A Novel Technique for Characterizing Polysialylated Proteins from Complex Mixtures

By

Tahlia Renee Lyne Derksen

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Physiology, Cell, and Developmental Biology

Department of Biological Sciences

University of Alberta

Abstract

Polysialic acid (polySia) is a long homopolymer of sialic acid residues that has profound consequences for the proteins it is attached to. Its expression is restricted in healthy human adults to the nervous, reproductive, and immune systems. It contributes to neuronal development and plasticity along with cell migration and immune suppression. PolySia is dysregulated in a variety of neurological diseases, mental health disorders, autoimmune diseases, and cancers. Abnormally high expression is associated with advanced disease and poor prognosis. However, many of the mechanisms of polySia in both health and disease remain poorly understood. This is partially due to not knowing which proteins are polysialylated and how they contribute to disease progression. As our analyses indicate that there are more polysialylated proteins than the handful documented in known and unknown sources, there is a need for further investigation.

We have developed a novel technique to identify polysialylated proteins from complex mixtures such as serum or cell lysate. In this method we have biotinylated a polySia lectin to immobilize on streptavidin agarose. The agarose beads can be added to complex mixtures to isolate polysialylated proteins. These isolated proteins undergo protein identification using mass spectrometry. This method improves upon immunoprecipitation techniques by allowing for vigorous washes with detergent.

The use of this method was validated by correctly identifying neural cell adhesion molecule (NCAM/CD56) as a polysialylated protein in the non-Hodgkin's lymphoma cell line NK-92. The utility has been further demonstrated by identifying novel polysialylated proteins from primary human T cells and invasive breast ductal carcinoma cell line MCF-7.

Along with the development of this technique, I have characterized polySia in the rare but fatal rheumatic disease scleroderma. This is the first documentation of polySia dysregulation in scleroderma. We have identified polySia both in scleroderma fibroblasts and serum and my novel technique will lead to the identification of the polysialylated proteins in scleroderma.

In addition to improving our understanding of the role polySia plays in health and disease, these identified proteins have the potential to be used for diagnostic and prognostic testing. This proteomics method is versatile and will be useful for identifying polysialylated proteins from various sources such as immune cells, cancers, and other diseases.

Preface

This research was performed under the supervision of Dr. Lisa Willis (Faculty of Biological Sciences) at the University of Alberta. Scleroderma research was conducted in collaboration with Dr. Mohamed Osman (Faculty of Medicine and Dentistry) at the University of Alberta. Proteomics experiments were performed in collaboration with Dr. Nichollas Scott (Department of Microbiology and Immunology) at the University of Melbourne.

A version of chapter 4 is awaiting submission as Makhsous, S.[*], Derksen, T.[*], Doll, M., Rodriguez, S., Scott, N.E., Willis, L.M. (2023). "Site-specific immobilization of the endosialidase improves its use as a tool to study polysialic acid." I was responsible for $EndoN_{DM}$ lectin data collection and assisted in writing the initial version of the manuscript. S.R. subcloned the lectin. Experiments were also performed by M.D., S.R., and N.E.S. L.M.W. was the supervisory author involved in concept formation and manuscript composition. All authors provided feedback on the manuscript.

Figure 5.1 A is published on biorxiv as Hunter, C., Derksen, T., Karathra J., Baker, K., Nitz, M., Willis, L.M. (2022). "A new strategy for identifying polysialylated proteins reveals they are secreted from cancer cells as soluble proteins and as part of extracellular vesicles." **doi:** https://doi.org/10.1101/2022.09.01.50623. I was responsible for assisting in method optimization. L.M.W. was the supervisory author involved in concept formation and manuscript composition. All authors provided feedback on the manuscript.

Data from chapter 3 is expected to be published in medical journal in the near future. All serum samples were obtained by Dr. Mohamed Osman. All fibroblasts were cultured by Lamia Khan. Some of the ELISA data was performed by Ana D'Aubeterre.

Data from chapters 5 is expected to be published in the future. I am responsible for all sample preparation and Dr. Nichollas Scott performed all mass spectrometry and analysis. The methods in chapter 2.6 were all performed and authored by Dr. Nichollas Scott. Dr. Carmanah Hunter is responsible for the initial research idea and further validation and mechanistic research on T cells in chapter 5.4.

