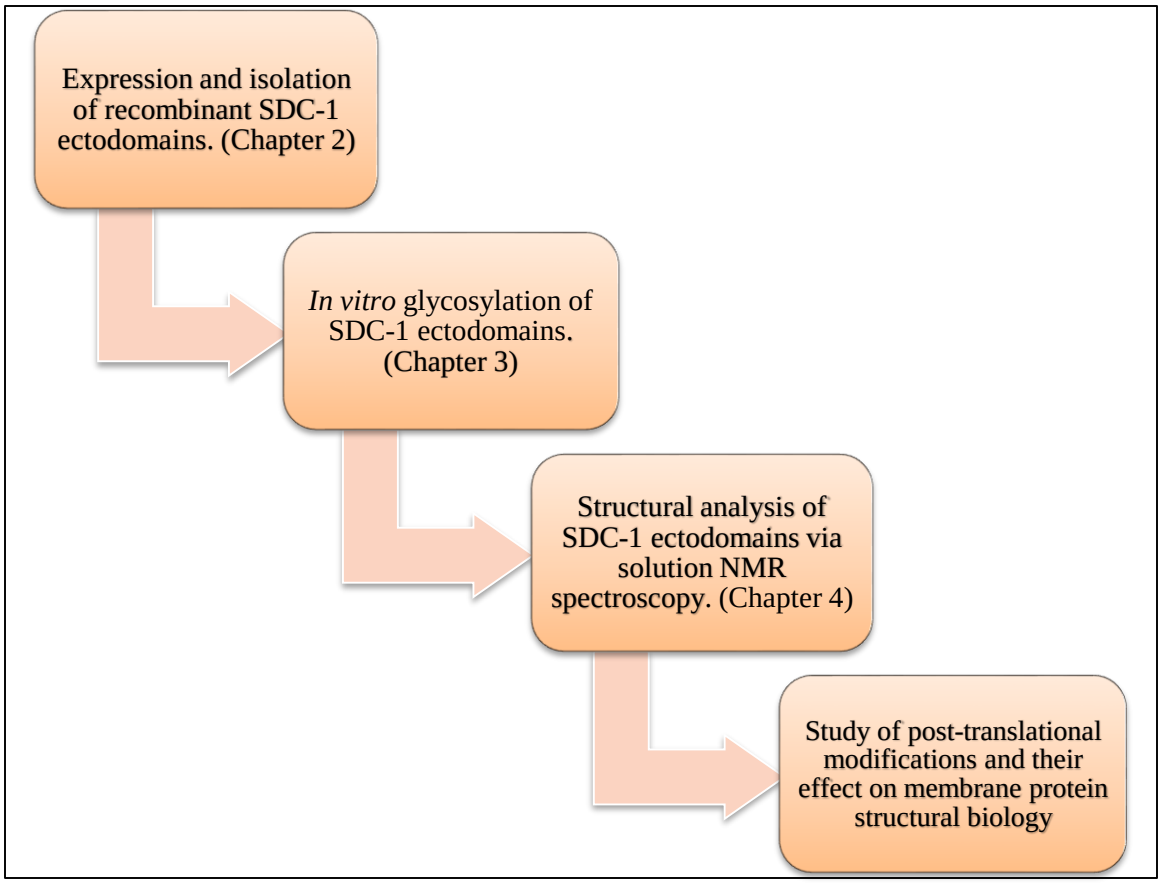

Figure 1.1: **Flow of dissertation research.**

1.1 Introduction

Integral membrane proteins are proteins that possess a transmembrane domain that is directly incorporated into or associated with lipid membranes. This class of proteins have many diverse functionalities that are essential for the biological cells. This includes transportation of ions and other solutes into and out of the cell, mediating communication between the cell and its environment, and acting as membrane-embedded enzymes for chemical catalysis.[1] Integral membrane proteins have been estimated to account for 20-30% of all genes expressed in most genomes and over 60% of all drug targets are membrane proteins.[2, 3] The amphiphilic nature of these proteins makes them exceedingly difficult to study due to problems that arise with overexpression, purification, and structural determination. Overexpressing membrane proteins in large quantities faces a multitude of hindrances including protein aggregation, misfolding and

premature cell death.[4] These issues are just a few of the reasons that make producing even the milligram amounts of protein needed for characterization studies an immense challenge. Protein aggregation and misfolding must also be limited to attain any useful structural information.

The very first membrane protein structure was determined in 1985 by X-ray crystallography[5]. According to *mpstruc*, a database of membrane proteins of known 3D structure, more than 1500 structures have been determined as of 2022. While this may seem significant, this accounts for less than 1% of the 196,000 structures that have been elucidated and posted to the RCSB Protein Data Bank as of 2022. Post-translational modifications such as glycosylation serve to further diversify the membrane proteome. In the 1960's, it was found that naturally occurring carbohydrates were commonly conjugated to proteins and by the 1970's it was evident that glycoproteins played a crucial role in biology.[6] While the area of membrane protein research is actively growing, it is imperative that advancements are made to discover novel and less problematic methods to study the structure and dynamics of this class of proteins. It is also essential for us to begin to understand exactly how glycosylation effects these proteins. These studies could be invaluable in future development of therapeutic drugs to combat various diseases.

The aim of this chapter is to give an overview for membrane proteins as well as the various concepts and methods we have used in our studies in the following chapters.

1.1.1 Membrane Proteins

Membrane proteins are a class of proteins that are incorporated into or associated with biological membranes. This class of proteins serve many diverse biological functions including transportation, signal transduction, and cell-anchoring. The three main types of membrane proteins are integral membrane proteins, peripheral membrane proteins, and lipid-linked proteins **(Figure 1.2).**[7] Integral membrane proteins have one or more transmembrane domains that are

directly incorporated into the hydrophobic portion of the lipid membrane. These transmembrane domains can span partially across the lipid bilayer in the case of an integral monotopic protein. A single integral protein may also make one or more passes completely through the lipid bilayer. Integral membrane proteins may also associate together in multiple subunits spanning the lipid membrane. Integral membrane proteins are tightly bound to membranes by hydrophobic forces and can only be separated from these membranes using agents such as detergents and organic solvents that disrupt the membrane. Peripheral membrane proteins, on the other hand, are a class of proteins that can be dissociated from the cell membrane by relatively mild procedures that leave the membrane intact. These proteins are found bound to the polar head groups of the lipid membrane or bound to an intracellular or extracellular portion of an integral membrane protein. The binding of peripheral membrane proteins often occurs through electrostatic interactions or hydrogen bonding. The final class of membrane proteins, lipid-linked proteins, result from a lipid bound through a covalent bond to a protein. These lipids are covalently-linked to a hydrophilic protein and serve to anchor the protein to the lipid membrane. [7] Together, membrane proteins allow cells and membrane-bound organelles to effectively interact and communicate with their respective environments. Without membrane proteins, many biological processes would be impossible.

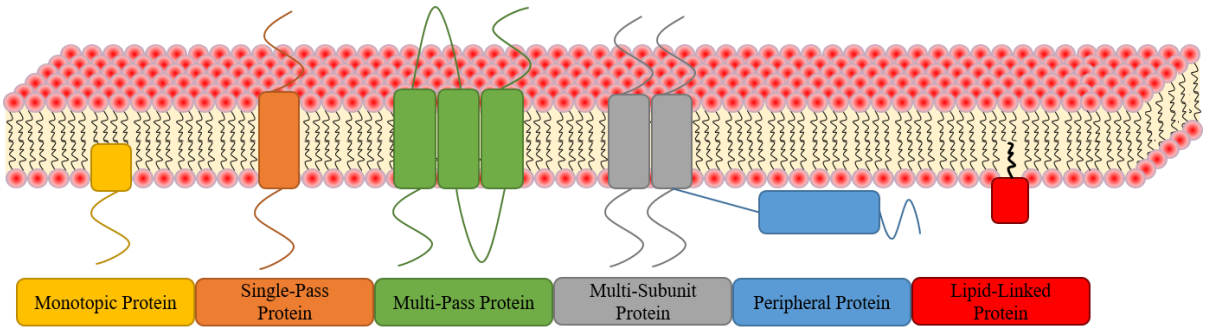

Figure 1.2: Membrane Proteins. Different types of membrane proteins and their association with the lipid bilayer.

Research in the membrane protein field has been steadily expanding to advance our understanding and find novel ways of studying these profoundly important classes of proteins. To

aid in these studies, this dissertation will contribute new methods for expressing, purifying, and characterizing the ectodomain of the integral membrane protein, Syndecan-1. These developments will lead to further knowledge on the communication and signaling aspect of membrane protein function. Further understanding of the signaling and binding pathways of transmembrane proteins will provide information advantageous to the development of novel therapeutic drugs and disease treatment pathways.

1.1.2 N-linked and O-linked Glycosylation

Glycosylation is one of the many post-translational modifications that can be used to further diversify the proteome. The process of glycosylation results in the attachment of a carbohydrate or glycan through a covalent bond to specific residues of a glycoprotein. Glycosylation has been proven to drastically alter the structure, dynamics, and specific interactions of various proteins.[8, 9] Glycosylation also often serves to regulate enzymatic activity and mechanisms of various proteases. [10, 11] Many membrane proteins are also glycoproteins with one or many sugars attached to the protein.[7] Therefore, when studying different membrane proteins, it is essential to also examine the effects of glycosylation on these proteins.

The two main types of glycosylation that can modify proteins are N-linked and O-linked glycosylation. There are also the glycosylation categories of C-mannosylation and glycosylphosphatidylinositol (GPI) anchors that should also be mentioned.[12] N-linked glycosylation is one of the most common types of PTMs that occurs in eukaryotic cells. This modification involves attachment of a glycan to the sidechain amide of an asparagine (N) residue and only occurs on asparagine residues within the amino acid consensus sequence of Asn-Xaa-Ser/Thr (N-X-S/T), where X is any amino acid other than proline followed by a serine or threonine (S/T). N-linked glycosylation can be catalyzed in eukaryotes by specific transferase enzymes. One such enzyme used in this dissertation to perform glycosylation is N-

glycosyltransferase. [13] O-linked glycosylation, most commonly occurring in mammals as an oligosaccharide attached to a serine or threonine through an N-Acetylgalactosamine (GalNAc) residue, is more complex with a larger variety of attachment sites. In addition, the likelihood of O-linked glycosylation depends on a variety of outside factors. O-GalNAc glycosylation often occurs on proteins as a reversible regulatory modification.[14] As depicted in **Figure 1.3** some polypeptides may have a mix of both N-linked and O-linked glycans attached. Traditionally, glycosylation has been considered a post-translational modification that is exclusive to only eukaryotes and archaea, however, it has recently been shown that a species of epsilon-proteobacteria exhibits N-linked glycans in its proteome.[15] This further cements the concept that glycosylation is an essential modification to proteins in all forms of life.

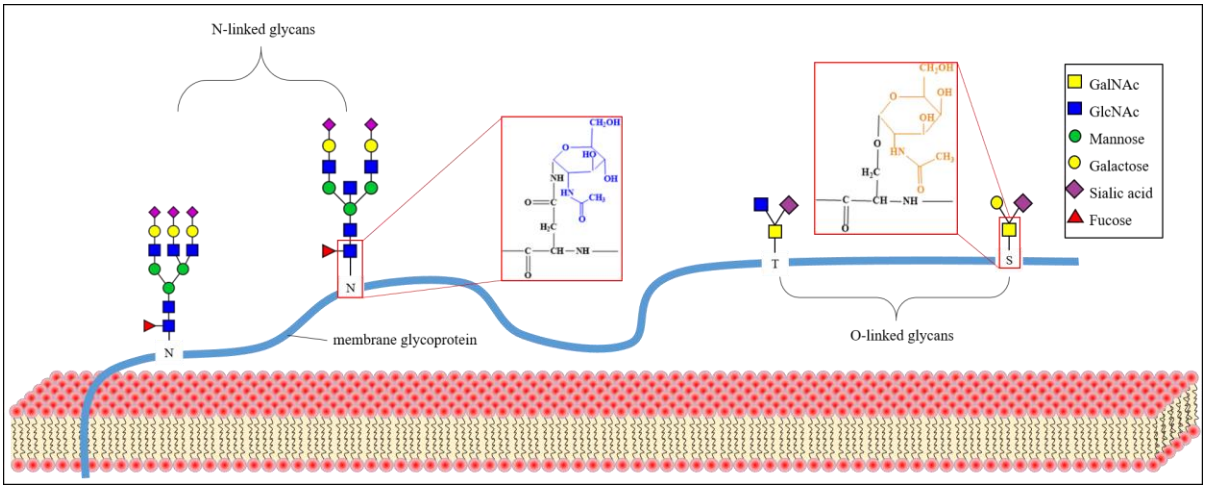

Figure 1.3: Integral Membrane Glycoproteins. Membrane protein with covalent attachment of glycans to the side chains of asparagine residues (N-linked) or serine/threonine residues (O-linked)

My work for this dissertation focused on the effects of N-linked glycosylation on the SDC-1 membrane glycoprotein. We were the first group to perform N-linked glycosylation on the ectodomain of SDC-1 *in vitro*. The efficiency of our glycosylation methodology on SDC-1 will be determined through liquid chromatography with tandem mass spectrometry (LC-MS/MS). From our initial work with *in vitro* glycosylation, we will further develop and refine our methodology to determine the ideal conditions for glycosylation efficiency. Forthcoming

experiments will allow the attachment of larger and more complex sugars of known structure to the core SDC-1 protein. This will provide highly homogenous glycoprotein samples with intact glycans attached to the core protein. This work will be essential for future structural and dynamics studies of glycoproteins.

1.1.3 Solution-State NMR Spectroscopy

Solution NMR can be used to study membrane protein structures, protein-protein interactions, protein-ligand interactions, protein-nucleic acid interactions, and protein dynamics.[16] This powerful analytical technique can also be used to examine the changes induced in the aforementioned properties by various post-translational modifications including glycosylation.[17] The concept of NMR spectroscopy relies on the properties of atoms with a non-zero nuclear spin state when these atoms are subjected to a magnetic field. When placed in a magnetic field, atoms with a non-zero nuclear spin state make transitions between different energy levels at specific resonance frequencies. These resonance frequencies exhibit slight variations or chemical shifts depending on the chemical environment around the individual atom. Therefore, individual atoms of the same type can be differentiated from one another and studied based on the unique changes in their respective resonance frequencies.[18]

By incorporating NMR active nuclei into a protein it is possible to study the orientation of individual atoms in the protein in 3-dimensional space. Atoms that possess an even number of both protons and neutrons, such as ^{4}He, ^{12}C, ^{16}O and ^{32}S have a spin state of zero and exhibit no changes when subjected to a magnetic field. Nuclei for which the number of protons plus the number of neutrons is an odd number result in a half-integer spin state of (1/2, 3/2, 5/2, etc.). Nuclei such as ^{2}H, ^{6}Li, ^{10}B, and ^{14}N, where the number of protons and neutrons are both odd numbers, possess a whole-integer spin state (1, 2, 3, etc.). Any atom with a spin state greater than ½ possesses an asymmetric nuclear charge distribution this gives them an electric quadrupole

moment in addition to their magnetic dipole moment. This causes atoms with greater than 1/2 spin to produce very broad NMR signals or no signals at all. Therefore, the most useful atoms for NMR studies are those with a spin state of 1/2. Atoms such as ^{1}H, ^{13}C, and ^{15}N possess nuclear spins of 1/2 and are the most essential isotopes for solution NMR studies on proteins.[19]

Depending on the length of a protein, one can examine different atoms and use various NMR experiments to correlate the chemical shifts of specific atoms to specific amino acids. For smaller proteins, this can usually be accomplished by examining ^{1}H-^{1}H correlations in two-dimensional (2-D) homonuclear experiments to assign chemical shifts to individual amino acid residues. As proteins become larger, many of these chemical shifts may begin to overlap or give increased line width in the spectra. Consequently, heteronuclear NMR experiments specifically examining correlations between ^{1}H, ^{13}C, and ^{15}N must be employed to study these larger proteins.[18, 19] A multitude of heteronuclear experiments are often employed when studying larger proteins, with each different experiment essentially giving researchers a few pieces of a spectral jigsaw puzzle. When enough spectral data has been accumulated, one can begin to mend together the different puzzle pieces to determine the orientation of each individual amino acid in a protein sequence in three-dimensional (3D) space.

The most common 2D heteronuclear experiment in protein NMR is ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectroscopy. This experiment uses the J-coupling between a ^{1}H proton directly attached to a ^{15}N nitrogen atom. Magnetization is first transferred from the proton to the nitrogen. The chemical shift is then evolved on the nitrogen and magnetization is subsequently transferred back to the proton for detection. This allows the plotting of chemical shifts of individual ^{15}N atoms and their corresponding protons together on a 2D spectrum, thus allowing data to be gathered on the individual backbone amides from each amino acid. The ^{1}H-^{15}N correlations from the side chains of tryptophan, asparagine, and glutamine are also visible in the resultant spectra.

While a vast amount of information can be extrapolated from the various 2D NMR experiments available, it is also necessary to employ 3D experiments designed to look into correlations between different ^{1}H, ^{13}C, and ^{15}N atoms to aid in structural elucidation studies, particularly as protein size increases.[20] First described in 1990, the four most widely-employed 3D heteronuclear experiments for backbone amino acid assignment are HNCA, HN(CO)CA, HNCO, and HN(CA)CO.[21] In an HNCA experiment, magnetization is initially passed from the ^{1}H proton of the backbone amide to the ^{15}N. Magnetization is then transferred from the ^{15}N to both the intraresidue alpha ^{13}C (C_{α}) and the C_{α} of the preceding residue. A 3D spectrum is then generated with two distinct ^{13}C chemical shifts for each ^{1}H-^{15}N chemical shift. The peak from the intraresidue C_{α} will be higher in intensity than the peak for the C_{α} on the preceding residue due to its closer proximity and stronger coupling to the ^{15}N. The HN(CO)CA experiment is complimentary to the HNCA experiment. In an HN(CO)CA experiment, magnetization is again transferred starting from the proton to the nitrogen of the backbone amide. Magnetization is then transferred from the nitrogen to the ^{13}C carbonyl carbon (CO) of the previous residue and then to the C_{α} of the same residue. Here only the chemical shifts for the ^{1}H, ^{15}N and C_{α} are evolved. When comparing HNCA and HN(CO)CA spectra, both spectra will exhibit the same peak correlating to the C_{α} of the previous residue. However, in an HN(CO)CA experiment, the stronger of the two HNCA C_{α} peaks is no longer visible. Visualizations of the magnetization transfer in HNCA and HN(CO)CA can be seen in **Figure 1.4A and Figure 1.4B** respectively.

Like HNCA and HN(CO)CA experiments, HNCO and HN(CA)CO experiments are complimentary towards one another. In an HNCO experiment, magnetization is transferred directly from the proton, to the nitrogen, and then to the immediately neighboring carbonyl carbon on the previous residue. This results in a single peak on the spectra correlating the ^{1}H-^{15}N to the ^{13}C carbonyl on the previous amino acid residue. Due to the nature of this magnetization transfer, the amides on the side chain of asparagine and glutamine residues are also visible. In an

HN(CA)CO experiment, magnetization is transferred from the backbone ^{1}H-^{15}N to both the C_{α} on the previous residue and the intraresidue C_{α}. Magnetization is then transferred from the C_{α} of each residue to the CO of that residue. Chemical shifts are then evolved for the ^{1}H-^{15}N and CO. This will generate a spectrum with two CO chemical shifts for every ^{1}H-^{15}N chemical shift, with the weaker of the two shifts correlating to the CO of the preceding residue. This weak CO shift will also be visible in the HNCO spectrum. The difference in strength between these two CO chemical shifts can be explained by the increased efficiency of magnetization transfer from the ^{15}N to the C_{α} to which it is directly connected. A visualization of the magnetization transfers for HNCO and HN(CA)CO can be seen respectively in **Figure 1.4C and Figure 1.4D.**

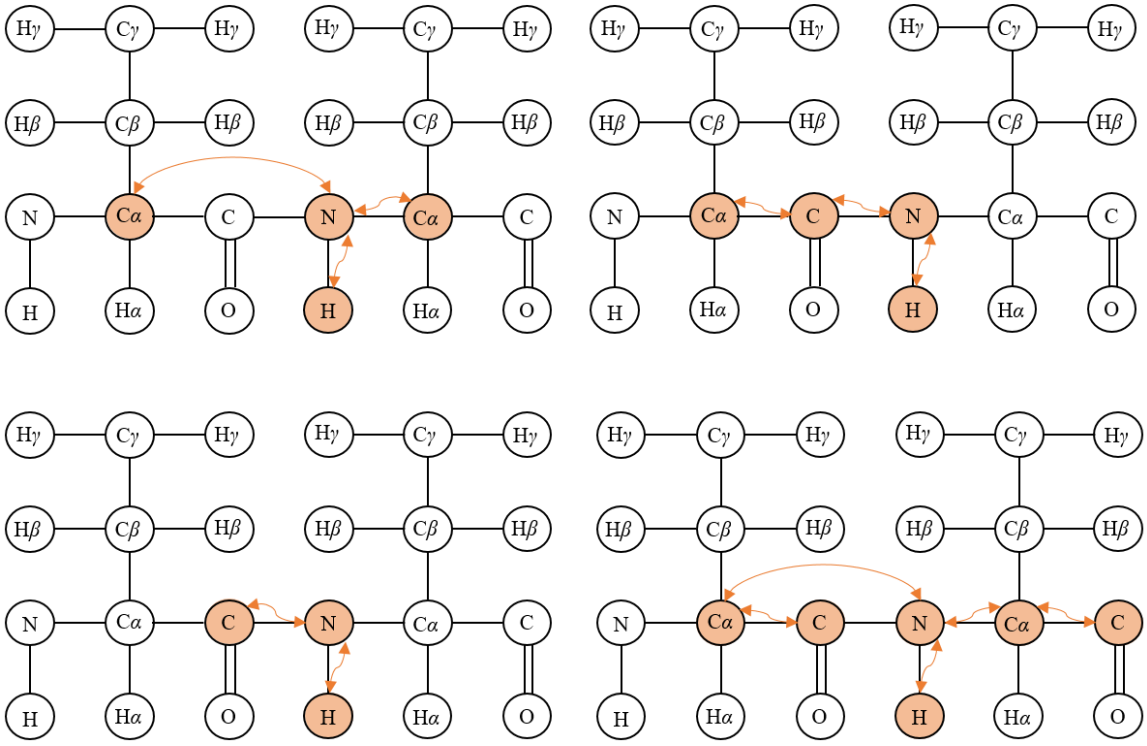

Figure 1.4: 3-D NMR Backbone Assignment Experiments. Visual representations of the magnetization transfer between ^{1}H, ^{15}N, and ^{13}C in four different 3D NMR experiments.

Our current work in this dissertation with ^{1}H-^{15}N HSQC experimentation has laid the foundation for future 3-D experiments such as the ones described in this chapter. This will lead to backbone assignment and future elucidation of secondary protein structure for the SDC-1 ectodomains. Upon elucidation of the core protein structure, it will then be possible to examine any changes as a response to glycosylation. This will elicit new insight into the role glycosylation plays in protein structure, function, and dynamics.

1.1.4 Expression and Purification of Recombinant Proteins

Studies on the structure and dynamics of proteins require the ability to produce and purify a protein of interest in large enough quantities to run some of the various experiments that were touched on in the previous section. When starting a project that requires recombinantly produced protein, it is important to select an optimal host that is uniquely suited to expressing the target protein for the desired application. Many proteins and enzymes require specific cellular mechanisms and post-translational modifications for proper folding and activity. If the recombinant protein is to be used in therapeutics or drug discovery, it often needs to be in the native form.[22] To this end, many eukaryotic hosts such as yeasts, fungi, insects, plants and mammalian cell lines have been utilized for recombinant therapeutic protein production and drug development.[23-27] For other research purposes, it may instead be necessary to produce relatively large quantities of the core recombinant protein sequence. This is most often accomplished by overexpressing recombinant proteins in *Escherichia coli*. A bacterial host is uniquely suited to producing core recombinant protein due to the fact that it lacks the necessary cellular mechanisms for most post-translational modifications. The characteristics of *E. coli* cells have made the bacteria a widely-used host for recombinant protein production. Ease of genetic modification and vast availability of genetic modification tools combined with *E. coli*'s exponential growth kinetics and simplistic cultivation techniques make the bacteria an ideal host for many recombinant protein studies.[28]

Recombinant protein expression in *E. coli* is accomplished through the utilization of a plasmid expression vector containing the DNA sequence that can be translated into the target protein. Expression vector options are vast and should be carefully selected to optimize expression and subsequent purification of the desired protein target.[29] Most popular expression vector options code for various fusion tags attached to the target protein rather than directly expressing the protein of interest. Commonly used fusion tags such as Glutathione S-transferases (GST), Maltose-binding protein (MBP), polyhistidine (His_6), and N-utilization substance protein A (NusA) can serve to aid in purification, enhance solubility, or aid in disulfide-bond formation.[30] For expression of recombinant proteins with highly hydrophobic transmembrane domains, attempted overexpression of these proteins in *E. coli* cells tends to lead to instability of the cell membrane and premature cell death, resulting in minimal expression levels of the desired protein. For these proteins, it is possible to use an expression vector that will fuse the target to a Tryptophan operon (Trp) leader protein sequence. This Trp leader fusion protein will sequester the insoluble target protein into readily formed inclusion bodies during overexpression and can vastly enhance expression levels.[31] Other fusion tags such as the consensus sequences recognized by Factor Xa or Tobacco Etch Virus (TEV) protease may also be incorporated into various expression vectors to provide an enzymatic pathway for the cleavage of fusion partners from the target protein during the purification process.[30] Aside from the described fusion partners, plasmid DNA also provides resistance to specific antibiotics, thus allowing the incorporation of these antibiotics in the cell culture medium and helping to ensure only the bacteria containing the plasmid DNA are able to grow in the media.[32]

Lastly, the strain of *E.coli* cells used for expression is also highly important for recombinant protein production. The most frequently used strains for recombinant protein production are BL21 and its various derivatives. These strains have been developed to be deficient of *ompT* and *lon* proteases, thus providing increased expression and stability of recombinant proteins.[33]

Various other strains have additional plasmid DNA with enhancements designed to aid in expression of genes that require a high frequency of rare tRNA sequences.[34] Like some of the various expression vectors that the recombinant protein DNA can be incorporated into, some bacterial strains also have genetics designed to enhance disulfide-bond formation and aid in the expression of toxic proteins.[35]

Many host strains can have a combination of the various properties mentioned above and have proven essential for recombinant protein expression. However, proper vector and fusion protein selection is equally as vital. For example, BSPH1, a human sperm-binding protein containing four disulfide-bonds was completely insoluble when expressed in BL21 (DE3) cells. When the same protein was expressed in Origami B (DE3)pLysS cells, solubility was increased, yet the protein was still forming misfolded aggregates. Various combinations of expression strains and vectors were tested before settling on a combination of Origami B (DE3)pLysS cells and an expression vector coding for a His_6-thioredoxin-tagged fusion partner. This combination generated a soluble, active, and easily purified form of the BSPH1 protein.[36] This highlights the importance of both expression plasmid and expression strain selection for the production of various recombinant protein in a protein-dependent approach.