[*] Authors contributed equally to the work

Acknowledgements

Firstly, I would like to extend my deepest gratitude to my supervisor Dr. Lisa Willis. She has been one of my greatest supporters and encouragers throughout my degree. Thank you for choosing to work with me. I have grown considerably during my time under your supervision, not only academically and professionally, but also personally. You have taught me so much more than just glycobiology and I look forward to using the skills you have helped me develop in my future life and career. I would also like to thank Dr. Glen Uhrig, my committee member, for his time and advice on my project.

Thank you to all the fantastic past and present members of the Willis lab: Dr. Carmanah Hunter, Sogand Makhsous, Muhammad Idrees, Bill Vouronikos and the many wonderful undergraduate students over the years. Thank you all for your friendship, encouragement, and support. It was a pleasure to work with you all. A special thank you to Sam Rodriguez and Ana D'Aubeterre for your prior contributions that led to the completion of my project. And to Ellen Hayhoe who will continue to carry the torch for this project after me.

Additionally, I would like to thank my family for getting me through this degree from afar. I am especially thankful for my partner Garrett Garlock for all your love, support, and encouragement through the difficult times and for celebrating with me through my accomplishments. I will be forever grateful to have you by my side. Thank you to Sara and Spencer Hilgendorf for your endless support, encouragement, and stability in my life.

Lastly, I would like to thank our cat, Taffeta, for keeping me company and bringing endless entertainment to my life.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Overview

Glycans cover the surface of all cells and play important roles in cell migration, receptor recognition and cell-to-cell interactions. Terminal sialic acid (Sia) sugars are often critical for this molecular recognition. Of particular interest is the glycan polysialic acid (polySia) because it plays many vital roles in both health and disease. PolySia has limited expression in healthy adults found on cell surface proteins as well as secreted glycoproteins. It is particularly important for synaptogenesis, immune system function and disease progression. For example, high amounts of polySia expressed in cancer tumors have been associated with increased metastasis and poor prognosis. However, the mechanisms of polySia function remain largely unknown, including many of the proteins to which polySia is attached. It is necessary to identify polysialylated proteins to better understand the functions of polySia in health and disease and to develop better diagnostics and treatments. We have developed a novel technique that allows us to identify polysialylated proteins from complex mixtures. We also demonstrated for the first time polySia is dysregulated in scleroderma and are working towards understanding how it contributes to the pathogenesis of this deadly disease.

1.2 Sialic acid in vertebrates

Sialic acids (Sia) are found in all vertebrates on cell surface proteins and lipids and most secreted proteins. Sias are a group of 9-carbon carboxylated monosaccharides found in many animal tissues typically on the non-reducing terminal end of glycans. The two major forms in most animals are *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc), which differ only by an additional alcohol in Neu5Gc. Humans, however, are only able to produce Neu5Ac due to a genetic mutation that produces an inactive truncated peptide in place of the enzyme responsible for making Neu5Gc[1]. Sia residues are attached to other sugars through α-2,3- or α-2,6-linkages and can be attached to other Sias through α-2,8 linkages.

Sia plays many roles in a variety of cellular processes, including signalling, migration, and intracellular interaction. Sia found on animal cell surfaces serves as a ligand for intrinsic and extrinsic lectins, antibodies, enzymes, and receptors[2,3]. For example, human CD22, a B-cell restricted receptor, has specificity for α-2,6-linked Neu5Ac while murine CD22 prefers Neu5Gc. CD22 itself is also sialylated with α-2,6-linked Sia to assist in aggregation of the receptor on the surface of B cells[4]. Red blood cell glycans are capped with Sia to stabilize them and prevent aggregation[2]. Sia has also been demonstrated to increase the half-life of secreted sialylated proteins in blood and milk by preventing removal of the proteins by the hepatic lectin, the Ashwell-Morell receptor [5]. Sia is vital to embryonic development as Sia enzyme knockout mice are embryonically lethal. Normal expression of Sia is very important for maintaining health due to its involvement in many biological processes.