Upon expression of a recombinant protein of interest, extraction and purification of the protein is necessary before moving on to further studies. If proper planning is used when designing the protein expression system, the subsequent isolation of the protein of interest can be done in a relatively simplistic manner that is efficient and yields large quantities of high-quality recombinant protein. In Chapter 2 of this dissertation, we will discuss the steps taken to generate the cells lines that produce recombinant SDC-1 ectodomains. We will then discuss the methodology to express and purify both the SDC-1 ectodomains and TEV protease. We will also describe the process used to subsequently cleave the SDC-1 fusion proteins with TEV protease.

1.1.5 Top-Down and Bottom-Up Protein Mass Spectrometry

The study of proteomics, or large-scale analysis of proteins, has been on the rise for the past several decades. Nearing the 2003 completion of the human genome project, there was a realization that the final product of any given gene is inherently much more complex than the gene itself.[37] This is in part due to the various post-translational modifications and genetic variations that can happen to any given protein. These changes are not seen at the gene-level, but may completely change a protein's function in the biological system.[38] The disconnect found between gene expression versus protein expression and function forced the scientific community to steer research toward the direct study of protein structure, function, dynamics, and interactions. This was likely a deciding factor for the selection of the 2002 Nobel prize winners for chemistry. The winners, Dr. John B. Fenn, Dr. Koichi Tanaka, and Dr. Kurt Wüthrich were far ahead of the curve. By the 1980s, they had already developed the methods required to directly study proteins. In 2002, ½ of the Nobel prize in Chemistry was awarded as ¼ jointly to Dr. Fenn and Dr. Tanaka for their development of ionization methods that allowed mass spectrometric analysis of biological macromolecules. The work by Dr. Tanaka led to the development of the matrix-assisted laser desorption/ionization (MALDI) technique, whereas Dr. Fenn developed electrospray ionization (ESI).[39] Both of these ionization techniques played an essential role in the research included in this dissertation. The other half of the 2002 Nobel prize was awarded to Dr. Kurt Wüthrich for his development of 3D heteronuclear solution NMR spectroscopy techniques used for determining the structure and dynamics of biological macromolecules.[39] The work done by Dr. Wüthrich with 3D NMR has likewise been crucial to this dissertation as described in the previous section.

First developed over 30 years ago, MALDI and ESI are both soft ionization techniques that, to this day, remain the two most important techniques for ionizing biomolecules. These "soft" ionization techniques allow the conversion of large biomolecules to gaseous ions without

fragmenting the molecules into various pieces, thus allowing mass spectroscopic analysis on intact proteins and peptides. MALDI works by embedding a target analyte into an acid matrix that readily absorbs UV light. A short, high energy laser pulse is then directed toward the matrix, consequently vaporizing and ionizing both the matrix and analyte. These ions are then directed by an electric field into the mass spectrometer[40] In ESI, the analyte solution is subjected to a high voltage electric field while passing through a capillary. The solution then exits the capillary in the form of highly-charged droplets. As the solvent droplets evaporate and shrink, the charge density of the ions within the droplet increases. The size of the droplets eventually reaches a critical point where the electric field strength within the droplet is strong enough to eject charged analyte ions at the surface of the droplet into the gas phase. These analyte ions are then sampled by an inlet and accelerated into the mass spectrometer.[41] While electrospray is used more often due to its ability to quantify an analyte via coupling to liquid chromatography (LC), MALDI still serves as arguably the best ionization method for studying intact membrane proteins.[42]

The study of protein via mass spectrometry comes in two main forms, top-down and bottom-up analysis. Top-down analysis is simply the analysis of intact proteins. After the full protein has been ionized and entered into the mass spectrometer, the mass-to-charge (*m/z*) ratio can then be directly analyzed, or the ions could be subjected to various fragmentation techniques in tandem mass spectrometry (MS/MS) where the resultant fragments are then studied.[43] Bottom-up analysis requires the protein to first be digested into peptides either enzymatically or chemically. These peptides are then ionized and resultant *m/z*'s are once again either directly examined or fragmentation is employed.[44]

When two or more mass analyzers are connected in sequence, it is possible to use MS/MS fragmentation to reveal more in-depth information about the analyte in question. An initial mass/charge analysis or ion isolation can be performed on parent ions in the first mass analyzer. The parent ions can then be fragmented in a collision cell and the resultant fragments analyzed in

the second mass analyzer.[45] For peptides, fragmentation patterns depend on the method of fragmentation, however, most fragmentation methods utilized such as collision-induced-dissociation (CID) and higher-energy collision dissociation (HCD) tend to fragment peptides and proteins around the peptide (amide) bonds.[46] For MS/MS analysis, it is possible to run the spectrometer in either data-dependent acquisition (DDA), or data-independent acquisition (DIA) mode. In a DIA, all peptide *m/z*'s that are identified in the first stage are subsequently fragmented during the second stage of MS/MS.[47] This is most useful for identifying all peptides that are in a sample as it does not require prior knowledge of specific peptide *m/z*'s. The downfall of DIA is that rare peptides are much less likely to be detected and therefore difficult to quantify. Using DDA, it is possible to scan for and fragment peptides in extremely specific *m/z* windows, thus allowing quantification of any specific peptide regardless of rarity.[48] Thus, bottom-up DDA MS/MS experiments such as the one depicted in **Figure 1.5** allow the quantification of any peptide so long as the *m/z* of the peptide is known beforehand.

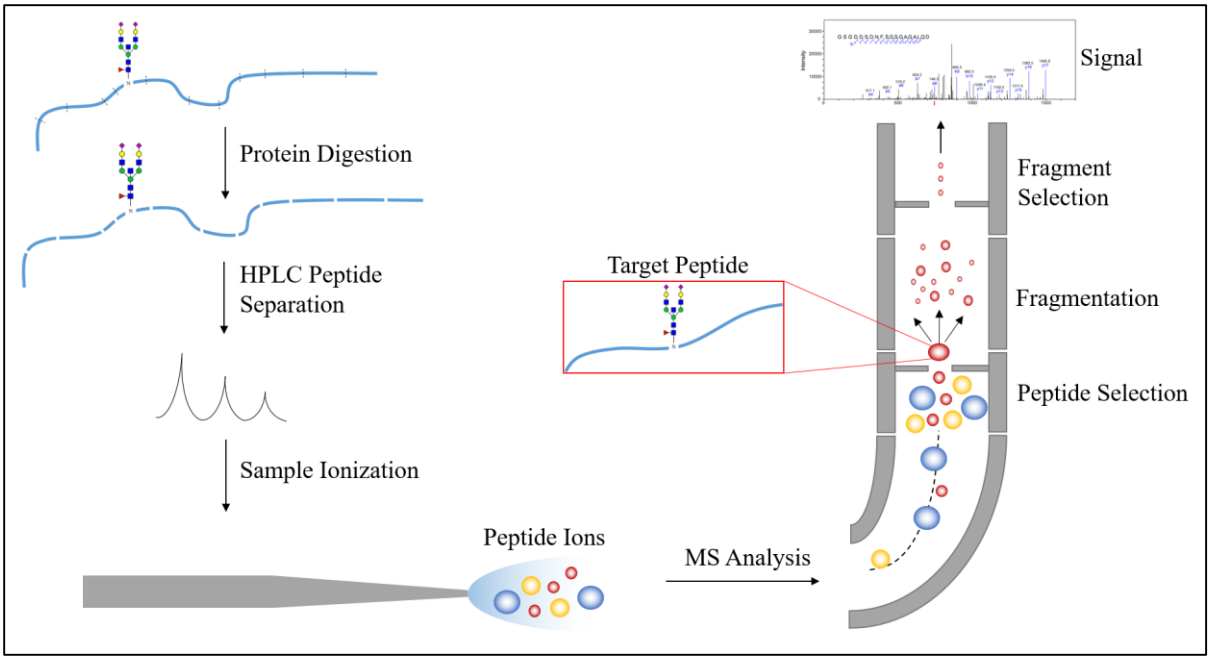

Figure 1.5: DDA MS/MS Experiment Flowchart. Experiment used to quantify the amount of glycosylated protein present in a protein sample. After the protein is enzymatically digested, the peptides are separated by HPLC, ionized, and injected into the mass spectrometer. The different mass analyzers then work in conjunction to separate a peptide of select *m/z*, fragment that peptide, then subsequently analyze the fragments generated.

In this dissertation, we have employed a top-down non-fragmenting methodology to confirm the molecular weights of our isolated SDC-1 ectodomains before and after cleavage. To this end, we used a combination of both MALDI and ESI. For our *in vitro* glysoylation studies, we utilized bottom-up LC – ESI – MS/MS methods with DDA to detect glycan attachment to the protein. A future DIA analysis will allow us to quantify the glycosylation as well.

1.1.5 Summary and Outlook

Membrane proteins are highly important in biological processes, with over 20% of all genes encoding membrane proteins. This class of protein is also the most prominent target for drugs and therapeutics with over 60% of all drugs on the current market targeting membrane proteins. This being said, membrane proteins make up less than 1% of all protein structures that have been elucidated to date. The material in this chapter highlights the importance of membrane protein

research and the necessity of gaining novel insight into their structures and functions. Membrane proteins can be near infinitely diversified by various post-translational modifications such as N-linked glycosylation. Consequently, we must gain a fundamental understanding of the different effects that glycosylation has on these proteins. The long-term goals of this project will be to explore new therapeutic pathways to counter disease and to develop novel methods of disease detection.

The following chapters will discuss the methodology used to design a bacterial expression system for recombinant Syndecan-1 ectodomains followed by a description of the procedures used to express and purify the SDC-1 ectodomains from these cells (Chapter 2). In chapter 3, I will discuss the technique used to achieve glycosylation of SDC-1 *in vitro* and how this can be quantified using LC-MS/MS analysis. Lastly, I will show our current progress made with NMR structural elucidation and I will elaborate on our future plans to analyze the structural effects on SDC-1 as a response to glycosylation (Chapters 4 and 5).

Abbreviations: Post-Translational Modifications (PTMs); Syndecan-1 (SDC-1); Nuclear Magnetic Resonance (NMR); Tobacco Etch Virus (TEV); N-Acetylgalactosamine (GalNAc); liquid chromatography with tandem mass spectrometry (LC-MS/MS); heteronuclear single-quantum coherence (HSQC); matrix-assisted laser desorption/ionization (MALDI); electrospray ionization (ESI); liquid chromatography (LC); tandem mass spectrometry (MS/MS); collision-induced-dissociation (CID); higher-energy collision dissociation (HCD); data-dependent acquisition (DDA); data-independent acquisition (DIA); Glutathione S-transferases (GST); Maltose-binding protein (MBP); polyhistidine (His_6); N-utilization substance protein A (NusA); Tryptophan operon (Trp); mass-to-charge (*m/z*)

CHAPTER II

EXPRESSION AND PURIFICATION OF SYNDECAN-1 ECTODOMAINS

2.1 Introduction

2.1.1 Biological Significance of the Syndecan Protein Family

The syndecans are a family of four integral cell membrane heparan sulfate proteoglycans (HSPGs) as seen in **Figure 2.1**. Heparan sulfate (HS) is a type of carbohydrate polymer known as a glycosaminoglycan (GAG). While the structure of HS is quite variable[49], this GAG is evolutionarily ancient with a composition that has remained relatively unchanged from *Hydra* to humans.[50] Cell-surface HSPGs first characterized by Hook et al. in 1984 are vastly important mediators in many cellular interactions.[51] Many ligands found in the extracellular matrix are known to bind heparan sulfate, which gave rise to the concept that membrane-associated HSPGs played a wide role in numerous cellular interactions. Investigations into HSPGs led to the discovery of the syndecans, which are the primary cell-surface HSPG synthesized by many cells. All members of the syndecan family consist of a core protein with multiple long, covalently attached GAG chains. All syndecans exhibit HS as the primary GAG chain, although syndecan-1 and syndecan-4 have been proven to display chondroitin sulfate chains as well. Each of the four proteins consist of an extracellular domain, a single transmembrane domain, and a short, mostly conserved cytoplasmic domain.[52]

While syndecan ectodomains exhibit only limited amino-acid sequence similarities, the transmembrane and cytoplasmic domains are highly conserved across all four proteins.

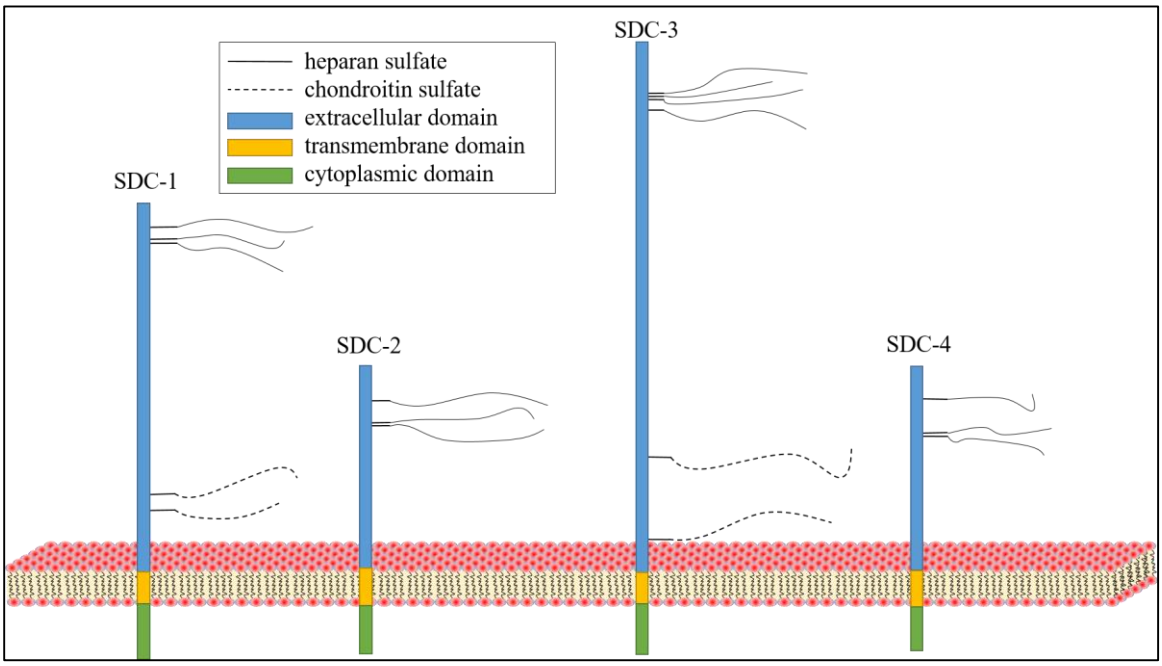

Figure 2.1: Syndecan Family of Proteins. All members exhibit a mostly-conserved transmembrane and cytoplasmic domain. Heparan sulfate is the predominant GAG, however SDC-1 and SDC-4 both exhibit chondroitin sulfate chains near the transmembrane domains

These proteins are immensely important in cell-signaling and stimuli response. However, with the exception of the syndecan-4 cytoplasmic domain, the 3D structures of the syndecan family are not well characterized.[53]

It is believed that intrinsic disorder is responsible for facilitating the large number of binding partners in the syndecan family.[54] In total, the syndecan family has 351 known binding partners. Out of all the syndecans, syndecan-1 has the largest number of known partners at 131.[53] Syndecan-1 is also predicted to have the greatest amount of intrinsic disorder with over 80% of its amino acids predicted to be disordered.[54] Comparing amino acid occurrences across multiple animal proteomes; glycine, alanine, threonine, glutamic acid, and proline were found to predominate in the syndecans. In contrast, amino acids that promote ordered structure such as cysteine, tryptophan, asparagine, and histidine were found to occur much less frequently in

syndecans compared to the rest of the proteome. In the same study, a highly-conserved aspartic acid-rich region located in the ectodomains of the syndecans was also identified.[55]

2.1.2 Functions of Syndecan-1

First cloned and named in 1989, syndecan-1 (SDC-1) is the primary HSPG exhibited by many epithelial cells.[56] SDC-1 is a very promising protein to dedicate research efforts towards as it has been shown by multiple studies that one of the primary functions of SDC-1 is in regulating the innate immune response system of the body. It has a hand in all parts of wound healing such as cell migration, proliferation, and inflammation.[57-59] In mouse models, knockout of the SDC-1 gene has proven to have major impacts on the body's ability to control the inflammatory response.[58] SDC-1 oversees inflammatory response by regulating leukocyte adhesion and migration. The protein regulates the generation and activity of chemokine gradients, which are in turn used to attract leukocytes to the site of infection or injury. Chemokines bind to the protein's ectodomain and, like other HSPGs, syndecan-1 may also directly potentiate or inhibit chemokine activities.[60]

As a part of the natural function of SDC-1, the ectodomain of the protein is cleaved and shed into the extracellular matrix. This soluble ectodomain retains its activity and binding partners.[61] The shedding of SDC-1 occurs to a small degree under normal physiological conditions, but this shedding can be dramatically increased in response to stimuli. Shedding of the SDC-1 ectodomain is one way cells are able to communicate over long distances, generally as a call for aid to an area affected by injury or disease. It has been shown that the SDC-1 ectodomain is more susceptible to sheddases after losing its heparan sulfate chains to heparanase indicating that these attached sugars play an important role in the accessibility of the membrane protein.[62]The most common sheddases for syndecans are matrix metalloproteinases (MMPs).[63]

SDC-1 has two main MMP cleavage sites; one located near the N-terminus between glycine-82 and leucine-83, the other site is only a few residues away from the transmembrane domain between glycine-245 and leucine-246. Some of these MMPs, such as MMP2 and MMP9, specifically cleave only at the glycine-82 location. MMP3 cleaves at the glycine-245 location. MMP7 may cleave at either location.[63] This indicates that a soluble syndecan-1 (sSDC-1) ectodomain may exist in sera as either the truncated form consisting of the N-terminus to glycine-82 (SDC-$1_{1\text{-}82}$) or the full-length form from the N-terminus to glycine-245 (SDC-$1_{1\text{-}245}$). The release of sSDC-1 ectodomains resulting from MMP cleavage is depicted below in **Figure 2.2**. It is also important to note that shed sSDC-1 ectodomain retains nearly all of its binding partners. However, upon loss of its heparan sulfate chains, sSDC-1 loses most of its binding partners[62]

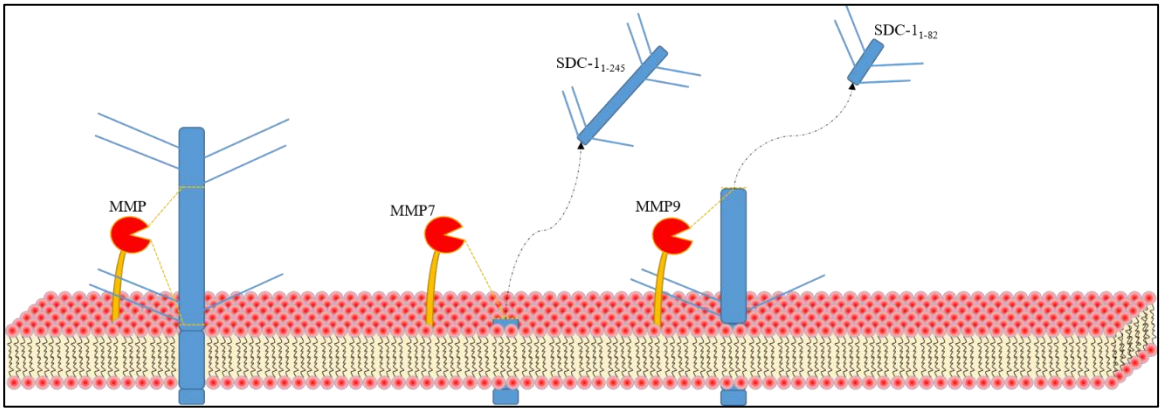

Figure 2.2: SDC-1 Cleavage. Syndecan-1 is cleaved by metalloprotease and the soluble ectodomain is subsequently diffused away from the cell membrane.

2.1.3 Literature Review

Although first cloned in 1989, studies on SDC-1 date back to around 1987 when it was first shown that an unidentified cell surface proteoglycan found in mouse mammary epithelial cells containing both heparan sulfate and chondroitin sulfate could be shed from the surface of the cells via cleavage of its ectodomain from the membrane associated domain.[64] It has been shown that the serum concentration of sSDC-1 increases with various cancers. Serum levels of sSDC-1 have already been established to be a useful biomarker in determining prognosis of multiple myeloma

patients.[65, 66] Some cancer lines have also been shown to overexpress certain MMPs. For example: breast cancer, prostate cancer, bladder cancer, and multiple myeloma all overexpress and/or activate either MMP2 or MMP9, both of which cleave SDC-1 specifically at the glycine-82 residue.[63, 67-70] These cancers have also all been noted for increased levels of sSDC-1 in sera.[65, 71-74] One can draw the conclusion that a majority of these sSDC-1 ectodomains are likely in the form of SDC-$1_{1\text{-}82}$. There have been no studies on sSDC-1 levels that have attempted to distinguish which part of the SDC-1 ectodomain has been released into patient sera.

The majority of studies involving SDC-1 have been performed with a goal of probing the effects of SDC-1 presence in conjunction with various disease states. As stated previously, SDC-1 loses a majority of its binding partners after a loss of its glycan chains. Therefore, to retain the protein's various binding affinities and interactions, it has been imperative for these studies to use glycosylated forms of SDC-1. One such study performed by Xu et al. used glycosylated SDC-1 ectodomains purified from murine mammary gland epithelial cells. This demonstrated that administration of these ectodomains to syndecan-1-null mice showed a significant inhibition of the inflammatory response in allergic lung inflammation. When these same mice were treated with SDC-1 ectodomains devoid of heparan sulfate, no inhibition effects were observed.[59] This again attests to the vital role played by glycosylation in the activity of SDC-1. However, no current studies have been able to probe exactly how glycosylation effects the structure or dynamics of this highly important membrane protein.

2.1.4 Summary and Outlook

SDC-1 is an integral membrane HSPG that plays a pivotal role in signaling and regulation of the inflammatory response system. Knowing that certain cancers and other ailments may generate specific lengths of the sSDC-1 ectodomain, it is important to be able to distinguish which part of the SDC-1 ectodomain is being shed. By producing these two versions of the shed sSDC-1 that

can be found after cleavage at the glycine-82 or glycine-245, we can test popular antibodies used in other studies to determine which antibodies, if any, are able to distinguish and show preferential binding for either of the ectodomain lengths.

Attempts in expression and characterization of the full-length SDC-1 protein face several challenges. Overexpression of integral membrane proteins (IMPs) in *E. coli* cells can be hindered by untimely cell death. This is due the IMPs overpopulating and destabilizing the cell membrane following induction of recombinant protein expression. IMPs are also exceedingly difficult to study using conventional techniques such as NMR spectroscopy and X-ray crystallography.[75] The insolubility of these membrane proteins necessitates experimental techniques for solution-based studies be altered to solubilize the protein. This is often accomplished via addition of detergents or lipids for solution-state NMR.

The purpose of the work in this chapter was to develop a method for the expression and purification of the two full length soluble ectodomain sequences of SDC-1 that are most likely to be found in patient serum. In this chapter, we used molecular cloning techniques to develop two different cell lines that would express these two ectodomain sequences incorporated into fusion proteins. The two protein sequences that were generated for this study along with the full-length SDC-1 transmembrane protein sequence can be found in **Figure 2.3**. The fusion partner attached to the two sSDC-1 ectodomains provides a straightforward route for purification and subsequent separation of the fusion tag from the sSDC-1 sequences. The work performed here was crucial for subsequent NMR structural elucidation and glycosylation studies.

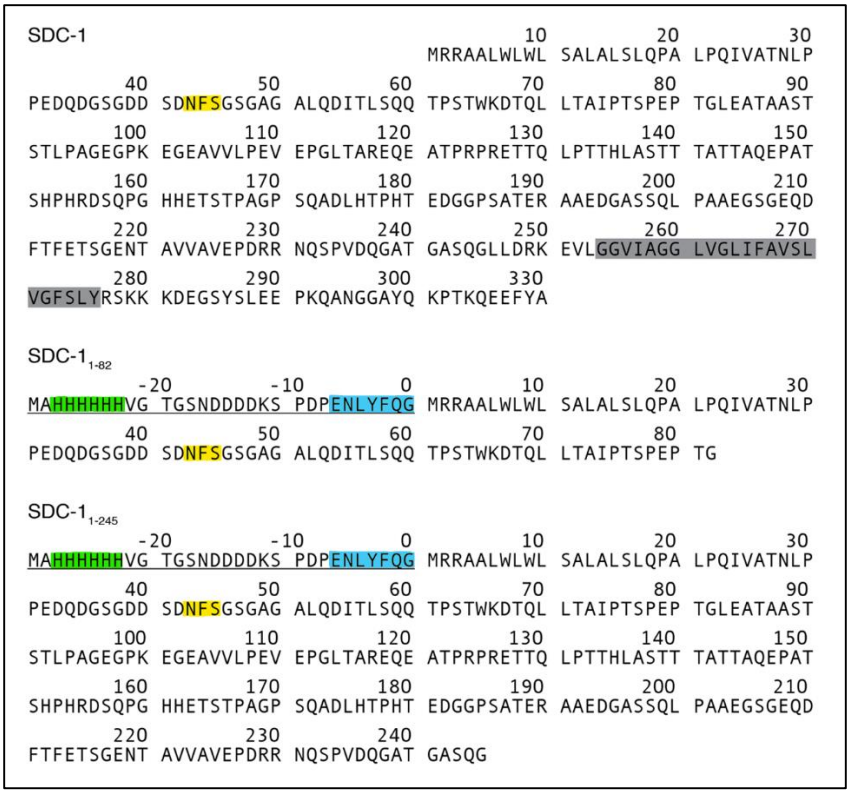

Figure 2.3: SDC-1 Sequences. The full length SDC-1 protein (top sequence) has a single transmembrane domain highlighted in grey. Full length SDC-1 exhibits a single N-linked glycosylation site at Asn 43 highlighted in yellow. This glycosylation site is also incorporated into our two ectodomain sequences SDC-$1_{1\text{-}82}$ and SDC-$1_{1\text{-}245}$. The ectodomain sequences both contain a 30 amino acid fusion partner underlined at the start of each sequence. This fusion partner incorporates a 6x His-tag highlighted in green along with a TEV protease consensus sequence highlighted in cyan.

2.2 Materials and Methods

2.2.1 Vector Selection and Vector DNA Purification

The pET-45b expression vector (**Figure 2.4** Novagen) was chosen to express our two syndecan constructs. This vector provides for the expression of an N-Terminal His-tagged fusion protein and includes a multiple cloning site region to clone target proteins with minimal extraneous amino acid sequences. This vector also provides the bacteria with carbenicillin resistance allowing the use of carbenicillin in growth media, thus helping to ensure there is no other bacterial contamination in the cultures.