Aberrant sialylation contributes to the pathology of numerous diseases. Some immune disorders are caused by autoantibody production against Sia or improperly glycosylated antibodies[2,3]. Changes in sialylation occurs in many cancers as well and are associated with tumor growth[3]. Malignant cells tend to overexpress Sia which may protect them from immune recognition and destruction. In particular, cancer-derived gangliosides have been reported to inhibit activation and migration of Langerhans cells, tissue-residing dendritic cells, and induce apoptosis within these immune cells[6]. Similarly, some pathogens use molecular mimicry to cover their surfaces in Sia preventing immune system recognition. Overall, Sia expression is major participant in healthy development and disease pathogenesis.

1.3 Polysialic acid structure and biosynthesis

Polysialic acid (polySia) is a long chain of α-2,8-linked sialic acid residues, ranging from 8 to >400, that possesses properties which differ from monosialic acid units. The unique properties of polySia can have profound implications for the cells and proteins it is on. PolySia can form both repulsive and attractive fields. Along with negative charge, the large size of polySia contributes to its hydrophilicity and large hydration radius that can physically prevent cell-cell and protein-protein interactions[7,8]. The glycan also creates an attractive field for positively charged molecules and can bind to small proteins and molecules, such as growth

factors, neurotrophins, and cytokines influencing cell migration and differentiation[7,8]. The multidimensional properties of polySia open up the probability for many different interactions and functions of polySia.

PolySia has been identified in various organisms, but with greatly restricted expression. First identified in bacterial capsules, various forms of polySia have since been characterized in the nervous and reproductive systems of echinoderms, fish, birds, and mammals[7]. Additionally, polySia is present in the immune system of mammals, where it exists on the cell surface of both innate and adaptive immune cells[7]. While polySia can exist in chains of α-2,8- and/or α-2,9-linked Neu5Ac, Neu5Gc, or 2-keto-3-deoxynononic acid or deaminated neuraminic acid (Kdn), in mammals it is exclusively α-2,8-linked Neu5Ac or Neu5Gc and exclusive α-2,8-linked Neu5Ac in humans.

Additionally, there are only a handful of known polysialylated proteins in mammals (Table 1.3). The highly well-known and greatly referenced polysialylated protein in vertebrates is the neural cell adhesion molecule (CD56). CD56 is one of the most widely expressed proteins in the mammalian nervous system[7,9,10] and can exist unpolysialylated or heavily polysialylated, with both states influencing cell-cell adhesion[9]. PolySia-CD56 has anti-adhesion properties, which confers a migratory phenotype that enables cell migration, thereby influencing synaptic plasticity and axonal growth[7]. PolySia-CD56 also exists in the reproductive and nervous systems of mammals[11,12], though its roles outside the nervous system remain less understood. More recently synaptic cell adhesion molecule (SynCAM)[13] and neuropilin-2 (NRP-2)[14] have also been identified as polysialylated in mammals. Additionally, there are publications indicating that E-selectin ligand 1[15], C-C chemokine receptor 7 (CCR7)[16,17], CD36[18] and the voltage gated sodium channel[19,20] are polysialylated in mammals. However, these proteins have yet to be confirmed with other methods. Knowing what proteins are polysialylated would improve our understanding of polySia biology.

Polysialylated protein	Physiological location	Potential implications	Refs
Neural cell adhesion molecule (CD56)	Brain, sperm, and natural killer cells	Overexpressed in tumors, associated with advanced disease	[7,11,12]
Polysialyltransferase ST8Sia2	Golgi of brain and heart muscle (The Human Protein Atlas[21])	Overexpressed in cancers, single nucleotide polymorphisms or deletion associated with schizophrenia, autism spectrum disorder, and bipolar disorder	[7,8,22]
Polysialyltransferase ST8Sia4	Golgi of most tissues (The Human Protein Atlas[21])	Overexpressed in cancers	[7,8,22]
Synaptic cell adhesion molecule (SynCAM)	Brain	Unknown	[13]
Neuropilin-2 (NRP-2)	Dendritic cells and macrophages	Unknown	[14,23]
E-selectin ligand 1	Brain	Unknown	[15]
C-C chemokine receptor 7 (CCR7)	Dendritic cells	Unknown	[16,17]
CD36	Breast milk	Unknown	[18]
Voltage gated sodium channel	Brain	Unknown	[19,20]

Table 1.3: Physiological location and potential implications of mammalian polysialylated proteins found in the literature.

The restricted expression of polySia is in part due to the regulation of the biosynthetic enzymes responsible for its synthesis. Two polysialyltransferase enzymes are responsible for polySia production in mammals, ST8Sia2 and ST8Sia4. These enzymes first autopolysialylate before producing chains of polySia on both N-linked and O-linked glycans of target proteins[22]. Polysialyltransferases are substrate specific which results a very limited number of proteins

being polylsialylated[7]. Both enzymes are individually capable of synthesizing polySia, but expression varies throughout development and adulthood. ST8Sia2 and ST8Sia4 are abundantly present during development, but ST8Sia2 decreases substantially and ST8Sia4 slowly declines after birth [9]. In adulthood, expression of both enzymes is limited and ST8Sia4 is the primary enzyme responsible for polySia in the reproductive and immune systems of mammals[7,12,24].

1.4 Polysialic acid in the reproductive system

PolySia is found in multiple places in the reproductive systems, though little research has been done to elucidate the molecular details underpinning its function. In females, polysialoglycoprotein is a major polysialylated protein in unfertilized eggs from Salmonidae fish and sea urchins[7]. PolySia has also been detected on yet to be identified glycoproteins in the human placenta where it contributes to cytotrophoblast migration[25]. Additionally, polySia has been detected in human and mouse breast milk on the scavenger receptor CD36, though this data has not yet been corroborated in other labs[18]. The potential role for polySia in breast milk has not been identified, though breast milk contains many other immuno- and neuro-modulatory glycans[26]. In males, polySia occurs on sperm in sea urchin and in mammals, though on different proteins. In sea urchin sperm, it has been characterised on flagellasialin and polysialogangliosides where it plays a role in sperm motility and fertilization[27]. In mammalian sperm, CD56 and ST8Sia2 are the predominant polysialylated proteins and may play a role in sperm motility, cryoprotection, and modulating the female immune system[12]. While the functions of polySia in reproduction remain elusive, we do see the glycan conserved across organisms in both female and male reproductive systems.

1.5 Polysialic acid in the mammalian nervous system

Most of our current understanding of the function of polySia comes from studies examining the murine nervous system. PolySia-CD56 has been extensively studied in the murine brain and is present systemically throughout the developing brain but limited in adults to areas of plasticity, particularly the olfactory and visual systems along with forebrain subventricular zone and dentate gyrus of the hippocampus[9]. Along with CD56, ST8Sia2 and ST8Sia4, SynCAM has also been identified as polysialylated in the developing mouse brain[13]. The majority of the literature focuses on polySia-CD56 even though other polysialylated proteins have been identified.

PolySia expression during embryonic and fetal growth is essential for nervous system development. Both ST8Sia2 and ST8Sia4 are abundantly expressed during early neurogenesis. PolySia plays important roles in precursor cell migration, not only on neuronal precursors but also on mesodermal and endodermal tissue precursors[28]. Many of these cells lose polySia with differentiation and maturation after migration is complete[10]. In mice, the removal of one polysialyltransferase is compensated for by the remaining enzyme. However, double knockouts of both enzymes lead to less than 20% survival rate 4 weeks after birth[29]. The fatal phenotype in the absence of the polysialyltransferases is indicative of the critical roles polySia plays during development.

In neonates, overall polySia levels decrease and expression is restricted in mammalian adult brains. Immediately post-birth, polySia and polysialyltransferase expression drastically diminishes[9]. ST8Sia2 expression becomes nearly non-existent and ST8Sia4 expression is greatly reduced but remains responsible for most polysialylation in healthy adults and polySia-CD56 remains the most abundant polysialylated protein. While CD56 continues to be expressed widely in the CNS, polySia-CD56 expression is restricted. High polySia expression can still be found in adult brains in areas of plasticity and repair, particularly in the olfactory bulb system and the hippocampus. PolySia-CD56 is important for axonal growth and regeneration along with synaptic plasticity. PolySia is also expressed in the visual system of the peripheral nervous system[8,10]. In the spinal cord, polySia expression can be observed on sensory neurons. It is also crucial for the migration of oligodendrocyte precursors, but myelination only occurs on polySia

negative axons[9]. Even though polySia expression is significantly lower post-birth, it still plays a crucial part in the adult nervous system.