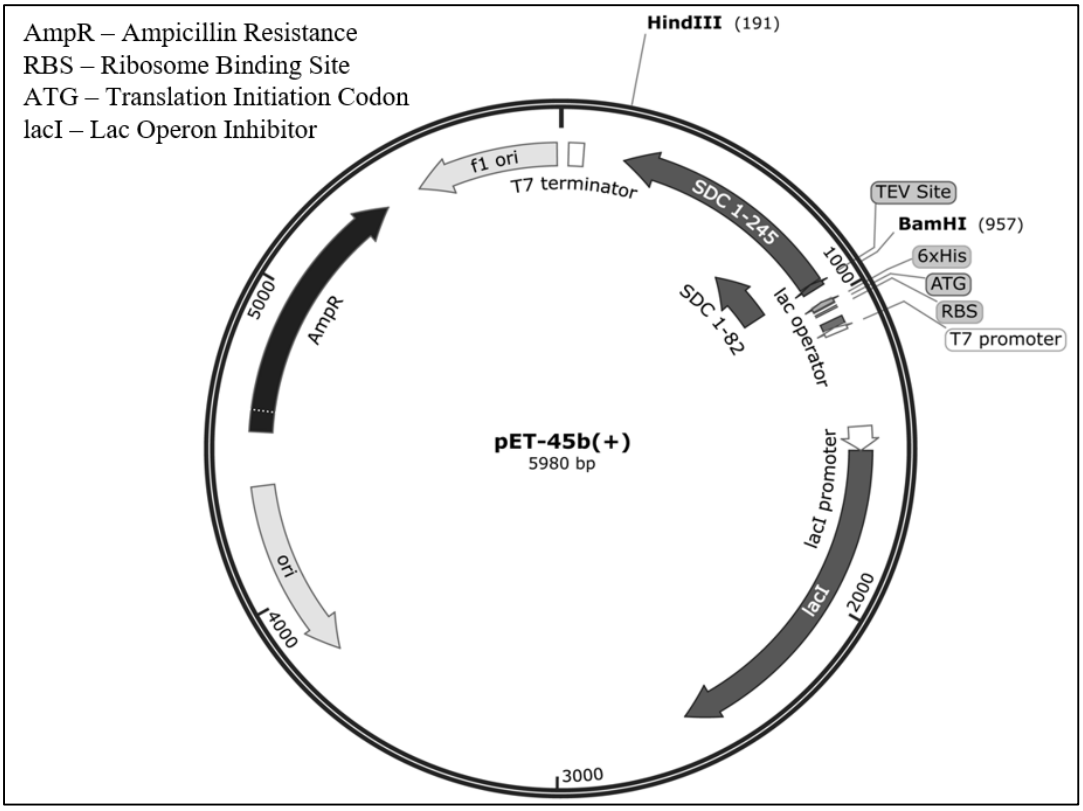

Figure 2.4: Customized pET-45b Expression Vectors. Two different expression vectors were generated for this project, one coding for SDC-$1_{1\text{-}82}$ and the other for SDC-$1_{1\text{-}245}$. Introduction of isopropyl β-D-thiogalactopyranoside (IPTG) to bacterial cells with these vectors will induce the expression of the two recombinant SDC-1 ectodomains.

The unmodified vector DNA was received as a spot on a filter paper disk. A small square was cut from this spot and incubated in a 1.5 mL microcentrifuge tube with 40 µL of DNA elution buffer from a GeneJET Plasmid Miniprep Kit (ThermoFisher) for 30 minutes. The solution was then extracted from the tube and placed into a fresh tube for storage. The vector DNA solution was then transformed into DH5α competent bacterial cells (Invitrogen) using a heat-shock transformation process. Five µL of vector DNA was mixed with 50 µL of the bacterial cells over ice. After 30 minutes on ice, the mixture was place in a water bath at 42°C for exactly 45 seconds, then immediately transferred back over ice. After 2 minutes on ice, 250 µL of SOC media (25 g/L Luria Broth, 2.5 mM KCl, .4% (w/v) sterile glucose, 20 mM $MgSO_4$) was added to the cells. The cells were then grown for 1 hour at 37˚C on a shaker (MaxQ 300, Thermo Fisher

Scientific) set to 220 rpm. One hundred microliters of cells were plated on LB/carbenicillin plates (25 g/L Luria broth, 15 g/L agar, 100 µg/L sterile-filtered carbenicillin) and incubated overnight at 37˚C. The following day, colonies were picked from the plates and deposited into a 50 mL conical tube containing 5 mL LB/carb media (25 g/L Luria broth, 100 µg/L sterile-filtered carbenicillin) These cultures were again placed on the shaker at 37˚C. After 2 hours, the cells were then spun down at 5000 rpm for 10 minutes using a tabletop centrifuge. The media was decanted and the DNA was extracted from the cell pellet using a GeneJET Plasmid Miniprep DNA extraction kit. The purified vector DNA was then transferred to a clean microcentrifuge tube and stored at -20˚C.

2.2.2 Primer Design and PCR Cloning of DNA Inserts

These studies examined the two significant portions of the SDC-1 protein that are cleaved in the process of ectodomain shedding. The two recombinant proteins produced for this study were derived from the *Homo sapiens* syndecan-1 amino acid sequence. SDC-$1_{1\text{-}82}$ consists of the N-terminal methionine residue through glycine-82. SDC-$1_{1\text{-}245}$ consists of the same N-terminal methionine through glycine-245.

To generate these polypeptides, DNA inserts to be ligated into the pET-45b expression vector were produced via polymerase chain reaction (PCR). The template DNA used for PCR was the TrpΔLE full-length SDC-1 fusion protein used in previous studies in the Cook laboratory. Primers depicted in **Figure 2.5** were designed in-house and purchased from Integrated DNA Technologies (idtdna.com).

Forward Primer	5' - GCG CGC GGA TCC CGA AAA CCT GTA TTT TCA GGG CAT G - 3'
SDC-1_{1-82} Reverse	5' - CGC CGC AAG CTT TTA TTA GCC GGT CGG TTC CGG GCT GGT CGG AAT CGC - 3'
SDC-1_{1-245} Reverse	5' - CGC CGC AAG CTT TTA TTA GCC CTG GCT CGC GCC GGT CGC GCC CTG ATC - 3'

Figure 2.5: Primers A single forward primer was used to produce both inserts. This primer coded for a BamHI restriction enzyme recognition site followed by a TEV protease consensus sequence, thus allowing for ligation into the vector and fusion protein cleavage after expression. The two different reverse primers inserted a pair of stop codons after amino acids glycine-82 for SDC-1_{1-82} and glycine-245 for the SDC-1_{1-245}. This stop codon was followed by a HindIII restriction site for ligation.

The protocol used for PCR was derived from guidelines obtained for the Platinum™ Pfx DNA Polymerase (Invitrogen). A 50 µL mixture containing 5 µL 10x Pfx buffer, 1.5 µL 10 mM dNTP, 1 µL 50 mM $MgSO_4$, 1.5 µL 10 µM forward primer, 1.5 µL 10 µM reverse primer, 1 µL Template DNA, 0.4 µL Pfx Polymerase, and 38 µL WFI quality water (Corning) was deposited into a ThermoGrid ™ PCR Tube (Denville Scientific). The mix was placed in thermocycler (Techne 3prime thermal cycler) set up for an initial denaturation at 94˚C for 5 minutes, followed by 30 cycles with the following sequence: denaturation at 94˚C for 15 s, annealing at 55˚C for 30 s, and extension at 68˚C for 1 min. The program included a final hold cycle at 4˚C.

After PCR, 10 µL of 6x DNA loading dye (ThermoFisher Scientific) was added and, after mixing, the amplified DNA was then loaded into the wells of a 1.5% agarose DNA gel (1.125 g agarose, 75 mL 1x TAE buffer, 7.5 µL SYBR Safe™) in 3 separate aliquots of 20 µL. Five microliters of TrackIt™ 100 bp DNA Ladder (Invitrogen) was added to an empty well and electrophoresis (Owl™ B1A) was performed at 80 mV for 1hr. Following electrophoresis, the gel was placed over a UV imaging system to visualize **(Figure 2.6 A)** and cut out the purified DNA inserts using a razor blade. Time spent under UV light was minimized to prevent DNA damage by turning off the lamp immediately after identifying and imaging the bands in the gel. These inserts were then extracted from the gels using a GeneJET gel extraction kit (ThermoScientific), following the given protocol. The purified DNA inserts were stored at -20˚C awaiting restriction digestion.

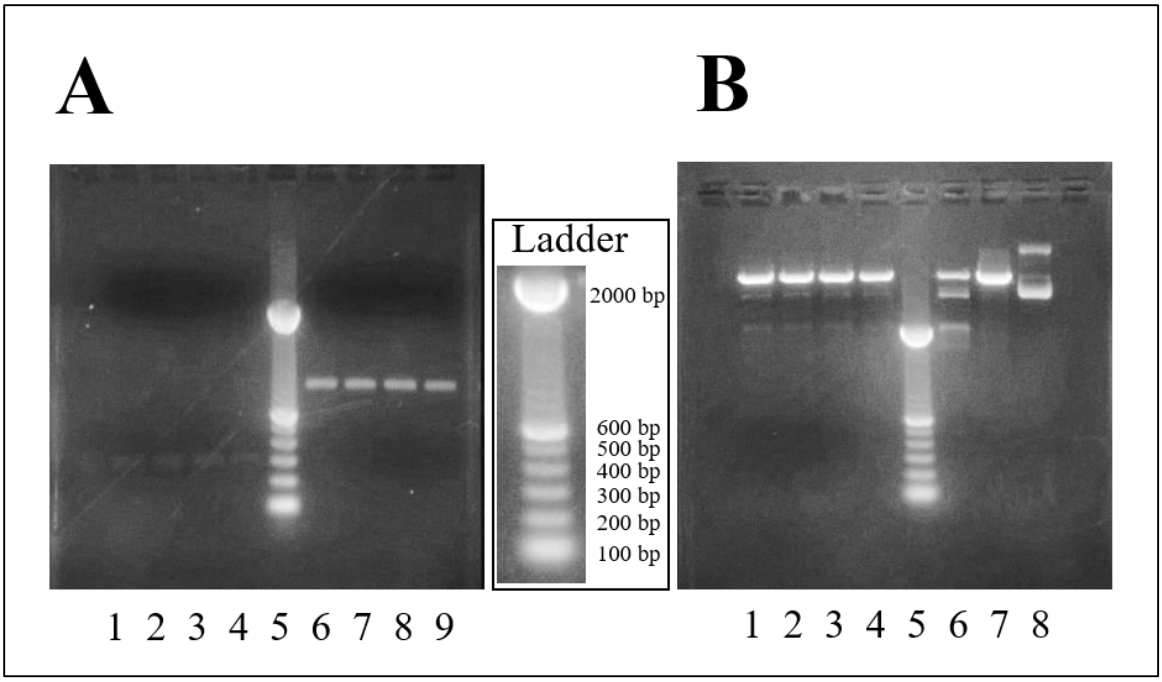

Figure 2.6: Agarose DNA Gel Images

A) Lanes 1-4 show the PCR products for our SDC-$1_{1\text{-}82}$ DNA insert. Lanes 6-9 are the PCR products for the SDC-$1_{1\text{-}245}$ insert. Lane 5 is a 100 base pair DNA ladder.
B) Lanes 1-4 are the pET-45b vector DNA after restriction digestion. Lane 5 is a 100 base pair DNA ladder. Lane 6 and 7 are controls where only one restriction enzyme was used to ensure activity of both enzymes. Lane 8 is undigested plasmid DNA.

2.2.3 DNA Digestion, Vector Ligation, and Cell Transformation

Restriction digestion was performed on both the plasmid and insert DNA by mixing 40 µL of DNA with 4 µL of each restriction enzyme (New England Biolabs), 10 µL of 10x Tango Buffer (ThermoFisher Scientific), and 42 µL of sterile water in a PCR tube. After an incubation period of 2 hours at 37˚C, 20 µL of 6x DNA loading buffer was added and the samples were loaded into a 1.25% agarose DNA gel (0.9375 g agarose, 75 mL 1x TAE buffer, 7.5 µL SYBR Safe™) for isolation of the cleaved vector and inserts. Electrophoresis settings remained unchanged and the desired DNA was again cut out after imaging **(Figure 2.6 B)** and purified using the gel extraction kit.

The elutions containing the digested DNA inserts from four separate rounds of digestion, gel electrophoresis, and DNA extraction were combined in a 1.5 mL Eppendorf tube and concentrated down using a centrifugal evaporator (Savant SpeedVac SC110 Vacuum Concentrator). Prior to ligation, the concentrations of the digested SDC-$1_{1\text{-}82}$ insert, SDC-$1_{1\text{-}245}$ insert, and plasmid were found to be 29 ng/µL, 32 ng/µL, and 20 ng/µL, respectfully. These DNA concentrations were determined using a NanoDrop spectrometer (ThermoScientific ND-1000) supplied by the Oklahoma State DNA/Protein Core Facility. The digested insert DNA was ligated into the digested vector using a 3:1 molar ratio of insert to vector DNA. The ligation mix was made by combining 5 µL of vector DNA with the appropriate molar amount of insert DNA, 2 µL of 10x ligase buffer, 1 µL of T4 DNA Ligase (ThermoScientific 5 Weiss U/µL), and diluted to a total volume of 20 µL with sterile water. The ligation mixtures were allowed to react at room temperature. After one hour, 5 µL of the ligation mix was added to 50 µL of DH5α cells in a 1.5 mL Eppendorf tube. The mixture was set on ice for 30 min. Heat shock transformation was then performed by placing the tubes in a water bath at 42˚C for 45 seconds immediately followed by placing them back in ice. After 2 minutes on ice, 250 µL of SOC media was added to the cells. The cells were then grown for 1 hr at 37˚C on a shaker set to 220 rpm. One hundred microliters of cells were plated on LB/carbenicillin plates and incubated overnight at 37˚C. The following day, colonies were picked from the plates and deposited into a 50 mL conical tube containing 5 mL LB/carb media. These cultures were again placed on the shaker at 37˚C. After 2 hours, the cells were then spun down at 5000 rpm for 10 minutes using a tabletop centrifuge. The media was decanted and the DNA was extracted from the cell pellet using a GeneJET Plasmid Miniprep DNA extraction kit. Half of the 50 µL DNA elution isolated from the extraction kit was sent to the Core Facility where it was confirmed that we had obtained the correct DNA sequence for the expression of both ectodomains.

Following DNA sequence conformation for both cell lines, 5 µL of the DNA elution was transformed into 50 µL of BL21(DE3) expression cells (Invitrogen) using the previously described heat-shock transformation process. The newly transformed cells were plated and colonies were picked the following day to begin initial expression trials and the making of cell stocks. The picked colony was added to 10 mL of LB/carb media in a 50 mL conical tube and grown at 37˚C with 220 rpm shaking. after 2 hours, 5 mL of cells were removed from the conical tube to use for DNA sequencing. The DNA was extracted using the same plasmid DNA extraction kit and protocol previously mentioned. The DNA sequence for both cell lines was again confirmed by the Core Facility. For making cell stocks, the remaining culture in the 50 mL tube was left to grow while monitoring the optical density at 600nm (OD_{600}). Once the OD_{600} reached 0.4, 3 mL of the culture was transferred to a 15 mL conical tube on ice and 1.02 mL of filter-sterilized 50% glycerol was added to make a final glycerol concentration of 17% in the cell stocks. Cell stocks were then deposited in 100 µL aliquots into steam-sterilized 1.5 mL Eppendorf tubes. To ensure cell growth and division was minimized while making the cell stocks, cell stock preparation was performed over ice. Once the stocks were aliquotted out, they were subsequently frozen at -80˚C.

2.2.4 Expression of Unlabeled SDC-1_{1-82} and SDC-1_{1-245}

For production of the recombinant proteins, a starter culture was made and, after 2 hours, 200 µL of this culture was used to inoculate an overnight growth of 100 mL of fresh LB/carb media in a 250 mL baffle-bottom flask. (Thermo Fisher Scientific) After 16 hours, 50 mL of the overnight culture was transferred into 950 mL of fresh LB/carb media in a 2 L baffle-bottom flask (Thermo Fisher Scientific). The culture was grown 2-3 hours while periodically monitoring the OD_{600} value. Once the OD_{600} reached 0.6, recombinant protein expression was then induced by adding 1 mL of 120 mg/mL isopropyl β-D-thiogalactopyranoside (IPTG) to the growth flask. IPTG, a molecular mimic of allolactose, turns the *lac* inhibitor off, allowing expression of the

recombinant protein in the vectors. Prior to induction, a 500 µL sample was removed (pre-induction sample) and saved for monitoring protein expression levels. The culture was then allowed to grow an additional four hours. Each hour, 1 mL samples were taken from the culture media to monitor the OD_{600} value. An appropriate amount (an amount corresponding to the same number of cells in a 500 µL sample with an OD_{600} of 0.6) of each sample was saved (post-induction samples) to monitor protein expression over time. The expression samples were each deposited into separate 1.5 mL Eppendorf tubes and spun down at 14,000 × g in a tabletop micro centrifuge for 5 minutes. The supernatant was drawn off and discarded with a disposable micropipette and the resulting cell pellet was frozen at -20˚C. Collecting and saving the samples in this manner allowed us to check expression levels of the recombinant protein while normalizing the total cell count in each sample. After four hours, the cells were harvested by transferring the cultures to 1 L polycarbonate centrifuge bottles and spinning them down at 6,500 rpm for 25 min at 4˚C (ThermoScientific, Sorval Lynx 4000 Centrifuge, with F10–4×1000 LEX rotor). The supernatant was discarded and cell pellets were frozen at -80˚C.

Recombinant protein expression was checked via SDS-page gel electrophoresis **(Figure 2.8)**. Two-hundred microliters of 1x LDS buffer (62.5 mM Tris-HCl pH 6.8, 25 mg/mL SDS, 1.25 mg/mL bromophenol blue, 10% (v/v) glycerol, 5% (v/v) 2-mercaptoethanol) was added to the Eppendorf tubes containing the cell pellets from the expression pre- and post-induction samples. The cells were lysed with a French press method utilizing a small-gauge disposable syringe. After lysis, the samples were boiled in a water bath for 10 minutes. A NuPAGE™ 4-12% Bis-Tris gel cassette (Invitrogen) was placed inside an electrophoresis apparatus filled with MES running buffer (50 mM MES, 50 mM Tris base, 0.1% SDS, 1 mM EDTA, pH 7.3). For each expression sample, a 15 µL aliquot was deposited into a well of the gel and a PageRuler™ prestained protein ladder (Thermo Fisher Scientific) was used to monitor the electrophoresis progress. After electrophoresis, gels were stained by depositing the gel in coomassie blue, microwaving for 30s,

and then rocking for 4 hours. Gels were then de-stained by exchanging the coomassie blue for de-stain solution (40% v/v methanol, 10 % v/v acetic acid)

2.2.5 Purification of SDC-1 Constructs

Harvested cell pellets were removed from the -80°C and resuspended with 30 mL of RS buffer (50 mM Tris-HCl, 1 mM NaN3, 1.5% (v/v) glycerol, pH 8.0) in a 50 mL conical tube. The cells were vortexed for 1 min and the resulting cell slurry was placed over ice prior to lysis. The cells were mechanically lysed via sonication using a Fisher Scientific 550 Sonic Dismembrator. The cells were lysed for 20 min, with sonication intervals of 2 sec "on" and 8 sec "off". The lysed cells were then spun down at 4˚C for 25 min at 16,500 rpm. The soluble SDC-1 fusion proteins were isolated in the supernatant solution. The supernatant was filtered through a 0.20-micron syringe filter into a 50 mL conical tube and placed in the 4˚C refrigerator awaiting Ni-NTA purification.

Ni-NTA affinity chromatography was used to selectively purify the 6x His-tagged SDC-1 fusion proteins. A 5 mL Ni-NTA resin bed (ThermoScientific HisPur™) gravity column pre-charged with 0.1 M $NiSO_4$ was prepared by washing the column with 25 mL of equilibrium buffer (50 mM Tris-HCl, 1 mM NaN_3, 1.5% (v/v) glycerol, 5 mM imidazole, pH 8.0). The lysis supernatant was removed from the refrigerator and made to be 5 mM imidazole by adding 300 µL of elution buffer (50 mM Tris-HCl, 1 mM NaN_3, 1.5% (v/v) glycerol, 500 mM imidazole, pH 8.0) to the tube. One quarter of a cell pellet or 7.5 mL of supernatant was added to the column. The column was then washed with 25 mL of wash buffer (RS buffer with imidazole: 75 mM imidazole for the SDC-$1_{1\text{-}82}$ and 50 mM imidazole for the SDC-$1_{1\text{-}245}$). After washing, 15 mL of elution buffer was added to the column and collected in 1 mL fractions. These fractions were examined using SDS-PAGE electrophoresis where it was determined that both of the proteins are completely eluted from the column within the 2-7 mL buffer fractions. After every round of protein purification, the

columns were stripped with strip buffer (20 mM Tris-HCl, 500 mM NaCl, 50 mM EDTA, pH 8.0) and recharged with $NiSO_4$ for the next round. The protein elution was then either dialyzed against water and lyophilized or stored in the elution buffer at 4˚C awaiting dialysis against another desired buffer.

To dry the protein, the elution was placed into 3.5 kDa molecular weight cutoff dialysis tubing (SnakeSkin™ 3.5MWCO 35 mm I.D. ThermoScientific) and dialyzed against 5 L of water in a beaker over a magnetic stirrer. The dialysis water was renewed twice, once after 30 minutes, once more after 2 hours, and then allowed to dialyze overnight. The contents of the dialysis bag were then transferred to a 50 mL conical tube, frozen with liquid nitrogen, and vacuum lyophilized (Labconco FreeZone 2.5L). On average, a 1 L LB growth yielded around 50 mg of dry SDC-$1_{1\text{-}82}$ and SDC-$1_{1\text{-}245}$ protein.

2.2.5.1 Tobacco Etch Virus (TEV) Protease Expression and Purification

After isolating the SDC-1 fusion proteins, TEV is used to cleave the fusion partner to separate the desired sSDC-1 ectodomain sequences. To this end, TEV protease must be expressed and purified for subsequent use in fusion protein cleavage. The main differences between purification of sSDC-1 and TEV are the buffers used and the fact that all TEV purification steps must be performed under cold conditions. An adaptation of the protocol by JE Tropea et al. for the purification of His6-TEV(S219V)-Arg5 protease was employed.[76] Recombinant TEV protease was expressed from BL21 *E coli* cells exactly as previously described for the two SDC-1 constructs. The cells were then resuspended in 30 mL of TEV resuspension buffer (20 mM Tris, 500 mM NaCl, 5% glycerol, 40 mM Imidazole, pH 7.8). Cells were then lysed in an ice water bath by sonication with intervals of 2s on, 8s off for 20 minutes. Cell lysate was then spun down at 16,500 rpm for 25 minutes at 4˚C. The resulting supernatant was then purified via a batch-binding Ni-NTA affinity chromatographic method.

To purify the TEV protease, the entire 30 mL of TEV supernatant is mixed with Ni-NTA resin (10 mL gel bed, ~20 mL slurry) that has been pre-equilibrated with TEV resuspension buffer. The mixture is then rocked at 4˚C for one hour. The slurry is then transferred to a 50 mL gravity flow column where the resin is collected and all liquid is allowed to flow through. The resin is then washed with 50 mL of cold TEV wash buffer (20 mM Tris, 500 mM NaCl, 5% glycerol, 80 mM Imidazole, pH 7.8). The TEV protease is then eluted from the column by adding 20 mL of TEV elution buffer (20 mM Tris, 500 mM NaCl, 5% glycerol, 500 mM Imidazole, pH 7.8). DTT is then added to the elution to a final concentration of 5 mM to help stabilize the protease and prevent autolysis.

2.2.5.2 TEV Cleavage of SDC-1 Fusion Proteins and Separation of Cleavage Fragments

TEV recognizes and readily cleaves proteins with a highly specific consensus sequence of amino acids. The optimum recognition site is as follows:

Glu-Asn-Leu-Tyr-Phe-Gln-[Gly/Ser] (ENLYFQ[G/S])

The most commonly used consensus sequence for cleavage of fusion proteins is ENLYFQ/G where the protease will cleave between the glutamine and glycine residues.[77] To cleave our fusion proteins, we first purified one cell pellet of previously expressed TEV protease. To ensure maximum cleavage efficiency, we subsequently used the TEV to cleave the fusion proteins immediately after purification. To this end, we mixed ½ of a cell pellet worth of TEV protease (10 mL of TEV elution) with one cell pellet (~50 mg of dry protein or 20 mL of Ni-NTA elution) of purified sSDC-1 fusion protein in a 50 mL conical tube. The mixture was then transferred into 3.5 kDa molecular weight cutoff dialysis tubing and placed in dialysis against 5L of TEV resuspension buffer. The cleavage mix was then allowed to dialyze overnight at 4°C.

The following day, the cleavage mix was transferred to a tub containing Ni-NTA resin (10 mL of resin bed ~20 mL of resin slurry) pre-equilibrated with TEV resuspension buffer. The mix was

then rocked for 1 hour at 4°C to ensure efficient binding of any polyhistidine tags to the resin material. The mix was then transferred to a gravity column and the flowthrough was collected. After cleavage, both TEV protease and the His-tagged fusion partner would remain bound to the affinity resin, while cleaved SDC-$1_{1\text{-}82}$ and SDC-$1_{1\text{-}245}$ would flow through the column. The separation of the fusion partner from recombinant syndecan-1 sequences is depicted below in **Figure 2.7**.

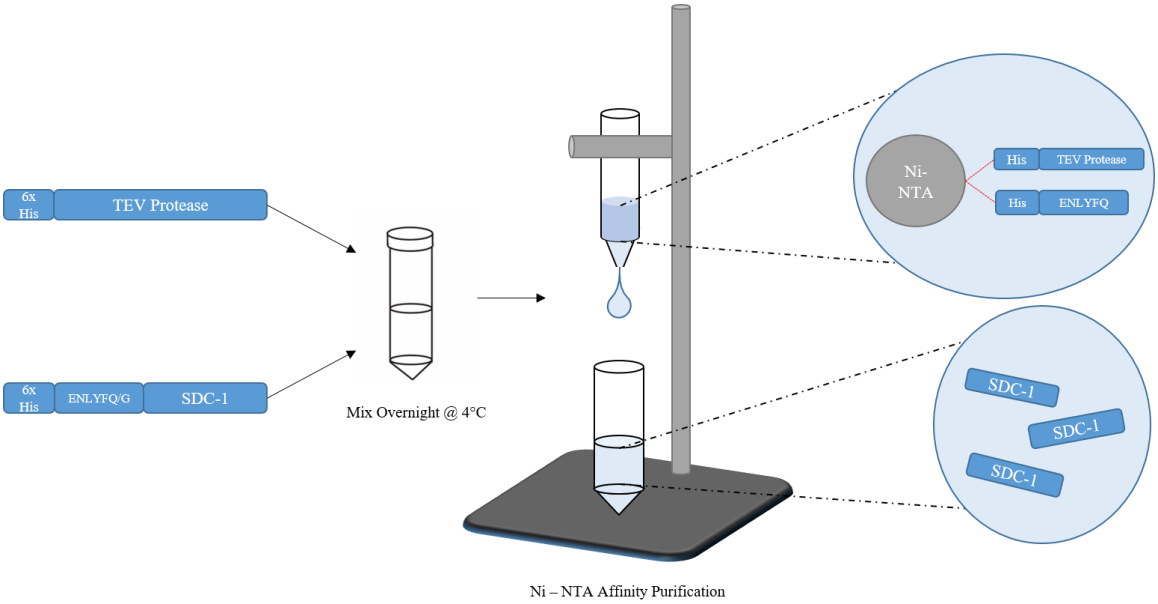

Figure 2.7: Cleavage of Fusion Proteins and Ni-NTA Purification. Ni-NTA resin is used for purification of His-Tagged proteins via affinity chromatography. Ni^{2+} ions are captured and immobilized via coordination with nitrilotriacetic acid (NTA) on agarose beads. The imidazole on the side chain of histidine residues has a strong affinity for Ni^{2+}. Therefore, any proteins that have a 6x histidine tag bind favorably to the column resin; allowing non-tagged proteins to be separated. The His-tagged proteins can then be eluted from the column by adding a buffer with increased amounts of imidazole to compete for Ni^{2+} binding sites.