1.6 Polysialic acid in the mammalian immune system

Many myeloid derived innate immune cells have been shown to express polySia. Dendritic cells (DCs), derived from monocytes, express polySia but not CD56. Neuropilin-2 (NRP-2) is an abundantly expressed protein on DCs that was discovered to be polysialylated on DCs[14]. Removal of polySia reduces CCL21-driven chemotaxis of DCs indicating that polySia-NRP-2 is needed for the migration of DCs towards the chemokine[16,30]. However, removal of NRP-2 alone does not inhibit migration and it was suggested that C-C chemokine receptor 7 (CCR7) acts as another polySia carrier on DCs[17]. CCR7 has previously been characterized as receptor for CCL21. PolySia on DCs also helps regulate T cell proliferation and release of proinflammatory cytokines[24]. Monocytes can also differentiate into macrophages that initially lose polySia expression. However, polySia can be re-expressed on NRP-2 and potentially additional proteins in macrophages[23]. A potential reason for loss of polySia with macrophage maturation, is the enhanced phagocytic ability in the absence of polySia[23]. While murine monocytes do not produce polySia, bone marrow myeloid cells do. CD56 is thought to be the only polysialylated protein expressed in murine bone marrow however, there is a small subset of CD56 negative cells expressing polySia[23]. Neutrophils express polySia and it can protect endogenous cells from histone cytotoxicity released in neutrophils extracellular traps (NETs). Not only does polySia bind to histones, but also to lactoferrin released during NETosis[31]. Both neutrophils and macrophages lose external polySia when they enter the peripheral blood circulation[11]. The role polySia plays in the innate immune system could be imperative to understanding immunodeficiencies and autoimmune diseases.

PolySia has also been characterized in lymphoid derived immune cells. Human natural killer (NK) cells express polySia primarily on CD56, although polySia-ST8Sia4 is also present in human NK cells. Mice do not express polySia or CD56 on their NK cells[24] which complicates the understanding of polySia in the human immune system from murine models. PolySia-CD56 expression in NK cells has been associated with the regulation of cytotoxicity. Activated NK

cells express higher levels of polySia and CD56 deletion reduces the tumor killing ability of NK cells[24]. While polySia clearly plays a role in NK cell function further investigation of polySia-CD56 specifics are still required. T cell progenitors in the bone marrow of mice express polySia that is needed for migration to the thymus[32]. Activated peripheral CD4+ T cells in humans have also been shown to express polySia on unknown proteins[33]. Many adaptive immune cells have not been investigated for polySia and performing such studies will greatly add to the understanding of polySia in the immune system.

To study the roles polySia might play as a ligand, receptors need to be present. Sialic acid-binding immunoglobulin-type lectins (Siglecs) are proteins that specifically to Sias and are found primarily on immune cells. While the binding activity of polySia to many Siglecs is unclear, Siglec-11 expressed on tissue macrophages and brain microglia, is a confirmed polySia receptor[34]. The expression of microglial Siglec-11 is specific to humans and has the highest binding affinity to polySia compared to human and chimpanzee tissue macrophage Siglec-11[34]. In a cellular system, removal of polySia from neural cells resulted in decreased binding to Siglec-11 but not complete prevention, potentially due to Siglec-11 binding oligoSia as well[34]. While Siglecs are specific to mammals, siglec-like proteins have been found in nonmammalian species. Given that most of the Siglec-polySia interactions remain unknown, methods for identifying polySia receptors are needed.

Much of the research regarding polySia function has been performed in non-human mammalian models, particularly mice. While having an animal model is important, there a few differences between mice and human polySia. Human polySia is made from α-2,8-linked Neu5Ac while murine polySia can be made from α-2,8-linked Neu5Gc or Neu5Ac[1]. Additionally, murine immune systems have a different cell composition than humans and contain NK cells that do not express polySia-CD56[24]. Lastly, mice do not express the same Siglecs as humans with a significant reduction of *Siglec* genes in mice[35]. Most human Siglecs do not have a true ortholog in mice. These characteristics eliminate the ability to fully apply polySia function discovered in mice to humans and vice versa.

1.7 Aberrant polysialic acid expression in disease

PolySia can also be aberrantly expressed in many neurodegenerative diseases, mental disorders, and cancers where immune system dysregulation occurs. Schizophrenia, bipolar disorder, and autism spectrum disorder have been described as having abnormal polySia involvement[8]. Single nucleotide polymorphisms or genetic deletion in the ST8Sia2 polysialyltransferase have been observed in all 3 disorders[8]. A relationship between polySia and Parkinson's disease, Alzheimer's disease, and Huntington's disease have also been documented, but the specific roles have not been characterized[8]. In multiple sclerosis, polySia is expressed in demyelinated axons but absent in plaques that show partial repair[36]. Due to the significant expression of polySia in the brain, dysregulation can be associated with many neurodegenerative diseases and mental disorders.

Cancers cells have specific phenotypic differences from healthy cells, such as increased DNA damage, resistance to apoptosis, metabolic reprogramming, and changes in glycosylation[37]. A common glycosylation change in cancer tumors is the overexpression of polySia which has been associated with increased metastasis and poor prognosis[38]. Cancerous tissues can be aberrantly expression polySia even if the healthy tissue lacks polySia expression as ST8Sia4 RNA has been detected in most tissues (Human Protein Atlas[21]). While the functions of polySia in cancer remain largely unknown, it has been shown to play a role in tumor migration and anti-adhesion of metastases. High expression of polySia can protect cancer cells from immune cell recognition as polySia is seen as "self" allowing cancer cells to avoid destruction[38]. Currently, polySia has only been characterized on CD56 in cancer, but polySia positive and CD56 negative tumors exist[39] suggesting polySia is on other proteins. Increased levels of polySia-CD56 have also been identified in the serum of cancer patients and correlate with disease severity[40,41]. Although, a return to healthy levels can be observed following treatment and remission[40,41]. Within tumors, polySia levels are highest in type III and type IV classifications, suggesting the potential use for prognostics[38,39]. Due to its involvement in embryogenesis and cancer, polySia can be considered an oncodevelopmental antigen[8]. However, lack of knowledge about how polySia works and what proteins it is attached to limits how we can use polySia as a diagnostic or prognostic marker of disease.

1.8 Systemic sclerosis

Systemic sclerosis (SSc) or scleroderma is a rare, but fatal rheumatic autoimmune disease with no cure and few treatment options. It is 6 times more likely to occur in females, however, males tend to rapidly progress to the most severe form of the disease[42]. Early diagnosis is essential to prevent irreversible fibrotic damage. Scleroderma diagnosis and classification can be difficult due to heterogeneity of the disease[42]. However, it can be characterized into two categories, early limited (elSSc), which primarily occurs in the skin of the hands and feet, and early diffuse (edSSc) which is more severe with myofibroblast differentiation occurring throughout the body. Those diagnosed with edSSc rarely live another 10 years due to the extensive fibrosis of internal organs. While bone marrow transplant for these patients is most common, it still does not greatly extend life expectancy and fibrotic damage cannot be repaired. Earlier diagnostics and more successful treatments are urgently needed to improve the lives of patients diagnosed with scleroderma.

Raynauds phenomenon is a common condition in scleroderma patients. It is characterized as a reversible vasospasm of the arteries leading to the whitening of fingers and toes, often when exposed to cold conditions. While not everyone diagnosed with Raynaud's will develop scleroderma, Raynauds often precedes scleroderma diagnosis[42]. Over half of those diagnosed with Raynauds along with abnormal capillary structure will develop scleroderma in 5 years[42]. Being able to predict who will progress to scleroderma from a Raynaud's diagnosis would be substantially beneficial to providing treatment and preventing severe organ damage.