2.2.6 Top-Down Mass Spectrometry

Isolation of the sSDC-1 ectodomains was confirmed by Top-Down mass spectroscopy. To this end, purified sSDC-1 elution from the Ni-NTA column elution was placed into 3.5 kDa MWCO dialysis tubing and dialyzed against 5 L of water. The dialysis water was exchanged for fresh water twice: once after 30 minutes of dialysis and again 2 hours later. The protein was then left to dialyze overnight. The following day, the protein was transferred to a 50 mL conical tube and lyophilized. Milligram amounts of protein powder were then resolvated in an appropriate solvent depending on which ionization method was being used.

2.2.6.1 MALDI Mass Spectrometry

Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) spectrometry was used to check the size of the isolated proteins before and after TEV cleavage during initial purification trials. The experiments were performed in the Oklahoma State DNA/Protein Core Facility using the Voyager DE Pro mass spectrometer (ABSCIEX) operated in the positive linear mode utilizing a BSA mass calibration standard. Around 0.1mg of the dried proteins were dissolved in a 250 µL mixture of 50% acetonitrile, 49.9% H_2O, 0.1% trifluoroacetic acid in a 1.5 mL Eppendorf tube. Solvation of the protein was aided by placing the tubes into a sonic bath (Branson 1800 ultrasonic cleaner) for 1 hour. The tubes were then spun down at 14,000xg for 5 minutes to pellet any undissolved protein. The dissolved protein was then pipetted into a new Eppendorf tube and mixed in a 1:1 ratio with a 250 µL matrix solution of sinapinic acid (10 mg of sinapinic acid in 50% ACN, 49.9% H2O and 0.1% TFA) for spotting on the target plate. Mass spectra were then acquired over the mass to charge (*m/z*) range of 1000-25,000 for the cleaved and uncleaved SDC-$1_{1\text{-}82}$ protein. For the SDC-$1_{1\text{-}245}$ protein, we scanned a range of 10,000-70,000 *m/z*. All spectra were captured using 30 laser shots per spectrum. **(Figure 2.12)**

2.2.6.2 ESI Mass Spectrometry and Deconvolution of Mass Spectra

After finalizing the sSDC-1 purification process, Electrospray ionization mass spectrometry was used to check the masses of the isolated sSDC-1 proteins. The experiments were performed in the Oklahoma State DNA/Protein Core Facility using a Thermo Scientific Orbitrap Fusion Tribrid mass spectrometer. Around 1 mg of the dried proteins were each solubilized in 100 µL of concentrated formic acid in a 1.5 mL Eppendorf tube. A sonic bath was then used to assist in dissolution of the proteins. After the proteins were solubilized, the solutions were diluted with 900 µL of a 50/50 (v/v) mixture of acetonitrile and water. The tubes were then spun down at 14,000 xg for 5 minutes to pellet any undissolved protein. The samples were subsequently pulled into a syringe and directly injected into the electrospray system. Mass spectra were then acquired over the mass-to-charge (*m/z*) range of 150-2,000 for the SDC-$1_{1\text{-}82}$ protein. For the SDC-$1_{1\text{-}245}$ protein, we scanned a range of 300-2,000 *m/z*. We collected ~200 spectral signals with a sampling rate of two samples per second and signal averaging was used to average the individual spectra into a single *m/z* spectrum. For the SDC-$1_{1\text{-}82}$ protein, the best spectrum was acquired using the quadrupole mass analyzer while rinsing the sample input line with concentrated formic acid. For the SDC-$1_{1\text{-}245}$ sample, the best spectrum was acquired using the Orbitrap mass analyzer with a 100x dilution of the original sample. The resulting charge distribution spectra were then plotted in UniDec[78] deconvolution software to generate the zero charge mass spectra.**(Figure 2.13)**

2.3 Results

In this study, we successfully used molecular cloning to integrate the sSDC-1 ectodomain coding sequences into an expression vector, which was confirmed by DNA sequencing. In the process of primer design of the DNA inserts we integrated an N-terminal TEV protease consensus sequence which allows for cleavage of the fusion protein without any need for mutations to the core protein sequence. This expression vector was then transformed into BL21expression cells and expression

tests to determine if the recombinant protein was being properly produced by the bacteria after induction with IPTG were performed. The resultant electrophoresis gels **(Figure 2.8)** showed excellent overexpression of protein after induction. However, both bands of sSDC-1 exhibited greater apparent molecular weights on these gels than as were expected. Knowing the plasmids contained the correct DNA sequences for the two sSDC-1 ectodomains, we then moved to purification of the protein.

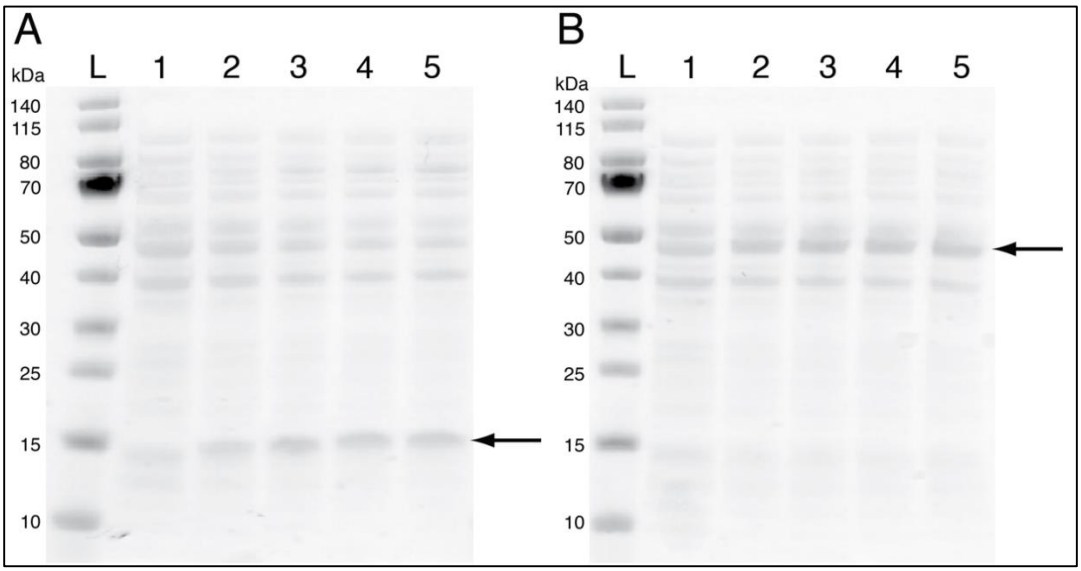

Figure 2.8: Expression of the two SDC-1 constructs. The expression of SDC-1_{1-82} (A) and SDC-1_{1-245} (B) can be seen in the two gels. Lane L contains the protein ladder with the sizes (in kDa) marked to the left of the gels. Lane 1 in each gel shows the endogenously expressed proteins of the bacteria prior to induction of recombinant protein production with IPTG. The following four lanes for each protein represent the 1hr, 2hr, 3hr, and 4hr time points after IPTG induction. Arrows highlight the band representing the overexpressed protein in both gels. The apparent molecular weight for the two fusion protein constructs are ~15 kDa and ~50 kDa respectively for SDC-1_{1-82} and SDC-1_{1-245}.

The purification of the His-tagged sSDC-1 fusion protein was made relatively easy with the use of nickel affinity chromatography. After cell lysis using sonication, we saved the resulting supernatant solution. Because we were working with the SDC-1 ectodomain sequences, we were able to make the assumption that the protein would likely reside in the soluble portion of the cell

lysate. We first purified the two constructs by adding a total of 15 mL of cell lysate to the 5 mL column and washing this with 15 mL of wash buffer containing 50 mM imidazole. After washing, we then eluted the protein from the column, collecting 1 mL fractions. We then ran an electrophoresis gel of the collected fractions **(Figure 2.9)**. The loss of protein in the flowthrough fraction of the purification was evident with both protein constructs and especially pronounced for the SDC-$1_{1\text{-}245}$ construct. To counter this loss of protein, we decreased the amount of protein being loaded onto the column by half in all subsequent purifications. Using this method, we still encountered some loss in the flowthough fraction. In order to optimize binding of the protein to the Ni-NTA, we removed ethylenediaminetetraacetic acid (EDTA) from the affinity column buffers to ensure the columns retained the strongest binding affinity possible. EDTA acts to inhibit protease and DNAse activity during cell lysis by chelating the essential metal cations required by these molecules. These same chelating properties of EDTA can also strip the affinity columns of nickel ions even at low concentrations and may affect the binding strength of the column.

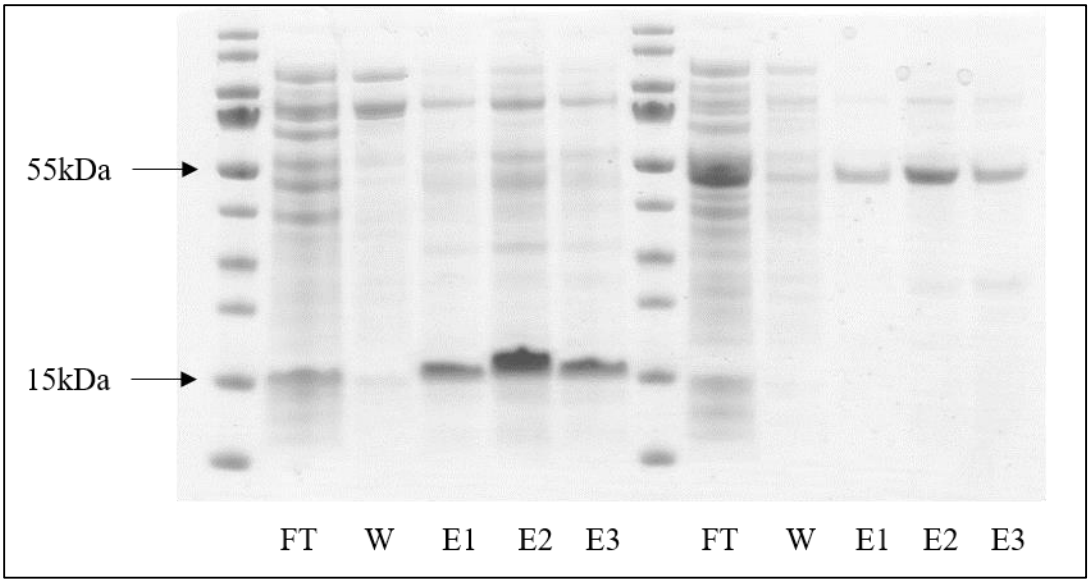

Figure 2.9: Initial Purification Trials. In this gel, the flowthrough fractions (FT) showed a loss of target protein for both constructs, however the SDC-$1_{1\text{-}245}$ (right) had a noticeably more pronounced loss in the flowthrough. The two wash fractions (W) and three elution fractions (E1-3) are also pictured on this gel. While the target protein was by far the most abundant in the elution fractions, there were still several other protein impurities coming through.

It was seemingly evident that the SDC-$1_{1\text{-}245}$ construct was associating much more weakly with the column than its shorter SDC-$1_{1\text{-}82}$ counterpart. After experimentation with varying concentrations of imidazole in the wash buffers, it was initially determined that the SDC-$1_{1\text{-}245}$ construct began eluting from the column with imidazole concentrations above 25 mM, whereas SDC-$1_{1\text{-}82}$ bound the column favorably with up to 100 mM imidazole in the wash buffer. However, this may not actually be the case: referring to **Figure 2.10 A**, a band of strongly expressed native protein can be found in the SDC-$1_{1\text{-}82}$ gel in the exact same molecular weight range as seen with the purified SDC-$1_{1\text{-}245}$ in **Figure 2.10 B**. This revelation allowed us to move forward and increase the imidazole concentration from 25 mM to 50 mM in the SDC-$1_{1\text{-}245}$ wash buffer to enhance the purity of the isolated protein. The gel results from our optimized purification process seen in **Figure 2.10** show that we were able to isolate our two fusion proteins with a very high degree of purity. The purity of our proteins were also later confirmed by both electrospray mass spectrometry and NMR.

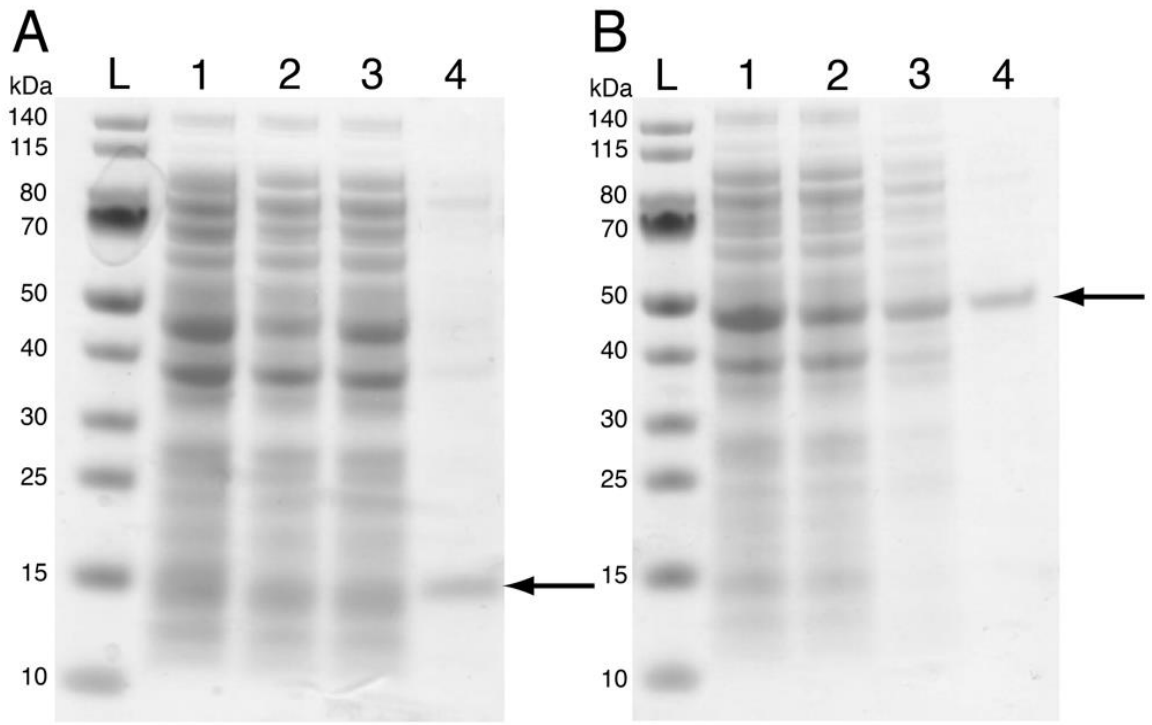

Figure 2.10: Optimized Purification. In these two gels, the sample loaded into lane 1 was cell lysate for each construct. Lane 2 represents the flowthrough fraction and lane 3 is the wash fraction. Lane 4 in each gel depicts purified SDC-$1_{1\text{-}82}$ (A) and SDC-$1_{1\text{-}245}$ (B).

It is also important to note that during the refinement of the purification process, the gel results still showed higher than expected masses for both protein constructs. The expected masses were 12 kDa and 28.7 kDa for SDC-$1_{1\text{-}82}$ and SDC-$1_{1\text{-}245}$ respectively, while the observed values were roughly 15 kDa for SDC-$1_{1\text{-}82}$ and 50 kDa for SDC-$1_{1\text{-}245}$. Being as both constructs contain a single cysteine residue, we hypothesized there may be some disulfide bond formations leading to the protein exhibiting itself in a dimer form on the electrophoresis gels. To address this, we lysed and purified the protein in the presence of reducing agent. In separate experiments we used 5 mM tris(2-carboxyethyl)phosphine (TCEP) and 5 mM dithiothreitol (DTT). The gel results of these experiments still exhibited the same higher than expected apparent molecular masses for both constructs **(Figure 2.11)**.

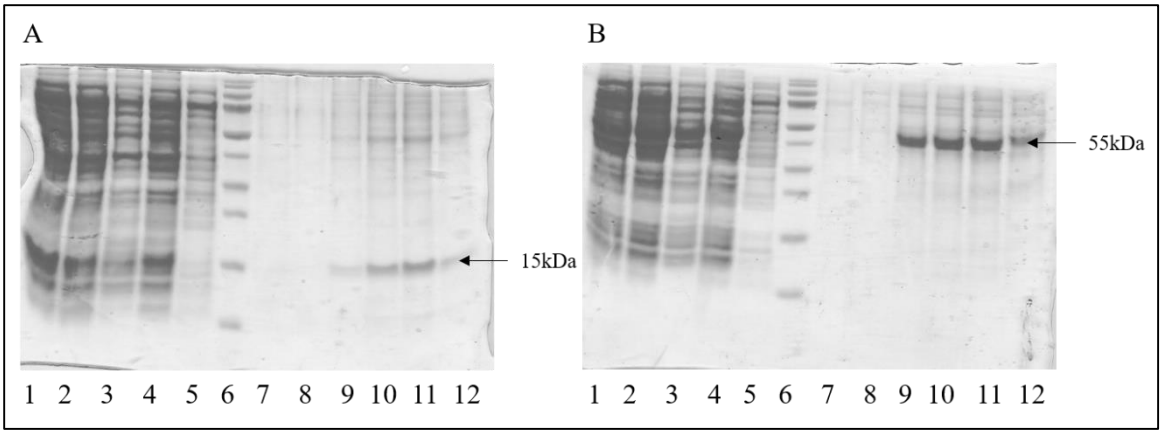

Figure 2.11: Gels with TCEP. Even when ran in the presence of a reducing agent, SDC-$1_{1\text{-}82}$ (A) and SDC-$1_{1\text{-}245}$ (B) both have a high apparent MW. Protein lysates prior to the addition of TCEP (lane 1) exhibit no noticeable changes after the addition of the reducing agent (lane 2). Lane 3 represents the 3-4 mL fraction of column flowthrough, while lane 4 is the 14-15 mL fraction. The noticeable increase in lost target protein in the latter part of the flowthrough lead us to decrease the amount of protein loaded onto the column by half. Lanes 9-12 show that our target proteins come off the column in the 2-7 mL fractions of the elution.

With the results of DNA sequencing we knew that the cell lines should both exhibit correct recombinant expression of the sSDC-1 constructs. Purification results showed that each line was indeed overexpressing a protein containing the required His-tag based off of the strong binding affinity of the isolated proteins. Knowing this, we moved on to deploying the use of mass spectrometry to determine if the purified samples contained the sSDC-1 proteins of the calculated molecular weights. Mass spectrometry results shown in **Figure 2.12 and Figure 2.13** demonstrated that the correct protein constructs had been expressed and purified from the recombinant cells. Comparing our results with previously published data on full-length SDC-1, it became apparent that other groups who have worked with full-length SDC-1 have also experienced this same anomaly with the apparent molecular mass on SDS-PAGE gels.[79-81]

The expected molecular masses of the two SDC-1 constructs were 12,026 Da and 28,700 Da. The expected molecular masses without the N-terminal methionine residue were 11,895 Da and 28,570 Da respectively. These latter expected masses aligned perfectly with the results from the in-tact protein mass spectra in **Figure 2.13**. It has been demonstrated that there is a high

likelihood of N-terminal methionine excision in *E. coli* when the methionine is immediately followed by an alanine residue.[82] This is likely the case for our protein constructs and explains the differences between expected and observed molecular weights.

After confirmation of the correct molecular mass for both constructs using mass spec, we moved on to TEV cleavage of the fusion proteins and subsequent separation of the cleavage products. We began using TEV lysis and purification methods as described by Tropea et al., but then switched to our own methods. Mass spectral analysis of the purified sSDC-1 after cleavage using the Tropea method vs. our own adopted method of TEV purification indicated similar protease activity for both. Our method of lysis and purification was adopted as it was simplified and less time consuming. For our purposes, it is unnecessary to add polyethylenimine to the sample prior to mechanical lysis. The Tropea method of purification also calls for Ni-NTA affinity purification followed by size exclusion chromatography. We found that, by appropriate adjustment of imidazole concentrations in the equilibrium, wash, and elution buffers (40 mM, 80 mM, and 500 mM respectively) during affinity chromatography, we were able to isolate the TEV protease without the need for further purification. The consensus sequence of ENLYFQG included in the SDC-1 fusion proteins is recognized by the TEV protease and cleaved between glutamine and glycine, leaving a single glycine residue attached to the N-terminal methionine of the two sSDC-1 core proteins. As seen in the mass spectrometry graphs in **Figure 2.12**, TEV cleavage of the sSDC-1 fusion protein was successful for both constructs. Graphs prior to cleavage show well isolated peaks for both fusion protein constructs, while graphs after cleavage are more noise-ridden. This could be due to impurities in the samples as the MALDI spectra depicted were taken before complete refinement of the purification process. This could also be due to the nature of the protein itself after TEV cleavage. After cleavage, the proteins lose several highly hydrophilic and ionizable amino acid residues. This could contribute to poor flight characteristics of the protein ions in a mass analyzer. The electrospray ionization results shown in **Figure 2.13** were the result

of ~200 spectra averaged together and exhibit much cleaner protein signals. Although many spectra were averaged here, it is important to note that nearly all individual spectra were also very clean. The only reason spectral averaging was utilized was to combat issues with spray stability. Unfortunately, we were not able to record ESI spectra of SDC-1 after TEV cleavage. Further experimentation is needed to find the proper sample conditions required to examine the cleaved sSDC-1 constructs via electrospray. Again this is likely due to the absence of the hydrophilic and easily ionizable residues in the fusion tag.

Also of note in **Figure 2.12**, our SDC-$1_{1\text{-}82}$ construct exhibited two distinct peaks both pre-cleavage and post-cleavage. This is likely due to a premature ribosomal detachment between the pair of glutamine residues located at positions 89 and 90 on the SDC-$1_{1\text{-}82}$ fusion protein sequence. The phenomenon of ribosomal drop-off with high levels of recombinant protein expression has been previously documented and is mRNA specific.[83] To counter this, in the future, the SDC-$1_{1\text{-}82}$ plasmid may need to be incorporated into a different *E. coli* strain such as BL21 Star™ (DE3) that is genetically engineered to promote mRNA stability.

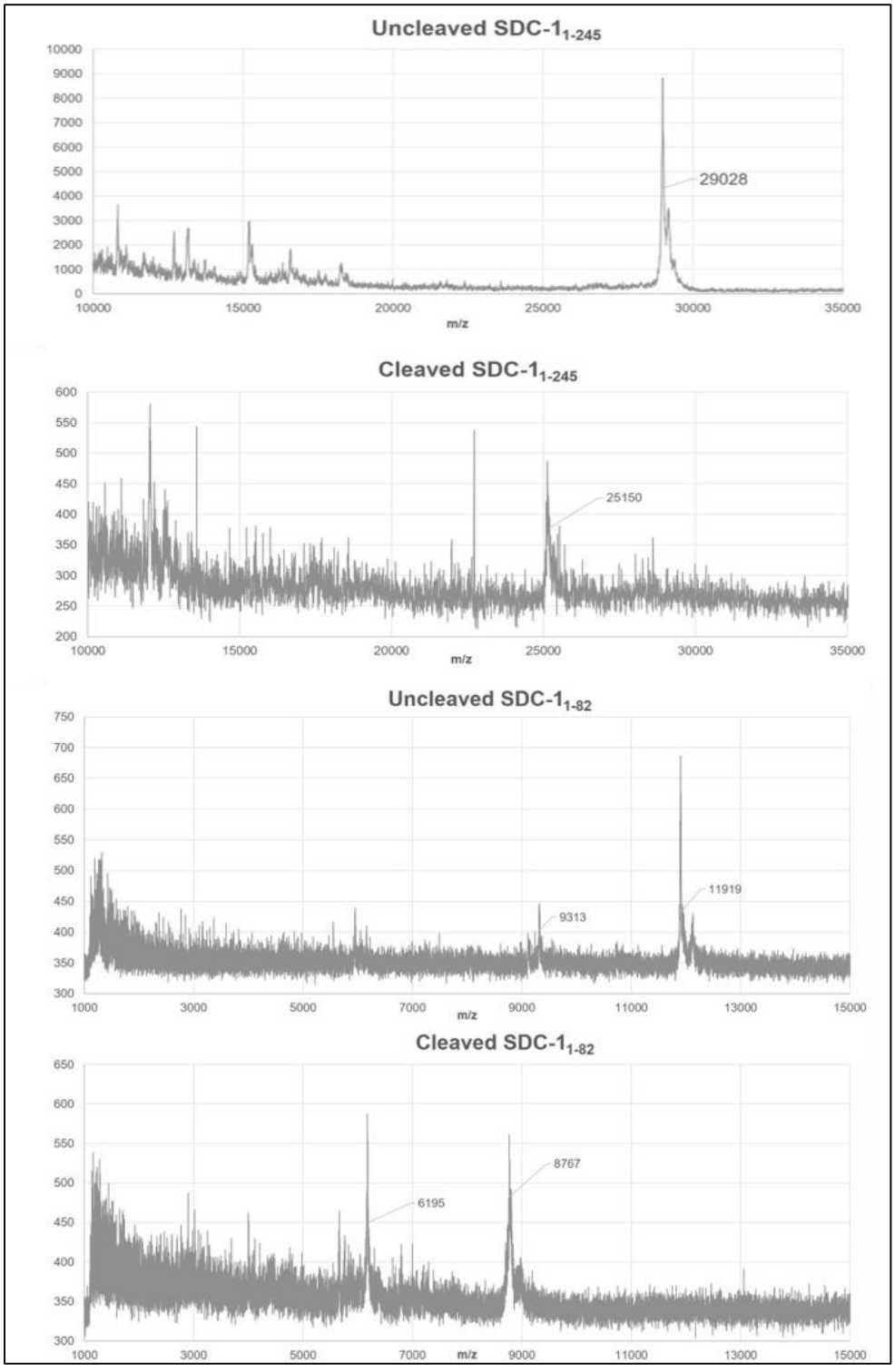

Figure 2.12: MALDI-TOF Results. In this figure, we can see the masses of our two SDC-1 constructs line up well with the expected molecular weights of 12 kDa for SDC-1_{1-82} and 28.7 kDa for SDC-1_{1-245} before cleavage. After cleavage, our expected MWs were 8.7 kDa and 25.4 kDa respectively. The resulting MWs after cleavage are also in good agreement with what we expected.

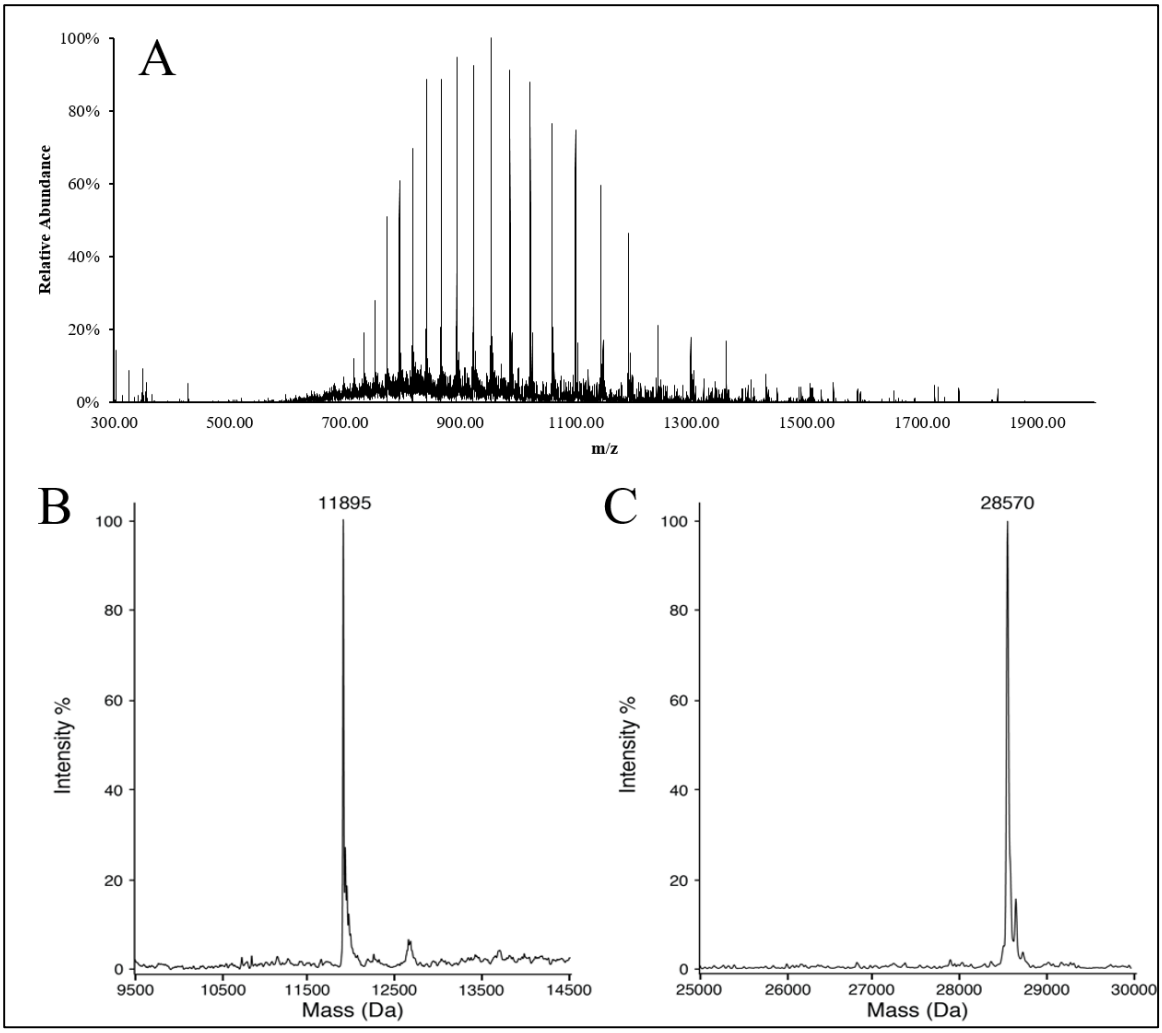

Figure 2.13: ESI-MS Results. In this figure, we can see an example of (A) the spectra showing multiple charge state distributions for one of our syndecan-1 constructs. When zoomed in, it is also possible to see individual isotopic distributions due to the natural abundance of several isotopes in our proteins. These spectra were then processed in a deconvolution software to generate the zero charge mass spectra of (B) SDC-$1_{1\text{-}82}$ and (C) SDC-$1_{1\text{-}245}$.

2.4 Discussion

The results of work done thus far show we have successfully developed a reliable and relatively simplistic method of production of recombinant SDC-1 ectodomain core protein in *E. coli* expression cells. Although bacterial cells lack the mechanisms required to add sugars to glycoproteins and research has proven the presence of GAG sidechains is highly important for the native function and binding properties of the SDC-1 ectodomain; we believe it is best to begin our understanding of this vital HSPG through studies of the core protein. As previously stated, if the heparan sulfate chains are lost, the core protein becomes more susceptible to cleavage factors. Therefore, it can be inferred that oftentimes when sSDC-1 levels are elevated, this may be in the form of the de-glycosylated core protein.

The results of a multitude of studies are in agreement that the syndecan family is highly enriched in intrinsic disorder. The proteins are able to adapt their structural conformity to fit various binding partners. Because of this intrinsic disorder, little experimental structural analysis data is available. Currently, no experimental NMR structural analysis of SDC-1 exists. Most, if not all, studies performed thus far involving SDC-1 have utilized protein isolated from eukaryotic cell lines. By growing the ectodomain of SDC-1 in E. Coli cells, we are able to reliably produce and purify homogeneous protein without the possibility of heterogeneous GAG sidechain formation. This provided us with the ability to resolve the first ever NMR spectra of the SDC-1 ectodomain described in the following chapter. The work done here has provided us with the ability to easily express and isolate SDC-1 ectodomains to be used in future studies involving the characterization and glycosylation of the protein.

Abbreviations: heparan sulfate proteoglycans (HSPGs); heparan sulfate (HS); glycosaminoglycan (GAG); syndecan-1 (SDC-1); matrix metalloproteinases (MMPs); soluble syndecan-1 (sSDC-1); X-ray crystallography (XRC); integral membrane proteins (IMPs); polymerase chain reaction (PCR); isopropyl β-D-thiogalactopyranoside (IPTG); Luria Broth (LB); optical density at 600nm (OD_{600}); molecular weight cutoff (MWCO); Tobacco Etch Virus (TEV); nitrilotriacetic acid (NTA); ethylenediaminetetraacetic acid (EDTA); tris(2-carboxyethyl)phosphine (TCEP); Dithiothreitol (DTT); Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF); electrospray ionization (ESI); acetonitrile (ACN); trifluoroacetic acid (TFA); mass to charge (*m/z*)

CHAPTER III

IN VITRO GLYCOSYLATION OF SDC-1 ECTODOMAINS

3.1 Introduction

3.1.1 Post-Translational Modifications of Proteins

There are only a certain number of genes coding for a limited number of proteins in any given genome. Proteins themselves consist of only 20 different amino acids, limiting their diversity. To further diversify the structure, function, and dynamics of the proteome, organisms can perform various post-translational modifications (PTMs) on the proteins that are produced. As the name implies, these modifications are chemical changes that a protein may undergo after translation.[84] PTMs are mainly covalent processing events which can be either reversible or irreversible and often occur through the use of specific enzymes.[85] Some examples of reversible PTMs include acetylation, amidation, hydroxylation, oxidation, phosphorylation, and glycosylation. Proteolytic cleavage events are one example of an irreversible PTM.[86] Post-translational modifications can take the 20 primary amino acids and diversify them into over 140 different residues as shown in **Table 1**.[87] Some proteins may undergo more than one different PTM and proteins that can undergo many PTMs are often also enriched in intrinsic disorder.

This is likely due to the reversible nature of many PTMs and the fact that while most PTMs are site or sequence dependent, PTMs such as glycosylation are also highly diverse and dependent upon various cellular conditions at the time of modification or sugar polymerization.[86]

Post translational modifications play a pivotal role in a broad spectrum of biological processes.[88] For example, acetylation of histones has been linked to transcriptional activation, gene silencing, DNA repair, and cell-cycle progression. Whereas phosphorylation of histones has been associated with transcription regulation, apoptosis, developmental gene regulation, and heat shock induced pathways.[89] Ubiquitination is another PTM in which proteins are marked by the ubiquitin protein. Ubiquitination is often used to tag improperly folded or aggregated proteins and is most commonly known for being the first step in the proteasomal degradation pathway. Dysfunction in the ubiquitin proteasome system has been linked to Alzhiemer's, Perkinson's, and Huntington's diseases.[7, 90]

Analysis and understanding of the direct effects of the various PTMs on proteins is highly important for development of novel methods to combat some of the most common ailments in the world such as heart disease, cancer, and neurodegenerative diseases. PTMs are vastly diverse and the study of PTMs on membrane proteins is made particularly difficult due to the amphiphilic nature of membrane proteins. PTMs *in vivo* are highly diverse and can have slight variations even between two identical protein sequences expressed in a single cell. This implies that many proteins isolated from cells capable of *in vivo* PTMs may be heterogeneous in nature and therefore difficult to study structural effect via NMR spectroscopy. An *in vitro* approach to studying the various PTMs holds great promise in determining direct effects of specific PTMs as *in vitro*-modified proteins can be made highly homogenous and allow for high resolution NMR spectroscopy and X-ray crystallography techniques.

Table 3.1: Post-Translational Modification List. The following table is a comprehensive list of all known PTMs that are known to modify the 20 amino acids along with example proteins that exhibit these PTMs.

Residue	Reaction	Example	Residue	Reaction	Example	Residue	Reaction	Example
Alanine (Ala, A)	N-Acetylation	*N*-Alpha-acetyltransferase	Cysteine (Cys, C)	S-Hydroxylation (S-OH)	Sulfenate intermediates; peroxiredoxins	Glutamine (Gln, Q)	Transglutaminatio n	Protein cross-linking
	Amidation	Pantothenate synthetase		Disulfide bond formation	Protein in oxidizing environments		Deamidation	Myelin basic protein
	N-Methylation	Ribosomal proteins		Phosphorylation	PTPases		Amidation	FMRFamide-related peptides
Arginine (Arg, R)	N-methylation	Histones		S-Acylation	Ras		N-Methylation	Ribosomal proteins
	N-ADP-ribosylation	GSa		S-Prenylation	Ras	Glycine (Gly, G)	C-Hydroxylation	C-terminal amide formation
	Citrullination/Deimination	Argininosuccinate synthase		Protein splicing	Intein excisions		N-Acetylation	*N*-Alpha-acetyltransferase
	N-Acetylation	*N*-Alpha-acetyltransferase		N-Acetylation	*N*-Alpha-acetyltransferase		Amidation	Glycine oxidase
	Amidation	Tachykinins		N-ADP-ribosylation	Glyceraldehyde-3-phosphate dehydrogenase		Cholesterol glycine ester	Hedgehog proteins
	Dihydroxylation	Steilins		Amidation	Cystein synthase A		N-Myristoylation	Protein Nef
	Hydroxylation	Carbon monoxide dehydrogenase large chain		S-Archaeol cysteine	Halocyanin	Histidine (His, H)	Phosphorylation	Sensor protein kinases in two-component regulatory systems
	Phosphorylation	Histones		Cysteine sulfinic acid (-SO2H)	Cysteine sulfinic acid decarboxylase		Aminocarboxypropylation	Diphthamide formation
Asparagine (Asn, N)	N-Glycosylation	N-Glycoproteins		Methylation	Crystallins		N-Methylation	Methyl CoM reductase
	N-ADP-ribosylation	eEF-2		N-Myristoylation	Genome polyproteins of several viruses		Amidation	VIP peptides
	Protein splicing	Intein excision step		Nitrosylation	Thioredoxins		Bromination	Sperm-activated peptide SAP-b
	Deamidation	Isomerization to isoaspartate and aspartate		N-Palmitoylation	Small cystein-rich outer membrane protein OmcA		Methylation	Actin
	Amidation	FMRFamide-related peptides		S-Palmitoylation	Myelin proteolipid proteins	Isoleucine (Ile, I)	Amidation	FMRFamide neuropeptides
	Hydroxylation	Hypoxia-inducible factor 1-alpha		S-Glutathionylation	Redox regulation via reversible glutathionylation		N-Methylation	Fimbial protein

Table 3.1: ***continued***

Residue	Reaction	Example	Residue	Reaction	Example	Residue	Reaction	Example
Aspartic acid (Asp, D)	Phosphorylation	Protein tyrosine phosphatases; response regulators in two-component systems	Glutamic acid (Glu, E)	Methylation	Chemotaxis receptor proteins	Lysine (Lys, K)	N-Methylation	Histone methylation
	Isomerization to isoaspartate	Protein-l-isoaspartate *O*-methyltransferase		Carboxylation	Gla residues in blood coagulation		N-Acylation by acetyl, biotinyl, lipoyl, ubiquityl groups	Histone acetylation; swinging-arm prosthetic groups; ubiquitin; SUMO (small ubiquitin-like modifier) tagging of proteins
	Deamidation	Beta-casein		Polyglycination	Tubulin		C-Hydroxylation	Collagen maturation
	N-Acetylation	*N*-Alpha-acetyltransferase		Polyglutamylation	Tubulin		O-Glycosylation	Adiponectin; *O*-glycoproteins
	Beta-methylthiolation	Ribosomal proteins		N-Acetylation	*N*-Alpha-acetyltransferase		N-acetylation	*N*-Alpha-acetyltransferase
	Hydroxylation	3-Hydroxyaspartate aldolase		Poly N-ADP-ribosylation	Poly (ADP-ribose) polymerase 1		Allysine	Elastin and collagen; lysyl oxidase
	cis-14-Hydroxy-10,13-dioxo-7-heptadecenoic acid aspartate ester	Nonspecific lipid transfer proteins		Amidation	Buccalin		Amidation	Histone-lysine *N*-methyltransferase EHMT1
Methionine (Met, M)	Oxidation to methionine sulfoxide	Methionine sulfoxide reductase		Deamidation followed by methylation	Methyl-accepting chemotaxis proteins		N6-1-carboxyethylation	Carbonyl reductases
	Oxidation to methionine sulfone	Catalase	Tryptophan (Trp, W)	C-Mannosylation	Plasma-membrane proteins		Dihydroxylation	Steilins
	Silent modification (conversion to aspartic acid)	Unstable hemoglobin, Hb Bristol [p67(E11) Val-Met]		Amidation	Neuropeptide-like proteins		Hydroxylation	Collagens
	N-Acetylation	*N*-alpha-acetyltransferase		Bromination	Mu-conotoxins		N-Myristoylation	Tumor necrosis factors
	Amidation	MIP-related peptides		C-Linked glycosylation	Properdin		N-Palmitoylation	Serine palmitoyltransferases
	N-Methylation	Ribosomal proteins		Hydroxylation	Alpha-ketoglutarate-dependent taurine dioxygenase		Trimethylation	Myosins

Table 3.1: ***continued***

Residue	Reaction	Example	Residue	Reaction	Example	Residue	Reaction	Example
Proline (Pro, P)	C-Hydroxylation	Collagen; HIF-1a	Tyrosine (Tyr, Y)	Phosphorylation	Tyrosine kinases/phosphatases	Phenylalanine (Phe, F)	Amidation	FMRFamide neuropeptides
	Dihydroxylation	Virotoxin		Sulfation	CCR5 receptor maturation		Hydroxylation	Adhesive plaque matrix proteins
	N-Acetylation	*N* -Alpha-acetyltransferase		*ortho* -Nitration	Inflammatory responses		N-Methylation	ComG operon proteins
	Amidation	Prothyroliberin		TOPA quinine	Amine oxidase maturation	Leucine (Leu, L)	Amidation	Myomodulin neuropeptides
	N-Methylation	*N* -Methylproline demethylase		N-Acetylation	*N* -Alpha-acetyltransferase		N-Methylation	Major structural subunit of bundle-forming pilus
Serine (Ser, S)	Phosphorylation	Protein serine kinases and phosphatases		Amidation	FMRFamide-related neuropeptides	Threonine (Thr, T)	Phosphorylation	Protein threonine kinases/phosphatases
	O-Glycosylation	Notch O-glycosylation		N-Methylation	General secretion pathway protein I		O-Glycosylation	*O* -Glycoproteins
	Phosphopantetheinylation	Fatty acid synthase		O-Glycosylation	S-layer protein SpaA		N-Acetylation	*N* -Alpha-acetyltransferase
	Autocleavages	Pyruvamidyl enzyme formation	Valine (Val, C)	N-Acetylation	*N* -Alpha-acetyltransferase		Amidation	Aurora kinase A
	N-Acetylation	*N* -Alpha-acetyltransferase		Amidation	MIP-related peptides		N-Decanoylation	Ghrelin
	O-Acetylation	*O* -Acetylserine (thiol) liase		Hydroxylation	Conophans		O-Octanoylation	Ghrelin
	N-ADP-ribosylation	Ras-related protein Rap-1b					O-Palmitoylation	Myelin proteolipid protein
	Amidation	Kallikrein-8					Sulfation	Cathepsin
	N-Decanoylation	Ghrelin					O-Acetylation	Inhibitor of nuclear factor kappa-B kinase subunit alpha
	O-Octanoylation	Appetite-regulating hormones; ghrelin						
	O-Palmitoylation	Myelin proteolipid protein						
	Sulfation	Retrograde protein of 51 kDa						

3.1.2 N-Linked Glycosylation

Glycosylation is arguably the most diverse form of protein post-translational modification. It also happens to be one of the most prominent PTMs with more than 50% of all polypeptides being modified by the addition of oligosaccharides to specific functional groups. Glycosylation is highly important in membrane proteins with a majority of membrane proteins exhibiting glycosylation at one or more sites. In membrane proteins, 99% of N-linked glycosylation sites are located in the extracellular region of the protein.[91] Glycosylation is known to affect protein folding structural integrity, and function.[92] Proper glycosylation is vital for many biological functions and genetic defects involved with glycosylation are often lethal in the embryo. Congenital glycosylation disorders have been linked to a multitude of health issues including defects in muscle development, seizures, growth retardation, hepatic fibrosis, hypoglycemia, developmental delay, mental retardation, immunodeficiency, and early death.[93]

Of the 4 main types of glycosylation, mentioned previously in Chapter 1, N-linked glycosylation is the most abundant.[94] N-linked glycan linkage predominantly occurs through the side chain amide of an asparagine residue in the sequence of Asn-Xaa-Thr/Ser (N-X-T/S) where Xaa is any amino acid other than proline. In very rare cases, N-glycosylation has also been shown to occur on asparagine residues in the sequence of Asn-Xaa-Cys.[91, 95] Although the N-X-T/S sequon is generally necessary for glycosylation to occur, this is not the only criteria.[96] For example, in mouse studies, a major prion protein with two N-linked glycosylation sites can be expressed in both the sciatic nerve and brain cells. This protein favors an unglycosylated form in the sciatic nerve (64% unglycosylated) and a glycosylated form in brain cells (66% diglycosylated).[97] This suggests that N-linked glycosylation is not only site-dependent, but also cell-type dependent. N-linked glycosylation sites are also secondary-structure-dependent with around 75% of glycosylated asparagine residues located in loop structures; suggesting glycan attachment is favored on the protein surface. [91]

N-glycans are assembled from nucleotide-linked building blocks on a lipid anchor via stepwise incorporation of monosaccharides by various glycosyltransferases. In eukaryotes, this process is performed on the cytosolic side of the endoplasmic reticulum (ER) membrane. After assembly, the lipid-anchored oligosaccharide is re-oriented to the luminal side of the ER where it serves as a glycosylation donor. A membrane-bound enzyme known as oligosaccharyltransferase (OST) then catalyzes the *en bloc* transfer of the oligosaccharide to the side chain amide of an asparagine in the N-X-T/S sequon of a glycoprotein. This process occurs co-translationally during protein biosynthesis on nascent polypeptide chains prior to final protein folding. Upon covalent linkage of the glycan to a protein, this glycan can be further modified by various processes in eukaryotes resulting in both species-specific and even cell type-specific diversity of the N-glycan.[98]

An alternative pathway for N-linked glycosylation has been characterized in bacterial cells. In this pathway, cytoplasmic N-glycosyltransferase catalyzes the attachment of a nucleotide-activated monosaccharide to an asparagine residue within the same consensus sequence of N-X-T/S. This results in a single glucose residue attached through the amide of the asparagine. Other glycosyltransferase enzymes have been identified that have the ability subsequently to polymerize the N-linked glucose.[99]

3.1.3 Literature Review

It has been established that glycosylation and glycoproteins play a vital role in biological systems. This has given rise to the fields of both glycomics and glycoproteomics. The term “glycomics” describes studies designed to define and characterize the complete set of glycans (or glycome) that a cell, tissue, or organism produces under specified environmental conditions. Whereas glycoproteomics describes this glycome in relation to the cellular proteome. The field of glycoproteomics attempts to determine which sites are glycosylated on the various glycoproteins along with the identification and quantitation of the glycan structures attached at these various

glycosylation sites.[100] Proteomics and glycomics employ the use of mass spectrometry to identify and quantitate the various proteins, glycans, and glycoproteins present in cellular biology. While the information gathered is important for a broad understanding of glycans, glycoproteins, and the various functions of both; mass spectrometry yields very little information about specific structural interactions of glycoproteins.

Many different analytical methods have been developed that aid in protein structural characterization. Circular dichroism (CD) spectroscopy, one of the most common methods used for secondary structure characterization, examines the ways a molecule absorbs left and right polarized light. Ultraviolet CD is used to quickly assess whether a protein secondary structure consists of mainly α-helix, β-sheet, or random coil structures.[101] Infrared difference spectroscopy has also proven useful for determining various protein secondary structure by examining vibrational changes of proteins in response to perturbation.[102] The two most-powerful tools for examining protein structure remain X-ray crystallography and NMR spectroscopy. However, due to the often heterogeneous nature of glycans and glycoproteins, the direct study of glycoprotein structure and dynamics presents many challenges. These issues are further compounded when studying membrane glycoproteins due to the amphiphilic nature of these proteins.

In recent years, the use of *in vitro* glycosylation for the study of glycoprotein structure and dynamics has shown great promise. Though the predominant pathway of N-linked glycosylation in eukaryotes is through OST, this membrane-bound enzyme has proven ineffective for glycosylation proteins after final folding. Consequently, the use of OST for *in vitro* glycosylation has been largely unsuccessful and requires the unfolding of the protein, disruption of protein secondary structure, or even fragmentation of the protein for glycosylation to occur.[103, 104] In addition, OST requires complex glycosyl donors that are largely inaccessible for large-scale *in vitro* glycosylation studies. Utilization of microsomal and solubilized OST has proven to have an increased enzymatic glycosylation activity, but only when using fairly short peptides as glycan

acceptors. Large proteins, even when chemically modified, have proven to be poor substrates for OST-catalyzed glycosylation and therefore cannot be converted to homogenous glycoproteins.[105]

N-linked *in vitro* glycosylation has been proven possible using the cytosolic N-glycosyltransferase (NGT) from *Actinobacillus pleuroneumoniae* described by Schwarz et al. This enzyme integrates some of the features of OST glycosylation to recognize the N-X-T/S consensus sequon. Alteration of this sequon to Q-L-T or N-P-T completely abolished glycosylation activity. Most importantly, NGT has the ability to attach a glucose to the asparagine residue *in vitro*. NGT was proven effective for achieving glycosylation *in vitro* for both synthetic acceptor peptides and the full-length AcrA protein, a common glycoprotein substrate of both bacterial and eukaryotic OSTs. These results were confirmed using both MS/MS analysis and solution NMR spectroscopy.[99]

In the Cook laboratory we have previously demonstrated that enzymatic activity of NGT is still present even in a detergent/lipid environment. This has allowed our lab to become the first to demonstrate *in vitro* glycosylation on both hydrophobic peptides and full-length integral membrane proteins in a lipid/detergent environment.[106] *In vitro* glycosylation using NGT has proven to be quantifiable using both LC-MS/MS and solution NMR with unpublished Cook Lab data showing NGT glycosylation efficiencies of >40%.

3.1.4 Summary and Outlook

The importance of post-translational modifications has been well known for decades. Glycosylation is one of the most prominent types of post-translational modifications and the importance of proper glycosylation of proteins has been well established. We know that glycosylation is highly important for proper functionality of glycoproteins. We also know that improper glycosylation or lack of glycosylation can lead to various different disease states.

However, attempts to study the direct effects of glycosylation on proteins using methods such as solution NMR have thus far proven relatively ineffective. This is due to the highly heterogeneous nature of glycosylation. To counter these issues with sample heterogeneity, most glycoprotein research has been performed by isolating a glycoprotein, cleaving the glycan chains from the protein, and subsequently examining the glycans and core proteins separately.[107]

In this chapter, I will discuss our methodology for the *in vitro* glycosylation of ectodomains from the SDC-1 membrane protein. By attaching sugars to a protein *in vitro,* we are able to establish control over exactly which sugar is being attached to the protein. As described in the previous chapter, SDC-1 is an integral membrane protein that is normally highly glycosylated in nature. Prior to the work performed in this dissertation, no researchers have ever attempted to glycosylate SDC-1 *in vitro*. Our work with *in vitro* glycosylation holds great promise for analysis of the role that glycosylation plays in membrane protein structural biology. The work done here will provide the highly homogenous samples required for determining the direct effects of glycosylation on the protein by solution NMR (Chapter 4).

3.2 Materials and Methods

3.2.1 *In Vitro* Glycosylation of SDC-1

For *in vitro* glycosylation experiments, our sSDC-1 constructs were first purified via Ni-NTA chromatography. The NGT enzyme used in the glycosylation was purified fresh via Ni-NTA chromatography the day before it was to be used in glycosylation and dialyzed overnight against glycosylation buffer. The Ni-NTA elution containing our purified sSDC-1 glycosylation target was dialyzed vs. 5 L of NGT glycosylation buffer (25 mM Tris, 150 mM NaCl, pH 8.0). The dialysis buffer was changed once at 30 minutes and again after 2 hours before being left to dialyze overnight. Concentrations of the protein were again determined using the A_{280}. For

glycosylation experiments, NGT was mixed with the target protein in a ~2:1 molar ratio while UDP glucose was used in a ~1000 molar excess.

For mass spectrometric experiments on unlabeled sSDC-1 protein, the concentrations of both proteins were made to be 0.3 mg/mL. These samples were evenly-split into two separate 250 µL aliquots and mixed with 250 µL of 25 µM NGT enzyme in glycosylation buffer. To start the glycosylation, 1 mg of UDP glucose was added to the 500 µL reaction mixture and the mixture was left to rotate overnight at room temperature (RT). For control samples, sSDC-1 and NGT were mixed without the addition of UDP glucose. Glycosylation was then observed using bottom-up LC-MS/MS analysis.

3.2.2 Sample Preparation and Enzymatic Digestion of SDC-1

Prior to LC-MS/MS analysis, glycosylated and unglycosylated control samples of sSDC-1 were ran on separate electrophoresis gels. These samples were ran on separate gels to prevent any possible cross-contamination during the electrophoresis process. After electrophoresis, gels were stained with coomassie blue and subsequently de-stained. Target protein bands were excised from the gels by razor blade and cut into ~1 mm^2 pieces. These gel pieces were then placed into 1.5 mL Eppendorf tubes with ~150 µL of gel pieces per tube. Gel pieces were then washed by adding 1 mL of cold 25 mM ammonium bicarbonate (ABC) in 50 % acetonitrile (ACN) and rocked for 1 hour. This washing step was repeated 3x and after the final wash, the wash solution was removed and gel pieces were stored overnight at -20 °C.

The following day, gel pieces were dehydrated by incubation in 1 mL of 100% ACN for 20 min. After removing the ACN with a pipette, the tubes were left open to dry at RT for 30 min. To reduce any disulfide bonds, the gels were then rehydrated in a solution containing 0.9 mL of 50 mM ABC with 0.1 mL of 28.9 mg/mL TCEP. Tubes were then vortexed and rocked at RT for 1 hour. After aspirating the reducing solution from the gels, alkylation was performed by adding

200 µL of 1.8 mg/mL 2-Iodoacetamide in 10 mM ABC. The alkylation reaction was then incubated at RT for 1 hour in the dark. The alkylation solution was then aspirated off and gels were rinsed 2x with 1 mL of 10 mM ABC. After rinsing, the gels were again dehydrated with 100% ACN, dried at RT, and then placed on ice.

Enzymatic digestion was tested using trypsin, chymotrypsin, and V8 in different samples.

For trypsin and chymotrypsin digestion:

300 µL of trypsinolysis solution (8 µg/mL trypsin in 10 mM ABC) was added to the tubes containing the dehydrated gel. The gel was allowed to swell on ice for 45 minutes. Afterwards, excess trypsin solution was discarded and replaced with just enough 10 mM ABC to cover the tops of the gels.

Chymotrypsin digestion was performed using the same method, but with 8 µg/mL of chymotrypsin in the digestion solution.

For V8 digestion:

V8 protease was removed from -80 °C storage and diluted to a final concentration of 17 µg/mL in cold buffered phosphate (0.1 M $NaxPO_4$, pH 7.0). 100 µL of the V8 solution was deposited into each tube and allowed to swell on ice for 30 minutes. Excess V8 solution was then aspirated and just enough buffered phosphate was added to wet the tops of the gel cubes.

All three digestions were allowed to proceed overnight at 37 °C.

The following day, the liquid from each digest was pipetted into a clean tube. Then, protein was extracted from the gels by adding 0.3 mL of 0.5% trifluoroacetic acid (TFA) and incubating for 2 hours at RT. Extraction was performed twice more using 0.2 mL of TFA and all extracts were pooled into the same tube. The tubes were then frozen at -80 °C and dried using rotary vacuum evaporation (speed vac).

Peptides were then resolvated with vortexing in 400 µL of 0.1% TFA and desalted using a C18 ZipTip®(Millipore Sigma). The ZT was first wetted 5x in 100% ACN and then 5x with 50% ACN / 0.1% TFA using a pipetman set at 200 µL. The ZT was then washed 10x in 0.1% TFA. Afterwards, 200 µL of the gel extraction solution was pulled into the ZT and using a new clean tube, peptides were bound to the column by 10 passes through the tip (120 µL pipet volume). After the 10^{th} pass, the solution was discarded and 10 more 120 µL passes were made in the original digestion tube to bind the remaining peptides. After peptides were bound to the ZT, the column was washed 5x with 0.1% TFA. After washing, 200 µL of 50% ACN / 0.1% TFA was pulled into the ZT and used to elute the peptides by pipetting up and down 10x in a clean tube. The eluate was then frozen and dried via speed vac.

3.2.3 Bottom-Up LC-MS/MS Analysis of SDC-1 Glycosylation

Peptide samples from digestion were resolvated in 50 µL of a 50% ACN / 5% formic acid mixture. The mix was then placed in an autosampler and 10 µL was used for LC injection. LC-MS/MS analysis was performed on a Thermo Fusion Tribrid mass spectrometer using a nanospray ionization needle and HPLC column.

Instrument settings were calibrated for a DDA analysis of all peptides present in samples. A master scan was acquired using the orbitrap mass analyzer set for a resolution of 120k and a scan range of 375-1575 *m/z* with 2^+ - 6^+ charge state isolation. Maximum injection time was set to 50 ms and AGC target was set to 500k. Second stage MS fragmentation (MS_2) scans were then

performed on parent peptides using both CID and HCD fragmentation in a data-dependent manner with a quadrupole isolation window of 0.8 *m/z*. HCD was performed using stepped collision energies of 30, 35, and 40 %. CID and HCD fragments were then detected by orbitrap at a resolution of 30k with a maximum injection time of 200 ms.

3.2.4 Peptide Mass Fingerprinting for Confirmation of Glycan Attachment

We used Byonic™ to examine the results from our bottom-up LC-MS/MS data. Byonic™ is a software package used for peptide and protein identification by tandem mass spectrometry.[108] We first edited a protein sequence database for all natively-expressed *E. coli* proteins to include the amino acid sequences of our recombinant sSDC-1 proteins. We also added the sequence of NGT to this database which helps to serve as a positive control in the fingerprinting analysis performed by Byonic™. For our instrument parameters, we selected a precursor mass tolerance of 10 ppm and a fragmentation type of both HCD and CID with a fragment mass tolerance of 10 ppm as well. Depending on the protease used to digest the sample, we adjusted our cleavage sites appropriately (Asp and Glu for V8; Lys and Arg for trypsin; Tyr, Phe, Trp, and Leu for chymotrypsin) and used a fully-specific digestion specificity search. For the fixed and variable modifications, we used the commonly recommended modifications. We also added a variable modification of Hexose / +162 at Asn residues, providing for the detection of glycose attached to the glycosylation site of our sSDC-1.

3.3 Results

Our first goal in this study was to establish which digestion enzyme yielded the best peptide fragments to examine the glycosylated ectodomains via bottom-up mass spectrometry. We performed a glycosylation reaction on one of our sSDC-1 constructs as described in the materials and methods section of this chapter. The glycosylation reaction mixture contained all three components required for glycosylation; the sSDC-1 target protein, NGT, and UDP-glucose. Our control samples lacked the required UDP-glucose for the glycosylation reaction. We then ran the products on an SDS-PAGE gel, excised the desired protein bands, and did triple enzymatic digest to determine which protease generates the best peptides for MS/MS analysis. We tested all three proteases that were readily available to us at the Oklahoma State Core facility: trypsin, chymotrypsin, and V8. Trypsin cleaves the protein on the C-terminal side of lysine and arginine residues. Chymotrypsin cleaves on the C-terminal side of tyrosine, phenylalanine, tryptophan, and leucine. V8 cleaves on the C-terminal side of aspartic and glutamic acid residues. Although the three proteases we used are generally highly specific and efficient for enzymatic cleavage of proteins. The efficiency of in-gel digestion is dependent on a number of factors; therefore, some cleavage location may be missed.[109] To confirm cleavage sites and determine the *m/z* of peptides containing the N-linked glycosylation site, we first ran a shotgun analysis on our control samples. This shotgun method allowed us to examine and fragment all novel peptides that were detected by the master MS scan. In a DDA shotgun instrumental method, any time a novel peptide is detected by the master scan, a stream of those peptides are directed toward MS_2 analysis. The instrument classifies novel parent peptides by their individual *m/z* and column elution time. Peptides that have the same *m/z* and that co-elute from the column together have a high likelihood of being identical peptides. Some peptides may exhibit a very similar *m/z* but elute from the column at different times, meaning these are not likely to be identical peptides. Therefore, the instrument will consider them both a novel peptide candidate for MS_2 analysis. MS_2 analysis

provided us a fingerprint of fragments for each peptide. We then used the Byonic™ software to take these fingerprints and generate a series of peptide spectrum matches. The software also calculated the probability of mismatch error for each peptide spectrum. The list of peptide spectrum matches from our first shotgun analysis can be seen for chymotrypsin digestion (**Figure 3.1 A)**, trypsin digestion **(Figure 3.1 B)**, and V8 digestion **(Figure 3.1 C)**. The results showed that trypsin is a poor match for monitoring *in vitro* glycosylation of sSDC-1 as none of the peptides detected included the glycosylation site. While chymotrypsin digestion did exhibit some peptide matches that contain the glycosylation site, the probability of error was relatively large: likely due to the large size of the peptide generated not flying well in MS analysis. V8 protease as the only candidate left is rather unique and generates multiple unique peptide fragments that contain the glycosylation site. It seems that cleavage efficiency of the enzyme was limited due to the high frequency of aspartate residues on the N-terminal side of the glycosylation site.

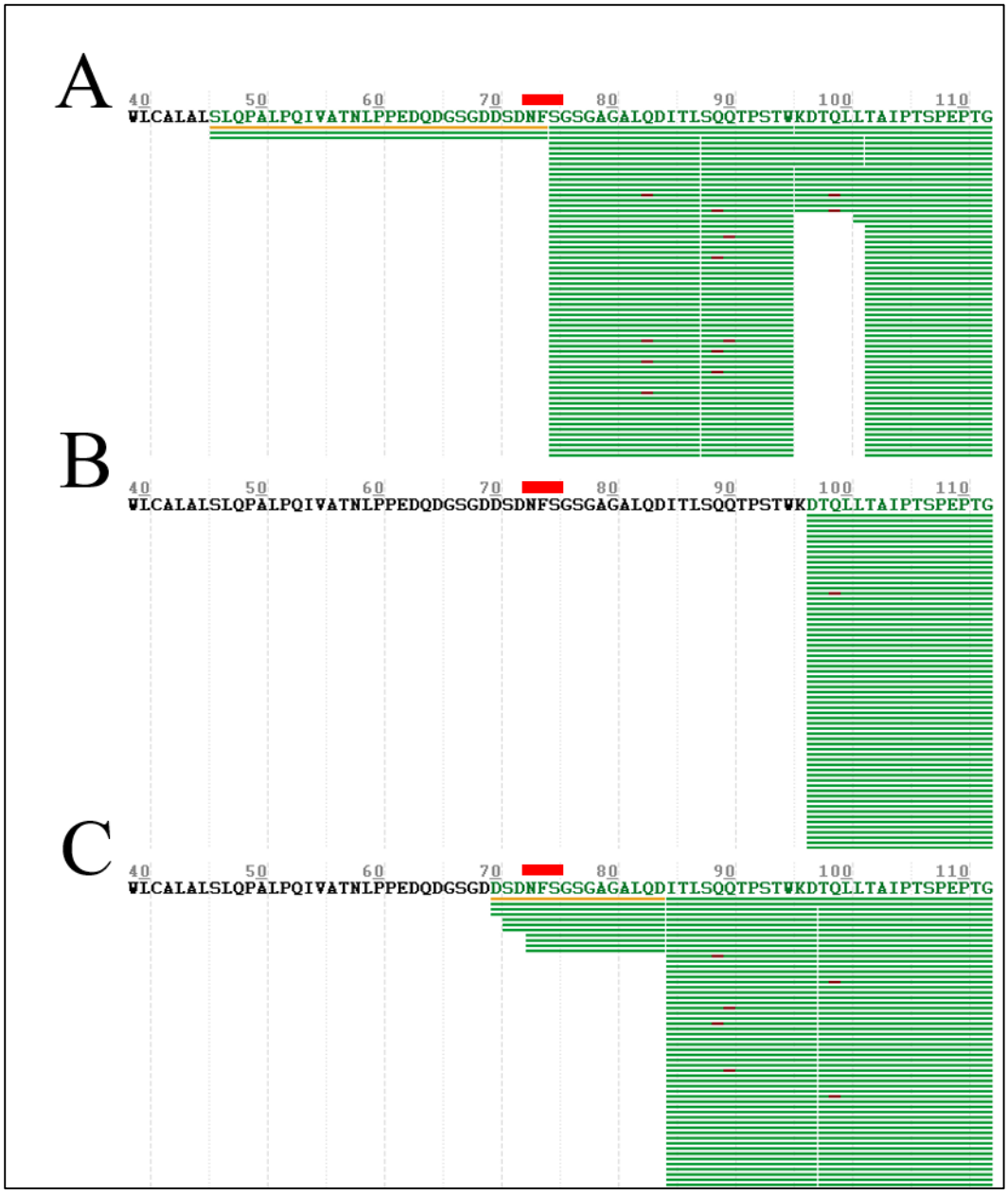

Figure 3.1: Peptide Spectrum Matches for Different Digestion Enzymes. Depicted here are summaries of peptide spectrum matches for (A) chymotrypsin, (B) trypsin, and (C) V8. Each green bar represents an individual peptide spectrum match from MS_2 analysis. The target glycosylation site of (NFS) is denoted by the thick red line.

After selecting V8 as our best protease candidate, we then performed LC-MS/MS on our glycosylation reaction. We first ran an initial shotgun analysis to detect any possible glycosylated peptides. Using the shotgun analysis, we attained definitive proof of success in our efforts to glycosylate sSDC-1 *in vitro* **(Figure 3.2)**. The resulting spectra depict a collection of the different b fragments of the parent peptides. A "b" fragment is a fragment of the parent peptide from the N-terminus to amino acid "b#". For example, our glycosylation site is located on amino acid #8 of the peptide sequence. Therefore, the b8 fragment of our target peptide contains the N-terminus of the peptide through amino acid #8, which is the asparagine glycosylation target. The spectral match for our unglycosylated control sample showed a mass of 748.3 Da for the b8 fragment containing our target asparagine residue. In the glycosylated sample, the mass of the b8 fragment had increased to 910.3 Da. This mass increase of 162 Da shows the glucose had been attached to the asparagine side chain. The masses of the b16 and b17 peptide fragments have also increased by the same 162 Da between the two samples, further proving successful glycan attachment.

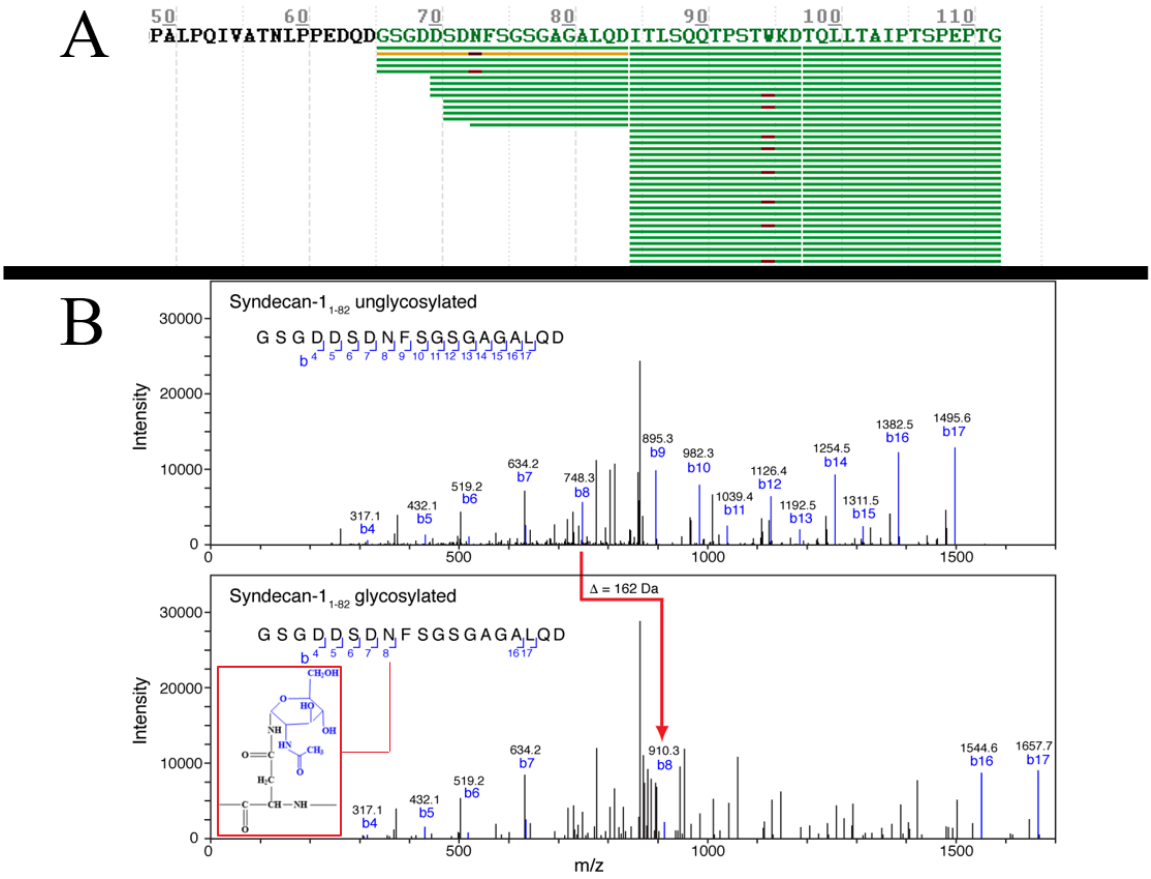

Figure 3.2: Glycosylation of SDC-1$_{1\text{-}82}$. In our shotgun analysis, we were able to detect two different instances representing glycan attachment to the glycosylation site in the resultant list of peptide spectrum matches (A). These are marked by the red dashes directly underneath the glycosylated asparagine for two separately-detected peptides. The peptide spectrum matches (B) for an unglycosylated peptide vs. a glycosylated peptide is shown below the peptide list. The *m/z* of the b8 fragment containing the glycoslation site has been increased by 162 Da from the unglycosylated MS/MS spectra to the glycosylated spectra. The b16 and b17 fragments were likewise shifted by the same 162 Da amount, further proving glycan attachment to the peptide.

3.4 Discussion

Through our work so far in this dissertation with sSDC-1, we have proven that the protein can be glycosylated *in vitro*. In the near future, our lab plans to both quantify and improve the efficiency of this process. This has laid the foundation for countless future studies into the effects of glycosylation on protein structural biology. The sSDC-1 ectodomain is uniquely suited to studying the effects of post-translational modifications on proteins due to the fact that it has only one N-linked glycosylation site and several O-linked glycosylation sites. This will allow us to eventually study the effects of single vs. multiple glycosylation on the structure and dynamics of the protein. By building the sSDC-1 glycoprotein starting with the core protein and subsequently attaching sugars *in vitro*, we have introduced a new level of control for the structural analysis of this complex and disordered protein.

In nature, the probability of attaching only a single glucose to an N-linked glycosylation site is likely relatively low. As we have established, glycosylation is often extremely complex and heterogeneous. The next steps of our *in vitro* glycosylation studies will be trans-glycosylation; an enzymatic replacement of the single glucose residue with a more complex sugar. This will allow us to study the effects due to the attachment of larger and more complex sugar chains to the glycoprotein.

Abbreviations: post-translational modifications (PTMs); endoplasmic reticulum (ER); oligosaccharyltransferase (OST); nuclear magnetic resonance (NMR); Circular dichroism (CD); N-glycosyltransferase (NGT); tandem mass spectrometry (MS/MS); Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE); ammonium bicarbonate (ABC); acetonitrile (ACN); trifluoroacetic acid (TFA); data-dependent analysis (DDA): Second stage MS fragmentation (MS_2)

CHAPTER IV

STRUCTURAL ANALYSIS OF ECTODOMAINS FROM SDC-1

4.1 Introduction

4.1.1 Solution State Protein NMR Spectroscopy

Solution NMR has proven a vital tool for the study of protein structure, interactions, and dynamics.[16] Individual atoms that are assembled into the various amino acids and glycans of glycoproteins can be differentiated from one another due to infinitesimal variations in the chemical environment surrounding these atoms.[18] Spectroscopic experiments described in chapter 1 such as ^{1}H-^{15}N HSQC, HNCA, and HN(CO)CA are made possible through scalar coupling, or J-coupling, which is a through-bond interaction between two nuclei. Scalar coupling provides the ability to examine a series of atoms that are directly connected to one another via covalent bonds. As the connectivity chain between these atoms becomes longer, magnetization transfer between the atoms becomes less efficient. Therefore, experiments using scalar coupling may only examine relatively short chains of interconnected atoms.[16]

Intraresidue correlation of protons is often accomplished via correlated spectroscopy (COSY) or total correlation spectroscopy (TOCSY) experiments. Both of these experiments utilize J-coupling for magnetization transfer. In a COSY experiment, signals between neighboring protons (usually up to four bonds away) are evolved. Whereas in a TOCSY experiment, all the correlations of all protons on any given side chain are evolved.[16] These experiments are often useful for determining the identity of residue side-chains.

The nuclear Overhauser effect (nOe) is a short range through-space interaction between atoms resulting from dipolar interactions. The strength of this effect is inversely proportional to the spatial distance between nuclei. The utilization of nuclear Overhauser effect spectroscopy (NOESY) experiments allows for the detection of interactions between the side-chain atoms of the amino acids in a protein. ^{1}H-^{1}H NOESY experiments are often used for the inter-residue correlation of protons on the side chains of residues.[16] This gives clues to the orientation of and distance between the side-chains of each residue.

All of the experiments described generate spectra that are highly sensitive to any structural perturbations. Therefore, the effects of glycan attachment to a glycosylation site or other post-translational modifications can be directly observed via NMR spectroscopy. The downside to studying glycosylation via solution NMR is the stringent sample homogeneity requirement. As stated previously, glycosylation is one of the most diverse forms of post-translational modification. Glycan presence and composition can change from protein to protein when expressed *in vivo*. However, using *in vitro* glycosylation, it is possible to maintain the rigid homogeneity requirements for these studies.

4.1.2 Literature Review

Syndecan-1 is considered to be intrinsically disordered based on observations from the amino acid content. However, there have been no direct studies of the structure via solution NMR. The only direct structural data available on SDC-1 comes from two different binding studies using extremely truncated peptides containing only 8 out of the total 330 amino acids in the protein.[110, 111] The first structural analysis performed by Liu et al. on SDC-1 examined the complex formed between an 8 amino acid peptide from the SDC-1 cytoplasmic domain (^{303}TKQEEFYA^{310}) and the Tiam1 PDZ domain. The structure of this complex was elucidated via X-ray crystallography (XRC).[110, 111] The only other structural data on a part of SDC-1 is from a study by Briggs et al. on a short 8 amino acid sequence from the ectodomain (^{202}AAEGSGEQ^{210}). In this study, the SDC-1 peptide structure was elucidated via XRC while in complex with Xylosyltransferase 1. Here, researchers were looking into the method for initiation of GAG biosynthesis and focused on a portion of the SDC-1 ectodomain that is able to accept an O-linked glycan.[110]

4.1.3 Summary and Outlook

The study of post-translational modification effects via NMR spectroscopy is nothing new. For example, phosphorylation (the enzymatic attachment of a PO_3 group to a Ser, Thr, or His) studies via NMR have been performed as far back as 1982.[112] While phosphorylation is common, with proteins often exhibiting the modification at multiple sites, it is also a small modification and therefore relatively easy to study via NMR spectroscopy. Acetylation, another highly-common PTM, is the attachment of an acetyl group to the positively charged N-terminus of a protein or to a lysine side chain. This PTM again results in a relatively small modification. Homologous sample conditions are easy to achieve, making changes in structural conformation simple to detect using NMR spectroscopy.[113]

Glycosylation, on the other hand, is just as common as phosphorylation and acetylation, yet substantially more complex. Acetylation and phosphorylation both add small functional groups to a few specific locations on proteins, while glycosylation often attaches long, complex oligosaccharide polymers to proteins. While N-linked glycosylation has proven to be an important modification that often occurs with high stoichiometry at potential N-glycan sequons, the probability of glycan attachment has been shown in a cell and protein-dependent manner.[91] To counter these issues with sample heterogeneity, most glycoprotein research has been performed by isolating a glycoprotein, cleaving the glycan chains from the protein, and subsequently examining the glycans and core proteins separately.[107]

Prior to the work performed in this dissertation, no resolved NMR spectral data for SDC-1 was known to exist. In this chapter, I will discuss our methodology for the NMR characterization of ectodomains from the SDC-1 membrane protein. The structure of SDC-1 has long been considered disordered and the glycans associated with the protein are highly heterogeneous as well. Without having first produced a bacterial cell line to produce only the core SDC-1 ectodomain sequences (Chapter 2), this work with solution NMR would have been impossible. These research efforts have allowed us to produce the first ever resolved NMR spectra for the SDC-1 ectodomain. Our work in this chapter with NMR spectroscopy combined with the previous chapter's advances in *in vitro* glycosylation has laid the foundation for numerous future studies that will allow us to further probe the structure and dynamics of this highly-important membrane glycoprotein.

4.2 Materials and Methods

4.2.1 Expression of SDC-1 Ectodomains Using Uniform Isotopic Labeling

Uniform isotopic labeling was employed to generate recombinant protein with NMR active carbon and nitrogen nuclei for our two sSDC-1 constructs. This was accomplished by growing the *E. coli* in the presence of isotopically-labeled minimal media (M9 media) (7 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 7 g/L D-glucose, 1 g/L ammonium sulfate, 1 mM $CaCl_2$, 10 mM $MgSO_4$, 1.25 g/L Luria broth, 100 mg/L carbenicillin). For uniformly-labeled ^{15}N protein samples, ^{15}N ammonium sulfate was incorporated into the M9 media. For ^{13}C labeling, the 7 g/L D-glucose was replaced with 1 g/L ^{13}C D-glucose in the M9 media.

A 5 mL LB/carbenicillin starter culture was inoculated with 10 µL of recombinant cells containing the plasmid for our target sSDC-1 construct. After 2 hours of shaking at 37 °C, 200 µL of the starter culture was transferred into 100 mL of M9 media in a 200 mL baffle-bottom flask and grown with shaking at 37 °C. After 16 hours, 50 mL of the bacterial culture was transferred to 950 mL of fresh M9 media and grown with shaking for ~3 hours at 37 °C while monitoring the OD_{600}. Recombinant protein expression was induced via the addition of 1 mL of 120 mg/mL IPTG once the OD_{600} value reached 0.4. Bacteria were then grown an additional 4 hours before being harvested by centrifugation.

Isotopically labeled proteins were then purified as previously described in chapter 2.

4.2.2 Solution NMR Sample Preparation and Spectroscopy

To prepare a sample of our sSDC-1 ectodomains for NMR spectroscopy, the protein was first purified via Ni-NTA affinity chromatography. This could then be directly dialyzed against NMR buffer. For cleaved protein samples, the protein was first purified, then cleaved by TEV, and subsequently isolated from a second Ni-NTA affinity purification. The purified sSDC-1 proteins

in the Ni-NTA elutions were then dialyzed against 5 L of 20 mM sodium phosphate at pH 7 in 3.5 MWCO dialysis tubing. The dialysis buffer was renewed once after 30 minute and then again 2 hours later before being left to dialyze overnight at room temperature. The following day, the sSDC-1 solution was transferred from the dialysis bag to a 50 mL conical tube. Concentrations of the sSDC-1 protein were determined using their absorptions at 280 nm. The extinction coefficients were 17990 $M^{-1}cm^{-1}$ for both SDC-1_{1-82} and SDC-1_{1-245} before cleavage of the fusion proteins and 16,500 $M^{-1}cm^{-1}$ for both constructs after cleavage. These extinction coefficients correlated to an absorbance of 1.496 au and 0.627 au respectively for the uncleaved SDC-1_{1-82} and SDC-1_{1-245} constructs at a concentration of 1 g/L. For the cleaved constructs, a 1 g/L concentration correlated to absorbances of 1.908 and 0.652 respectively for SDC-1_{1-82} and SDC-1_{1-245}

The sSDC-1 constructs were then concentrated to a desired concentration for NMR analysis by centrifugal filtration. A 3000 Nominal Molecular Weight Limit (NMWL) centrifugal filter was used for SDC-1_{1-82} while a 10,000 NMWL centrifugal filter was used for SDC-1_{1-245}. For our NMR experiments, SDC-1_{1-82} was brought to a final concentration of 2 g/L while SDC-1_{1-245} was concentrated to 4.7 g/L. These samples were then made to be 10% v/v D_2O. The samples were then adjusted to a desired pH via the addition of concentrated HCl or NaOH and pH was tested using pH indicator strips.

NMR experiments were performed utilizing a Bruker DMX 500 MHz spectrometer (www.bruker-biospin.com) A triple resonance $^{1}H/^{13}C/^{15}N$ TXI high-resolution solution NMR probe was used. A pulse sequence of hsqcetgpsi2 (www.bruker-biospin.com) was calibrated and the water peak was referenced to 4.7 ppm. The ^{1}H-^{15}N HSQC experiments were run with 4096 t_2 points and 256 t_1 points. Spectral data was processed using nmrPipe[114] and visualized in Sparky.[115]

4.3 Results

In expressing isotopically-labeled proteins for NMR studies, we saw excellent expression levels of sSDC-1 even in minimal media. On average, we were able to isolate about 25 mg of either construct from 1 L of M9 growth media. With only 1-3 mg of protein needed for NMR studies, a single 1 L growth produces sufficient protein for several different studies. These labeled proteins were then used to generate highly resolved ^{1}H-^{15}N HSQC spectra for both SDC-$1_{1\text{-}82}$ and SDC-$1_{1\text{-}245}$.

In our initial preparations of sSDC-1 NMR samples, we lyophilized the protein and resolvated the samples in buffer. We tested the solubility of the dried proteins in two common NMR buffers, Sodium Phosphate (NaPi) and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES). Our sSDC-1 constructs were soluble in NaPi down to a pH of around 5.0 and in HEPES to a pH of 6.0. Below these pH values, the proteins began to precipitate. Interestingly, some of the protein redissolved upon addition of NaOH.

To gain some insight into why our proteins were insoluble at lower pH values, we used Expasy ProtScale to generate a Kyte-Doolittle hydropathy plot of the SDC-$1_{1\text{-}245}$ protein sequence with our fusion protein attached. **(Figure 4.1)** As noted in this hydropathy plot, a small portion of the fusion protein near the N-terminus (amino acids 38-44) may exhibit some slightly hydrophobic tendencies. This corresponds to a sequence in the ectodomain that is highly enriched in hydrophobic leucine residues. We hypothesized that this region of the two fusion proteins combined with their isoelectric points at pH 4.4 were contributing to poor solubility in low pH ranges. Because lower pH in NMR often means higher resolution, we hypothesized that the addition of detergents would stabilize the protein at low pH values and aid in NMR resolution.

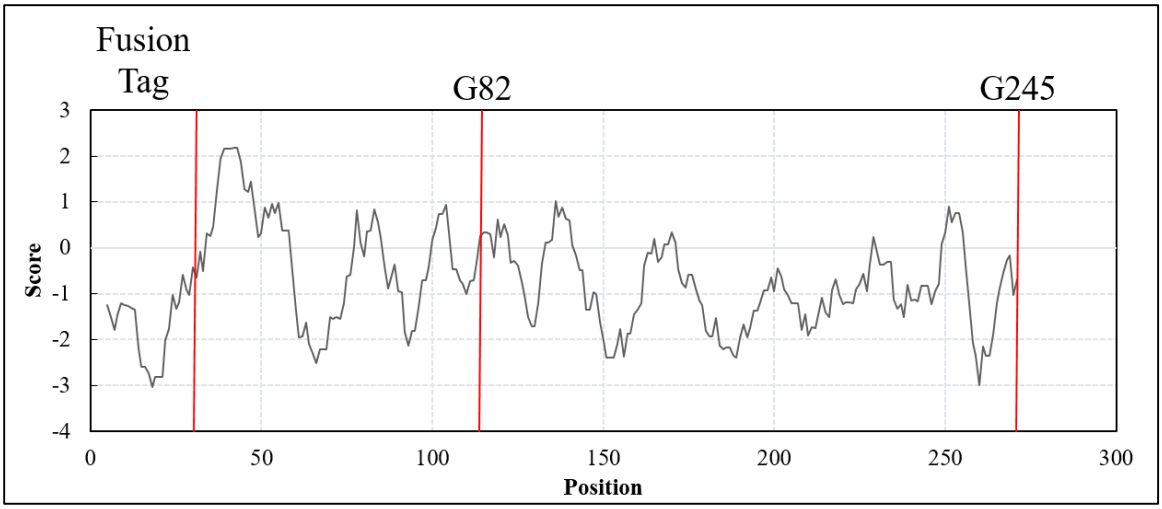

Figure 4.1: Kyte-Doolittle Hydropathy Plot. The hydropathy plot of a protein gives an indication of hydrophobicity/hydrophilicity of the protein sequence. The more positive the score, the more hydrophobic. Amino acid sequences that have a hydropathy score of greater than 1.8 are often found in transmembrane portions of the protein. The hydropathy plots for SDC-1 with our fusion protein attached is depicted here. A small portion of the fusion protein near the N-terminus (amino acids 38-44) may exhibit some slightly hydrophobic tendencies. In this region of the protein lies a sequence of LWLWLCALAL. The large number of semi-consecutive leucine residues in this region likely contribute greatly to the solubility issues experienced with these two proteins. After TEV cleavage a largely hydrophilic fusion protein is lost and solubility issues are further compounded.

To stabilize the protein at lower pH values, we tested the solubility of the protein in the presence of several NaPi-buffered detergents **(Figure 4.2)**.

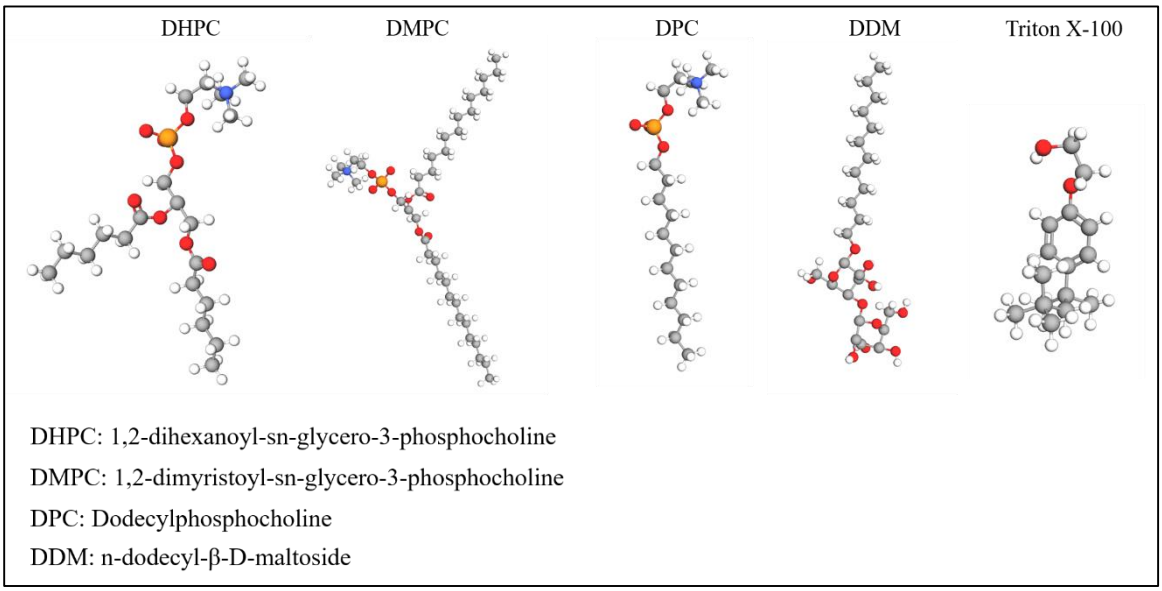

Figure 4.2: Detergents Tested to Aid Solubility. Detergents are often used to help solubilize proteins with hydrophobic regions. We tested the solubility of our sSDC-1 proteins in these 5 different detergents.

Prior to the addition of detergents, we were getting poorly resolved NMR spectra with many overlapping chemical shifts **(Figure 4.3)**. After addition of detergents the protein was able to remain soluble down to a pH of 3.0. We ran several ^{1}H-^{15}N HSQC experiments with the protein incorporated into our various detergents to determine which detergent provided the highest resolution. This allowed for the collection of our first resolved ^{1}H-^{15}N HSQC spectra for the SDC-$1_{1\text{-}82}$ ectodomain after selecting DHPC as the best detergent candidate **(Figure 4.4).**

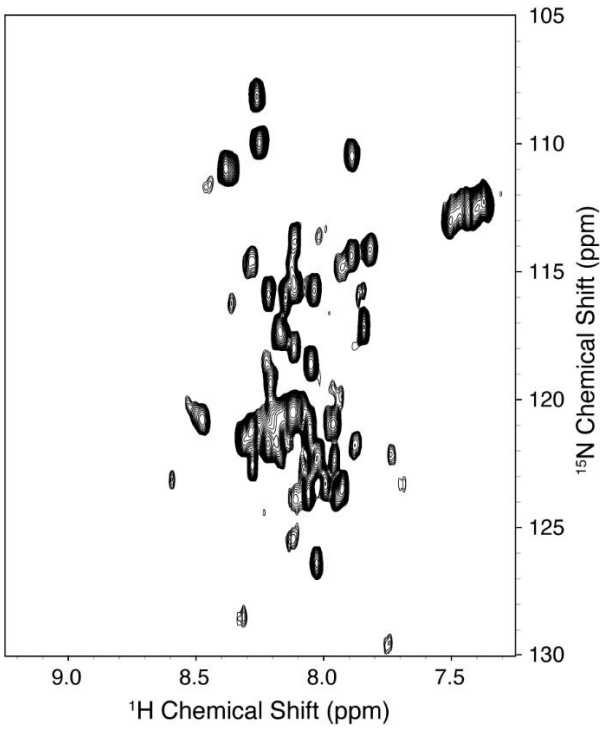

Figure 4.3: Low Resolution ^{1}H-^{15}N HSQC Spectrum of SDC-1$_{1\text{-}82}$. Prior to the addition of detergents, many of our proton-nitrogen chemical shifts were overlapping at pH 5.0

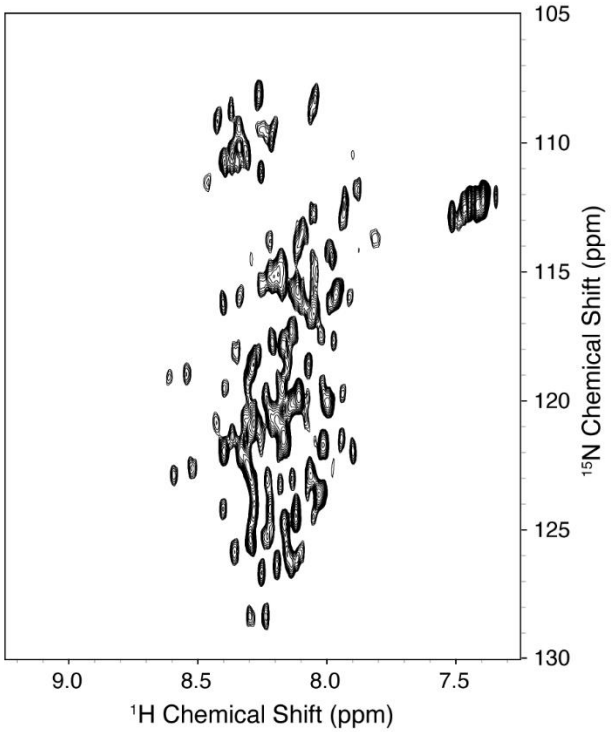

Figure 4.4: Higher-Resolution ^{1}H-^{15}N HSQC Spectrum of SDC-1$_{1\text{-}82}$ in DHPC. Resolution was markedly increased for the SDC-1$_{1\text{-}82}$ construct when incorporated into 125 mM DHPC micelles at pH 4.0

Even with the addition of detergents to our SDC-$1_{1\text{-}245}$ construct, the resolution left something more to be desired **(Figure 4.5 A)**. To further increase our resolution, we decreased the spectral width in the nitrogen dimension from 6,000 hz to 1,600 hz. We also increased the number of time points in the proton dimension from 2000 to 4000. We then ran the exact same sample again using these new parameters. The resulting SDC-$1_{1\text{-}245}$ spectra **(Figure 4.5 B)** showed excellent resolution of a majority of individual residues.

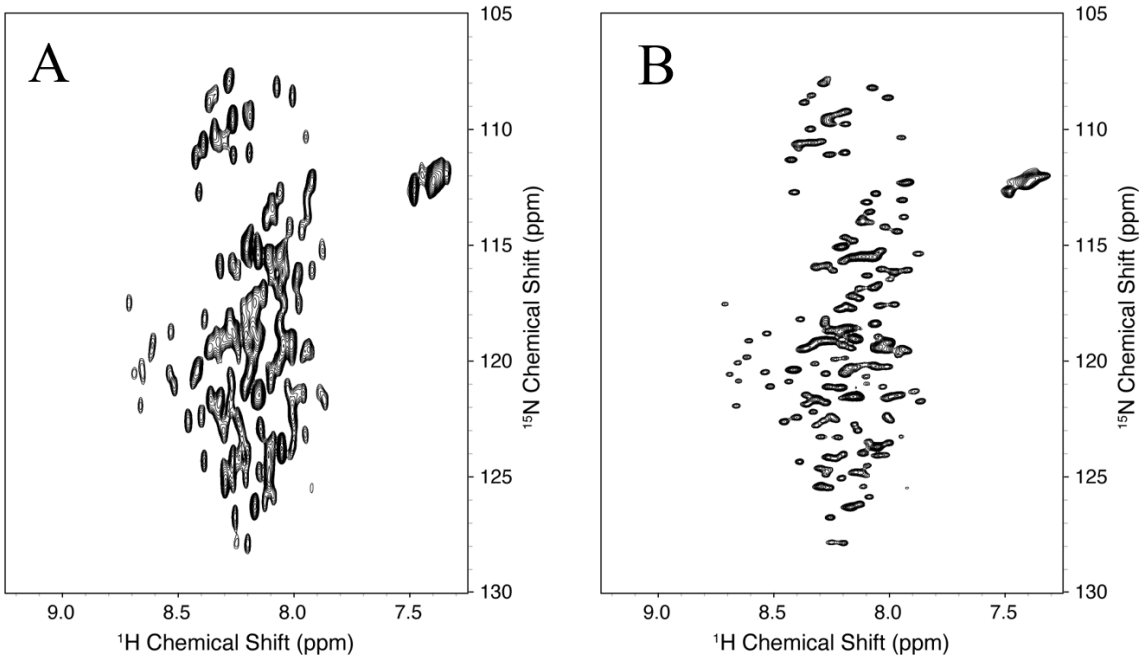

Figure 4.5: Resolution Increases due to Changes in Experimental Parameters.

A) SDC-$1_{1\text{-}245}$ was incorporated into 125 mM DHPC micelles at pH 4.0 and a ^{1}H-^{15}N HSQC spectra was acquired.

B) Spectral width in the ^{15}N dimension was decreased and the number of ^{1}H time points was doubled. The resulting resolution was significantly improved.

For both of our sSDC-1 constructs, we noticed early on that there were seemingly drastic changes in solubility after lyophilization of the proteins. The proteins exhibited excellent solubility in the Tris-HCl based resuspension buffer used for lysis. However, after drying the proteins, both constructs had difficulty dissolving in the same Tris-HCl buffer that we originally purified them in. We attempted to force the proteins into solution using sonication. After over five hours of sonication, we found that less than 1 mg of the 5-6 mg weighed out for the sample actually went into the 1 mL buffer solution.

We hypothesized that the lyophilization process was inducing aggregation of the proteins. Lyophilization is generally used to provide a long-term protein storage solution that limits degradation and aggregation.[116] However, studies have shown that lyophilization can actually have the opposite effect on some proteins, inducing aggregate formation.[117] These solubility issues forced to try a number of different lipid/buffer combinations in attempts to dissolve the dried protein. While we did get some resolved NMR spectra from these efforts as shown on the previous pages, we eventually made the decision to abandon these methods of lyophilization and resolvation of the proteins. Instead, we adopted a route that allowed the protein to remain soluble for the entirety of the purification and NMR sample preparation process. This completely eliminated solubility issues of the lyophilized protein and allowed the sample to remain in an aqueous NaPi buffer at a range of pH conditions without the use of lipids or detergents. We then began experiments to determine the best NMR buffer conditions that yielded spectra with excellent resolution and good peak dispersion.

After selecting a sodium phosphate buffer for NMR samples, we ran a series of experiments to determine the effect of pH on sample resolution and peak shift. In this series of experiments, we used a single sample of SDC-$1_{1\text{-}245}$ and ran multiple ^{1}H-^{15}N HSQC experiments on the aforementioned sample, adjusting the pH with HCl from 6 to 5 to 4**(Figure 4.6 A-D)**, then back up to 6 again with subsequent addition of NaOH **(Figure 4.7 A&B).** While we found that a pH of

4 yielded the best resolution, peak shifts in the spectra that were observed may be indicative of pH-dependent structural changes. Interestingly, some of these structural changes at lower pH values are reversible upon addition of NaOH to the sample.

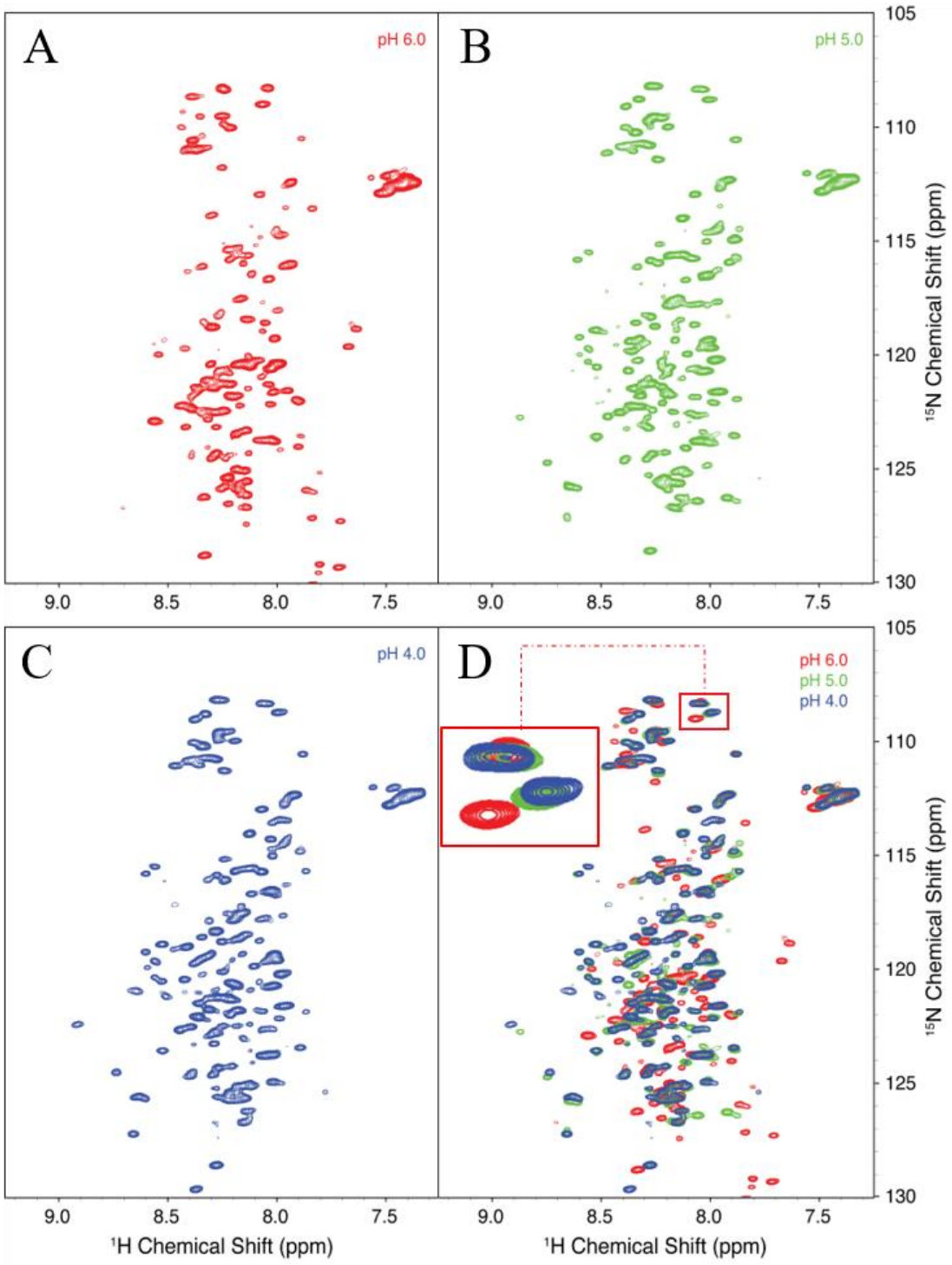

Figure 4.6: Adjustment of pH from 6 down to 4. An increasing number of residues and higher resolution can be seen as pH is adjusted from (A) pH 6, to (B) pH 5, to (C) pH 4. When overlaid (D), many of the chemical shifts from the residues in our protein overlap at different pH values. However, some of these residues exhibit a noticeable chemical shift change such as the residue highlighted in (D).

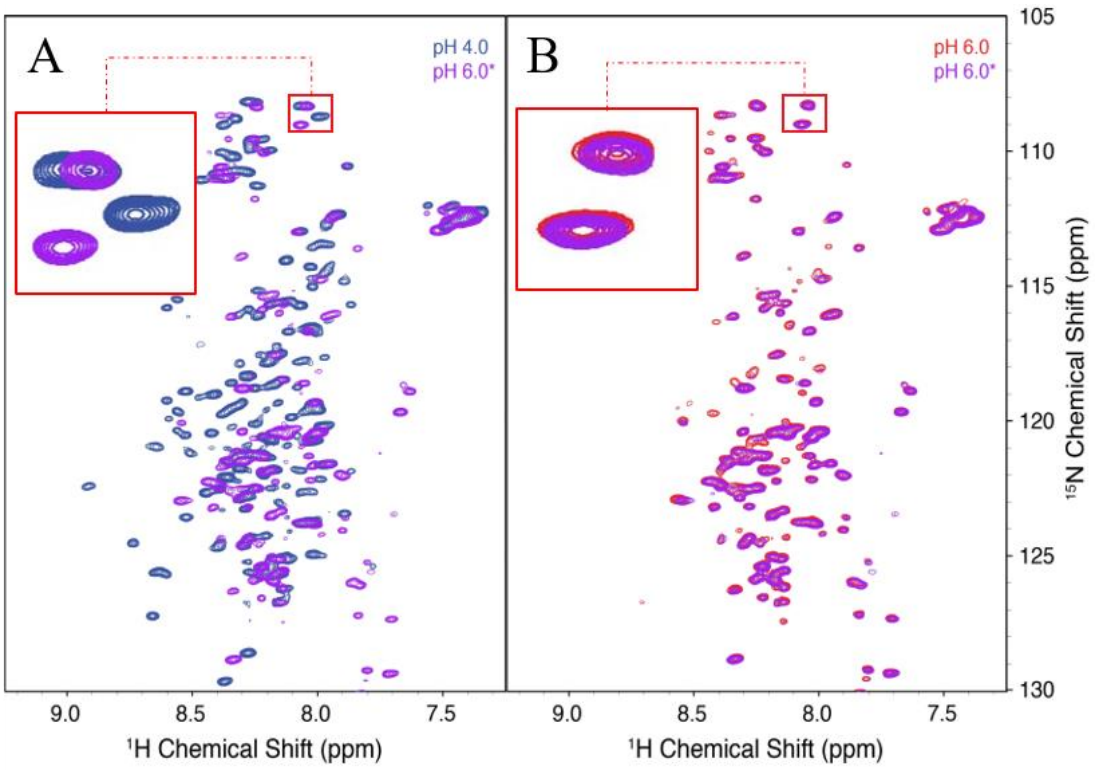

Figure 4.7: Adjustment of pH from 4 back to 6. In (A), when the pH is adjusted from 4 (blue) back up to 6 (purple) by addition of NaOH, a significant number of signals are lost. In (B) we can see an overlay of the pH 6 spectra at the start of the experiment (red) with the pH 6 spectra at the end of the experiment (purple). Notice that many of the residues which exhibited different chemical shifts at the lower pH values returned to their original positions (highlighted). This is indicative of changes in structure that are reversible and adaptive to different environmental pH conditions.

Spectra that were recorded at pH 6 for SDC-$1_{1\text{-}82}$ **(Figure 4.8 A)** and SDC-$1_{1\text{-}245}$ **(Figure 4.8 B)** have great resolution and peak dispersion indicating the protein samples are homogeneous and well structured. An overlay of the spectra **(Figure 4.8 C)** shows that a number of the peaks line up quite well and indicate that these residues are most likely in the shared N-terminal portion of the protein.

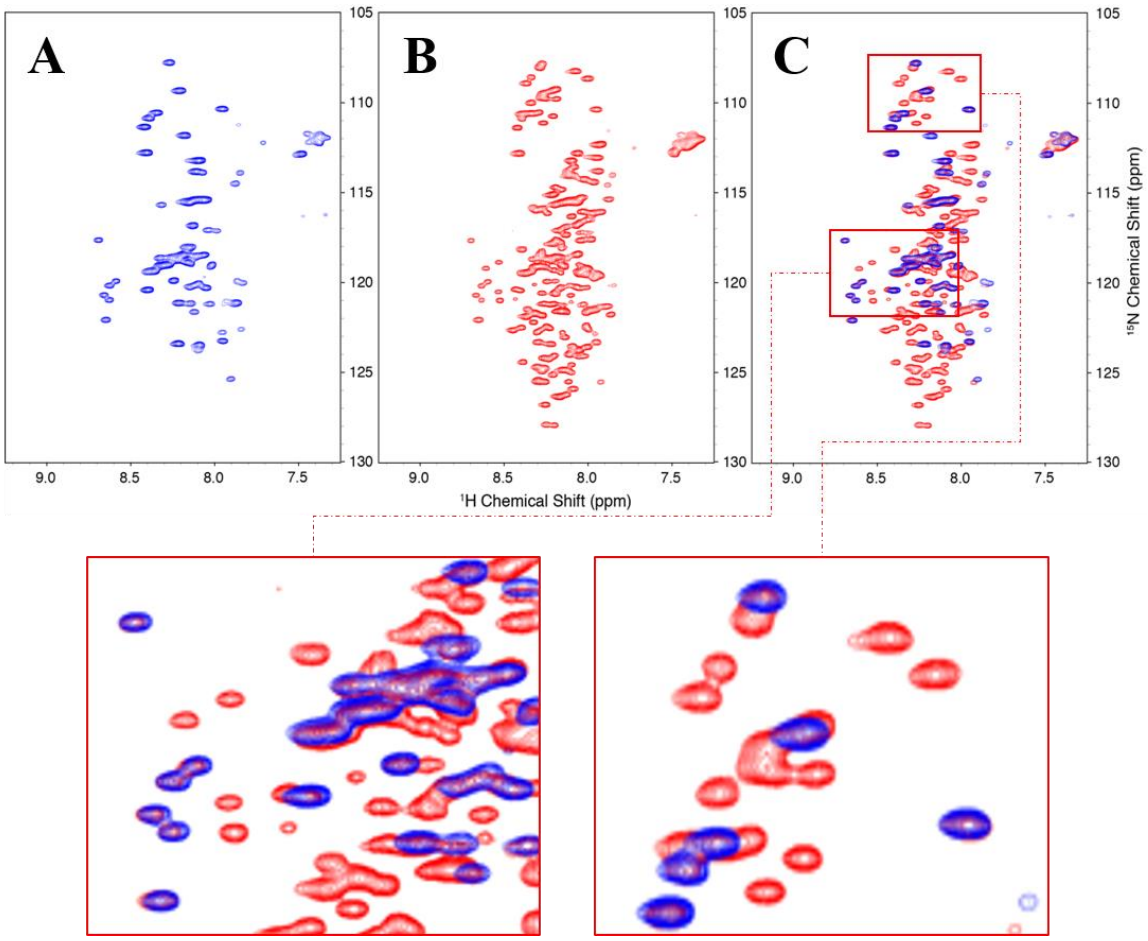

Figure 4.8: ^{1}H-^{15}N HSQC Spectra of SDC-1 constructs. Shown here are the acquired ^{1}H-^{15}N HSQC spectra for the two syndecan constructs. The SDC-$1_{1\text{-}82}$ (A) and SDC-$1_{1\text{-}245}$ (B) spectra are reliably reproducible and show good resolution even though the proteins are considered to be intrinsically disordered. The two spectra can be seen overlaid (C) where a majority of the residues in the SDC-$1_{1\text{-}82}$ protein overlay well with their counterparts in the SDC-$1_{1\text{-}245}$ sequence.

4.4 Discussion

The solubility of the SDC-1 ectodomain allows relatively simplistic sample preparation for solution NMR under physiological pH conditions. Our initial HSQC results indicate that both sSDC-1 ectodomains exhibit changes in chemical shifts as the pH environment around the protein changes. This may be an indication that there is a pH dependence of protein structure. Therefore, the ability to record resolved spectra under physiological pH conditions is highly important.

Having successfully developed this method of production, purification, and preparation of the isotopically-labeled sSDC-1 core protein, we plan to perform more in-depth NMR characterization studies of the sSDC-1 ectodomain fragments including the effects of *in vitro* glycosylation on structural adaptations and dynamics.

Abbreviations: Nuclear Magnetic Resonance (NMR) Correlated Spectroscopy (COSY); Total Correlation Spectroscopy (TOCSY); Nuclear Overhauser Effect (nOe); Nuclear Overhauser Effect Spectroscopy (NOESY); Post-Translational Modification (PTM); X-ray Crystallography (XRC); 3000 Nominal Molecular Weight Limit (NMWL); Sodium Phosphate (NaPi); 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Introduction

In this final chapter, I will conclude my findings in the previous chapters and discuss the various future direction in which this research is heading.

5.2 Chapter 2: Expression and Purification of Syndecan-1 Ectodomains

In this chapter, I have described the step by step methods used to generate two different recombinant soluble ectodomains from the SDC-1 protein. We began with molecular cloning to generate our recombinant sSDC-1 DNA and ended by demonstrating isolation of these two distinct ectodomains in high-purity. Rather than expressing the full length membrane protein, we focused on expression of the soluble ectodomain. This ectodomain also happens to be the most biologically relevant part of SDC-1. The ectodomain has over 100 known binding partners, most of which it retains upon being shed and sequestered away into the extracellular matrix. We have demonstrated that we are able to isolate large amounts of this ectodomain with ease; thus providing a straightforward pathway for other scientists to express and purify this protein for their own studies.

Shed sSDC-1 has already been used as a biomarker for various diseases and cancers. However, no studies have been able to examine exactly which part of the SDC-1 ectodomain is being shed. We hope to soon perform antibody binding studies that will determine which current market antibodies, if any, have the ability to distinguish the two sSDC-1 ectodomains. If no antibodies are available, we plan to use our sSDC-1 DNA to have sets of antibodies generated that are able to distinguish the two different ectodomains in patient sera. We also hope to develop a highly-sensitive method for detection of sSDC-1 in other bodily fluids such as urine or saliva samples. It has already been established that high levels of SDC-1 in blood sera can be correlated to various disease states and cancers. Essentially, when large amounts of sSDC-1 are being shed in a person's body, then that means something is wrong. Development of a highly sensitive non-invasive method of detecting shed sSDC-1 could lead to a novel pathway for determining overall health and wellbeing of patients.

5.3 Chapter 3: *In Vitro* Glycosylation of SDC-1 Ectodomains

In chapter 3, I have reported our initial successes glycosylating the SDC-1 ectodomain utilizing *in vitro* methods. We demonstrated that we were able to attach a nucleotide-linked glucose to the single N-linked glycosylation site present on the SDC-1 ectodomain using NGT *in vitro*. While we have not yet quantified the glycosylation percentage, we will soon be able to. The efficiency of our glycosylation will be determined by quantifying the amount of unglycosylated vs. glycosylated peptides in our LC-MS/MS chromatograms.

In nature, it is likely rare to find eukaryotic glycoproteins with only a single glucose attached to an N-linked glycosylation site. Because of this, we are already looking into performing *in vitro* trans-glycosylation studies using the Endoglycosidase A enzyme. This will allow us to catalyze the transfer of a larger oxazoline sugar moiety in place of the single N-linked glucose that NGT

attaches. These studies using *in vitro* glycosylation in combination with solution NMR will prove vital for gaining a fundamental understanding of the effects of glycan attachment in biology.

5.4 Chapter 4: Structural Analysis of Ectodomains from SDC-1

In this chapter, I have discussed our current progress made in regards to understanding the structural biology of the SDC-1 ectodomain. In this study we determined how to prepare homogenous, soluble, isotopically-labelled sSDC-1 protein sequences for study via NMR spectroscopy. We determined that upon lyophilization, the sSDC-1 has a propensity for aggregation. These sSDC-1 aggregates are difficult to resolubilize unless under detergent conditions. This is likely due to the Leucine-rich domain near the N-terminus. After coming to this realization, we were able to devise a process that allows the sSDC-1 proteins to remain soluble all the way through NMR sample preparation. We then were able to record high-resolution spectra of both SDC-$1_{1\text{-}82}$ and SDC-$1_{1\text{-}245}$ in a water-soluble environment under physiological pH conditions.

In future studies, we hope to expand upon the structural elucidation of the SDC-1 ectodomain via NMR spectroscopy. By using a combination of heteronuclear HNCA, HN(CO)CA, HNCO, and HN(CA)CO experiments we will be able to assign resonances for each individual amino acid in the protein backbone. We can then utilize other experiments that take advantage of J-coupling and the nuclear Overhauser effect to help determine the 3-dimensional structure of the ectodomain. If indeed the structure is largely disordered as predicted, the protein will still make an excellent candidate for dynamics studies. Upon further structural elucidation, we plan to examine the structural responses as a response to glycosylation.

REFERENCES

1. Cournia, Z., et al., *Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory.* J Membr Biol, 2015. **248**(4): p. 611-40.

2. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* J Mol Biol, 2001. **305**(3): p. 567-80.

3. Lundstrom, K., *Latest development in drug discovery on G protein-coupled receptors.* Curr Protein Pept Sci, 2006. **7**(5): p. 465-70.

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

5. Deisenhofer, J., et al., *Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3A resolution.* Nature, 1985. **318**(6047): p. 618-24.

6. Dwek, R.A., *Biological importance of glycosylation.* Dev Biol Stand, 1998. **96**: p. 43-7.

7. Voet, D., *Biochemistry*. 2011. p. 399-406.

8. Meher, B.R., et al., *Glycosylation Effects on FSH-FSHR Interaction Dynamics: A Case Study of Different FSH Glycoforms by Molecular Dynamics Simulations.* PLoS One, 2015. **10**(9): p. e0137897.

9. Pena, C.E., Jr., M.G.S. Costa, and P.R. Batista, *Glycosylation effects on the structure and dynamics of a full-length Cel7A cellulase.* Biochim Biophys Acta Proteins Proteom, 2020. **1868**(1): p. 140248.

10. Aikawa, J., et al., *Effects of glycosylation on the secretion and enzyme activity of Mucor rennin, an aspartic proteinase of Mucor pusillus, produced by recombinant yeast.* J Biol Chem, 1990. **265**(23): p. 13955-9.

11. Goettig, P., *Effects of Glycosylation on the Enzymatic Activity and Mechanisms of Proteases.* Int J Mol Sci, 2016. **17**(12).

12. Blom, N., et al., *Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence.* Proteomics, 2004. **4**(6): p. 1633-49.

13. Kong, Y., et al., *N-Glycosyltransferase from Aggregatibacter aphrophilus synthesizes glycopeptides with relaxed nucleotide-activated sugar donor selectivity.* Carbohydr Res, 2018. **462**: p. 7-12.

14. Van den Steen, P., et al., *Concepts and principles of O-linked glycosylation.* Crit Rev Biochem Mol Biol, 1998. **33**(3): p. 151-208.

15. Langdon, R.H., J. Cuccui, and B.W. Wren, *N-linked glycosylation in bacteria: an unexpected application.* Future Microbiol, 2009. **4**(4): p. 401-12.

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

17. Petrescu, A.J., et al., *The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding.* EMBO J, 1997. **16**(14): p. 4302-10.

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

19. Patching, S.G., *NMR-Active Nuclei for Biological and Biomedical Applications.* Journal of Diagnostic Imaging in Therapy, 2016.

20. Grzesiek, S., et al., *1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma.* Biochemistry, 1992. **31**(35): p. 8180-90.

21. Kay, L.E., et al., *Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990.* J Magn Reson, 2011. **213**(2): p. 423-41.

22. Lagasse, H.A., et al., *Recent advances in (therapeutic protein) drug development.* F1000Res, 2017. **6**: p. 113.

23. Mattanovich, D., et al., *Recombinant protein production in yeasts.* Methods Mol Biol, 2012. **824**: p. 329-58.

24. Ikonomou, L., Y.J. Schneider, and S.N. Agathos, *Insect cell culture for industrial production of recombinant proteins.* Appl Microbiol Biotechnol, 2003. **62**(1): p. 1-20.

25. Ward, O.P., *Production of recombinant proteins by filamentous fungi.* Biotechnol Adv, 2012. **30**(5): p. 1119-39.

26. Wurm, F.M., *Production of recombinant protein therapeutics in cultivated mammalian cells.* Nat Biotechnol, 2004. **22**(11): p. 1393-8.

27. Plasson, C., et al., *Production of recombinant proteins in suspension-cultured plant cells.* Methods Mol Biol, 2009. **483**: p. 145-61.

28. Vieira Gomes, A.M., et al., *Comparison of Yeasts as Hosts for Recombinant Protein Production.* Microorganisms, 2018. **6**(2).

29. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Front Microbiol, 2014. **5**: p. 172.

30. Francis, D.M. and R. Page, *Strategies to optimize protein expression in E. coli.* Curr Protoc Protein Sci, 2010. **Chapter 5**: p. Unit 5 24 1-29.

31. Diefenderfer, C., et al., *Reliable expression and purification of highly insoluble transmembrane domains.* Anal Biochem, 2009. **384**(2): p. 274-8.

32. Dmitriev, P.V. and Y.S. Vassetzky, *A set of vectors for introduction of antibiotic resistance genes by in vitro Cre-mediated recombination.* BMC Res Notes, 2008. **1**: p. 135.

33. Gottesman, S., *Minimizing proteolysis in Escherichia coli: genetic solutions.* Methods Enzymol, 1990. **185**: p. 119-29.

34. Burgess-Brown, N.A., et al., *Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study.* Protein Expr Purif, 2008. **59**(1): p. 94-102.

35. Bessette, P.H., et al., *Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 13703-8.

36. Lefebvre, J., G. Boileau, and P. Manjunath, *Recombinant expression and affinity purification of a novel epididymal human sperm-binding protein, BSPH1*. Mol Hum Reprod, 2009. **15**(2): p. 105-14.

37. Graves, P.R. and T.A. Haystead, *Molecular biologist's guide to proteomics.* Microbiol Mol Biol Rev, 2002. **66**(1): p. 39-63; table of contents.

38. Melby, J.A., et al., *Novel Strategies to Address the Challenges in Top-Down Proteomics.* J Am Soc Mass Spectrom, 2021. **32**(6): p. 1278-1294.

39. Tabet, J.C. and S. Rebuffat, *[Nobel Prize 2002 for chemistry: mass spectrometry and nuclear magnetic resonance].* Med Sci (Paris), 2003. **19**(8-9): p. 865-72.

40. Dreisewerd, K., *The desorption process in MALDI.* Chem Rev, 2003. **103**(2): p. 395-426.

41. Ho, C.S., et al., *Electrospray ionisation mass spectrometry: principles and clinical applications.* Clin Biochem Rev, 2003. **24**(1): p. 3-12.

42. Meier-Credo, J., et al., *Top-Down Identification and Sequence Analysis of Small Membrane Proteins Using MALDI-MS/MS.* J Am Soc Mass Spectrom, 2022. **33**(7): p. 1293-1302.

43. Cui, W., H.W. Rohrs, and M.L. Gross, *Top-down mass spectrometry: recent developments, applications and perspectives.* Analyst, 2011. **136**(19): p. 3854-64.

44. Zhang, Y., et al., *Protein analysis by shotgun/bottom-up proteomics.* Chem Rev, 2013. **113**(4): p. 2343-94.

45. Rajawat, J. and G. Jhingan, *Chapter 1 - Mass spectroscopy*, in *Data Processing Handbook for Complex Biological Data Sources*, G. Misra, Editor. 2019, Academic Press. p. 1-20.

46. Pejchinovski, M., et al., *Comparison of higher energy collisional dissociation and collision-induced dissociation MS/MS sequencing methods for identification of naturally occurring peptides in human urine.* Proteomics Clin Appl, 2015. **9**(5-6): p. 531-42.

47. Hu, A., W.S. Noble, and A. Wolf-Yadlin, *Technical advances in proteomics: new developments in data-independent acquisition.* F1000Res, 2016. **5**.

48. Guan, S., et al., *Data Dependent-Independent Acquisition (DDIA) Proteomics.* J Proteome Res, 2020. **19**(8): p. 3230-3237.

49. Simon Davis, D.A. and C.R. Parish, *Heparan sulfate: a ubiquitous glycosaminoglycan with multiple roles in immunity.* Front Immunol, 2013. **4**: p. 470.

50. Sarrazin, S., W.C. Lamanna, and J.D. Esko, *Heparan sulfate proteoglycans.* Cold Spring Harb Perspect Biol, 2011. **3**(7).

51. Hook, M., L. Kjellen, and S. Johansson, *Cell-surface glycosaminoglycans.* Annu Rev Biochem, 1984. **53**: p. 847-69.

52. Carey, D.J., *Syndecans: multifunctional cell-surface co-receptors.* Biochem J, 1997. **327 (Pt 1)**: p. 1-16.

53. Gondelaud, F. and S. Ricard-Blum, *Structures and interactions of syndecans.* FEBS J, 2019. **286**(15): p. 2994-3007.

54. Peysselon, F., et al., *Intrinsic disorder of the extracellular matrix.* Mol Biosyst, 2011. **7**(12): p. 3353-65.

55. Leonova, E.I. and O.V. Galzitskaya, *Cell communication using intrinsically disordered proteins: what can syndecans say?* J Biomol Struct Dyn, 2015. **33**(5): p. 1037-50.

56. Saunders, S., et al., *Molecular cloning of syndecan, an integral membrane proteoglycan.* J Cell Biol, 1989. **108**(4): p. 1547-56.

57. Angsana, J., et al., *Syndecan-1 modulates the motility and resolution responses of macrophages.* Arterioscler Thromb Vasc Biol, 2015. **35**(2): p. 332-40.

58. Voyvodic, P.L., et al., *Loss of syndecan-1 induces a pro-inflammatory phenotype in endothelial cells with a dysregulated response to atheroprotective flow.* J Biol Chem, 2014. **289**(14): p. 9547-59.

59. Xu, J., et al., *Endogenous attenuation of allergic lung inflammation by syndecan-1.* J Immunol, 2005. **174**(9): p. 5758-65.

60. Teng, Y.H., R.S. Aquino, and P.W. Park, *Molecular functions of syndecan-1 in disease.* Matrix Biol, 2012. **31**(1): p. 3-16.

61. Manon-Jensen, T., Y. Itoh, and J.R. Couchman, *Proteoglycans in health and disease: the multiple roles of syndecan shedding.* FEBS J, 2010. **277**(19): p. 3876-89.

62. Stepp, M.A., et al., *Syndecan-1 and Its Expanding List of Contacts.* Adv Wound Care (New Rochelle), 2015. **4**(4): p. 235-249.

63. Manon-Jensen, T., H.A. Multhaupt, and J.R. Couchman, *Mapping of matrix metalloproteinase cleavage sites on syndecan-1 and syndecan-4 ectodomains.* FEBS J, 2013. **280**(10): p. 2320-31.

64. Jalkanen, M., et al., *Cell surface proteoglycan of mouse mammary epithelial cells is shed by cleavage of its matrix-binding ectodomain from its membrane-associated domain.* J Cell Biol, 1987. **105**(6 Pt 2): p. 3087-96.

65. Aref, S., T. Goda, and M. El-Sherbiny, *Syndecan-1 in multiple myeloma: relationship to conventional prognostic factors.* Hematology, 2003. **8**(4): p. 221-8.

66. Seidel, C., et al., *Serum syndecan-1: a new independent prognostic marker in multiple myeloma.* Blood, 2000. **95**(2): p. 388-92.

67. Sheen-Chen, S.M., et al., *Serum levels of matrix metalloproteinase 2 in patients with breast cancer.* Cancer Lett, 2001. **173**(1): p. 79-82.

68. Guan, K.P., et al., *Serum levels of endostatin and matrix metalloproteinase-9 associated with high stage and grade primary transitional cell carcinoma of the bladder.* Urology, 2003. **61**(4): p. 719-23.

69. Xie, T., et al., *Association between MMP-2 expression and prostate cancer: A meta-analysis.* Biomed Rep, 2016. **4**(2): p. 241-245.

70. Li, H., et al., *The relationship between MMP-2 and MMP-9 expression levels with breast cancer incidence and prognosis.* Oncol Lett, 2017. **14**(5): p. 5865-5870.

71. Kanika Yadav, H.C., S. Chandra, G. Negi, S.K. Verma, *Evaluation of serum syndecan-1 level in patients of multiple myeloma.* Blood Heart and Circulation, 2018. **2**.

72. Szarvas, T., et al., *Circulating syndecan-1 is associated with chemotherapy-resistance in castration-resistant prostate cancer*. Urol Oncol, 2018. **36**(6): p. 312 e9-312 e15.

73. Su, G., et al., *Shedding of syndecan-1 by stromal fibroblasts stimulates human breast cancer cell proliferation via FGF2 activation.* J Biol Chem, 2007. **282**(20): p. 14906-15.

74. Sanaee, M.N., et al., *Soluble CD138/Syndecan-1 Increases in the Sera of Patients with Moderately Differentiated Bladder Cancer.* Urol Int, 2015. **94**(4): p. 472-8.

75. Jamaleddine, M., et al., *Expression, purification, and structural analysis of the full-length human integral membrane protein gamma-sarcoglycan.* Protein Expr Purif, 2020. **167**: p. 105525.

76. Tropea, J.E., S. Cherry, and D.S. Waugh, *Expression and purification of soluble His(6)-tagged TEV protease.* Methods Mol Biol, 2009. **498**: p. 297-307.

77. Raran-Kurussi, S., et al., *Removal of Affinity Tags with TEV Protease.* Methods Mol Biol, 2017. **1586**: p. 221-230.

78. Marty, M.T., et al., *Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles.* Anal Chem, 2015. **87**(8): p. 4370-6.

79. Weitzhandler, M., et al., *The cell surface proteoglycan of mouse mammary epithelial cells. The extracellular domain contains N terminus and a peptide sequence present in a conditioned medium proteoglycan.* J Biol Chem, 1988. **263**(15): p. 6949-52.

80. Kojima, T., N.W. Shworak, and R.D. Rosenberg, *Molecular cloning and expression of two distinct cDNA-encoding heparan sulfate proteoglycan core proteins from a rat endothelial cell line*. J Biol Chem, 1992. **267**(7): p. 4870-7.

81. Dews, I.C. and K.R. Mackenzie, *Transmembrane domains of the syndecan family of growth factor coreceptors display a hierarchy of homotypic and heterotypic interactions.* Proc Natl Acad Sci U S A, 2007. **104**(52): p. 20782-7.

82. Frottin, F., et al., *The proteomics of N-terminal methionine cleavage.* Mol Cell Proteomics, 2006. **5**(12): p. 2336-49.

83. Bonnin, P., et al., *Novel mRNA-specific effects of ribosome drop-off on translation rate and polysome profile.* PLoS Comput Biol, 2017. **13**(5): p. e1005555.

84. Mann, M. and O.N. Jensen, *Proteomic analysis of post-translational modifications.* Nat Biotechnol, 2003. **21**(3): p. 255-61.

85. Ramazi, S. and J. Zahiri, *Posttranslational modifications in proteins: resources, tools and prediction methods.* Database (Oxford), 2021. **2021**.

86. Pejaver, V., et al., *The structural and functional signatures of proteins that undergo multiple events of post-translational modification.* Protein Sci, 2014. **23**(8): p. 1077-93.

87. Uversky, V.N., *Post-Translational Modification*, in *Brenner's Encyclopedia of Genetics*. 2013, Elsevier. p. 425-428.

88. Xu, H., et al., *PTMD: A Database of Human Disease-associated Post-translational Modifications.* Genomics Proteomics Bioinformatics, 2018. **16**(4): p. 244-251.

89. Ramazi, S., A. Allahverdi, and J. Zahiri, *Evaluation of post-translational modifications in histone proteins: A review on histone modification defects in developmental and neurological disorders.* J Biosci, 2020. **45**.

90. Bedford, L., et al., *Is malfunction of the ubiquitin proteasome system the primary cause of alpha-synucleinopathies and other chronic human neurodegenerative disease?* Biochim Biophys Acta, 2008. **1782**(12): p. 683-90.

91. Zielinska, D.F., et al., *Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints.* Cell, 2010. **141**(5): p. 897-907.

92. Xiong, Y., et al., *Effects of N-Glycosylation on the Structure, Function, and Stability of a Plant-Made Fc-Fusion Anthrax Decoy Protein.* Front Plant Sci, 2019. **10**: p. 768.

93. Reily, C., et al., *Glycosylation in health and disease.* Nat Rev Nephrol, 2019. **15**(6): p. 346-366.

94. Pitti, T., et al., *N-GlyDE: a two-stage N-linked glycosylation site prediction incorporating gapped dipeptides and pattern-based encoding.* Sci Rep, 2019. **9**(1): p. 15975.

95. Vance, B.A., et al., *Multiple dimeric forms of human CD69 result from differential addition of N-glycans to typical (Asn-X-Ser/Thr) and atypical (Asn-X-cys) glycosylation motifs.* J Biol Chem, 1997. **272**(37): p. 23117-22.

96. Gavel, Y. and G. von Heijne, *Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering.* Protein Eng, 1990. **3**(5): p. 433-42.

97. Bradford, B.M., et al., *Dramatic reduction of PrP C level and glycosylation in peripheral nerves following PrP knock-out from Schwann cells does not prevent transmissible spongiform encephalopathy neuroinvasion.* J Neurosci, 2009. **29**(49): p. 15445-54.

98. Schwarz, F. and M. Aebi, *Mechanisms and principles of N-linked protein glycosylation.* Curr Opin Struct Biol, 2011. **21**(5): p. 576-82.

99. Schwarz, F., et al., *Cytoplasmic N-glycosyltransferase of Actinobacillus pleuropneumoniae is an inverting enzyme and recognizes the NX(S/T) consensus sequence.* J Biol Chem, 2011. **286**(40): p. 35267-74.

100. Rudd, P.M., et al., *Glycomics and Glycoproteomics*, in *Essentials of Glycobiology*, th, et al., Editors. 2022: Cold Spring Harbor (NY). p. 689-704.

101. Andrews, S.S. and J. Tretton, *Physical Principles of Circular Dichroism.* Journal of Chemical Education, 2020. **97**(12): p. 4370-4376.

102. Lorenz-Fonfria, V.A., *Infrared Difference Spectroscopy of Proteins: From Bands to Bonds.* Chemical Reviews, 2020. **120**(7): p. 3466-3576.

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

104. Kronquist, K.E. and W.J. Lennarz, *Enzymatic conversion of proteins to glycoproteins by lipid-linked saccharides: a study of potential exogenous acceptor proteins.* J Supramol Struct, 1978. **8**(1): p. 51-65.

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

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

107. Unione, L., et al., *NMR of glycoproteins: profiling, structure, conformation and interactions.* Curr Opin Struct Biol, 2021. **68**: p. 9-17.

108. Bern, M., Y.J. Kil, and C. Becker, *Byonic: advanced peptide and protein identification software.* Curr Protoc Bioinformatics, 2012. **Chapter 13**: p. Unit13 20.

109. Goodman, J.K., et al., *Updates of the In-Gel Digestion Method for Protein Analysis by Mass Spectrometry.* Proteomics, 2018. **18**(23): p. e1800236.

110. Briggs, D.C. and E. Hohenester, *Structural Basis for the Initiation of Glycosaminoglycan Biosynthesis by Human Xylosyltransferase 1.* Structure, 2018. **26**(6): p. 801-809 e3.

111. Liu, X., et al., *The structure of the Tiam1 PDZ domain/ phospho-syndecan1 complex reveals a ligand conformation that modulates protein dynamics.* Structure, 2013. **21**(3): p. 342-54.

112. Vogel, H.J., W.A. Bridger, and B.D. Sykes, *Frequency-dependent phosphorus-31 nuclear magnetic resonance studies of the phosphohistidine residue to succinyl-CoA synthetase and the phosphoserine residue of glycogen phosphorylase a.* Biochemistry, 1982. **21**(6): p. 1126-32.

113. Kumar, A., V. Narayanan, and A. Sekhar, *Characterizing Post-Translational Modifications and Their Effects on Protein Conformation Using NMR Spectroscopy.* Biochemistry, 2020. **59**(1): p. 57-73.

114. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.

115. Lee, W., M. Tonelli, and J.L. Markley, *NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy.* Bioinformatics, 2015. **31**(8): p. 1325-7.

116. Moorthy, B.S., et al., *Predicting protein aggregation during storage in lyophilized solids using solid state amide hydrogen/deuterium exchange with mass spectrometric analysis (ssHDX-MS).* Mol Pharm, 2014. **11**(6): p. 1869-79.

117. Roughton, B.C., et al., *Protein aggregation and lyophilization: Protein structural descriptors as predictors of aggregation propensity.* Comput Chem Eng, 2013. **58**(2013): p. 369-377.

VITA

Austin Ray Anderson

Candidate for the Degree of

Doctor of Philosophy

Dissertation: EXPRESSION, CHARACTERIZATION, AND *IN VITRO* GLYCOSYLATION OF ECTODOMAINS FROM THE MEMBRANE PROTEIN SYNDECAN-1

Major Field: Chemistry

Biographical:

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

Completed the requirements for the Bachelor of Science in Chemistry at Northwestern Oklahoma State University, Alva, Oklahoma in 2017.

Experience:

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

- Research Fields – Protein Biochemistry, Molecular Cloning, Expression and Purification of Recombinant Membrane Proteins, Solution-State NMR Spectroscopy, MALDI-TOF Mass Spectrometry, Electrospray Ionization Mass Spectrometry, Tandem Mass Spectrometry
- Teaching – General chemistry 1&2, Chemistry for Engineers, Organic Chemistry 1&2 Private Tutoring, Physics 1&2 Private Tutoring

Professional Memberships:

American Chemical Society (2014-present)