The understanding of scleroderma pathogenesis is very limited. Environmental and genetic elements are thought to initiate scleroderma disease progression through vascular changes, inflammation, autoimmunity, and fibrosis[42]. DNA damage causes fibroblasts to differentiate into irregular myofibroblasts that express α-smooth muscle actin and are apoptosis resistant. These damaged fibroblasts overproduce collagen starting in the skin and progressing to internal organs reducing the elasticity of affected tissues[42]. As the matrix stiffens from fibrosis, this sends survival signals to irregular myofibroblasts[43]. This continued increase in fibrosis damages capillaries causing larger, disorganized capillaries in elSSc and complete loss of capillary structure in edSSc (Fig 1.8). As the endothelium becomes damaged, an inflammatory

response is mounted, and immune cells begin infiltrating. Many immune cells release cytokines that stimulate myofibroblast function[43]. This increased myofibroblast stimulation further exacerbates the immune response leading to immune dysregulation[42] and potentially immunocompromised patients. The specifics for immune dysregulation are highly unexplored. NK cell involvement in pathogenesis has conflicting results, but overall cytotoxicity appears diminished in scleroderma and NK abnormalities have been confirmed[43–45]. T cells have been shown to be autoreactive in scleroderma and produce pro-fibrotic cytokines[46]. Autoantibodies from B cells have been used in diagnostics, some are thought to induce fibroblast activation, but many of their functions in pathogenesis remain undetermined[46]. A better understanding of scleroderma pathogenesis and immune system involvement will lead to earlier diagnosis and better treatments.

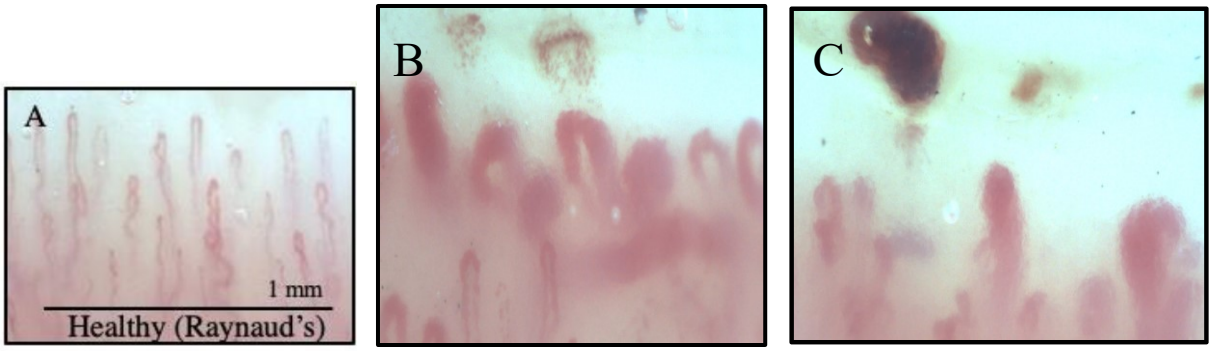

Figure 1.8: **Capillaroscopy of healthy (A), elSSc (B), and edSSc (D) nailfold capillaries.** Images from Dr. Mohamed Osman (Faulty of Medicine and Dentistry, University of Alberta)**.**

Abnormal fibroblasts in scleroderma share many phenotypes with abnormal cells in cancer. DNA damage within fibroblasts and cancer cells contributes to metabolic reprogramming[37,42] (Osman unpublished data). Healthy myofibroblasts are rare due to undergoing apoptosis after the healing process is complete, but like cancer cells metabolic reprogramming leads to an apoptotic resistance. It has also been noted that signalling pathways in scleroderma involve many known oncogenic proteins suggesting similarities in disease progression[47]. Along with the similarities at a cellular level, scleroderma patients are at a higher

risk of developing cancer[47] theoretically due to the pro-oncogenic microenvironment created by scleroderma development. To prevent further disease damage, it is imperative that we understand the cellular mechanisms of both cancer and scleroderma.

1.9 Tools and techniques to study polysialic acid

An antibody, mAb735, has been developed that has high specificity and binding activity for polySia[48,49]. Chains of 8 or more Sia residues form a helical shape and mAb735 specifically recognizes this shape for α-2,8-linked polySia[7]. It is useful for staining polySia in cells and tissues for microscopy and polysialylated proteins in immunoblots[50]. However, special conditions requiring a positively charged nylon membrane are needed for blotting free polySia[51]. Until recently, mAb735 antibody was not commercially available and previous polySia antibodies were either less specific or specific for polySia-CD56.

A bacteriophage endosialidase (EndoN) specific for cleaving polySia is another useful tool. This enzyme has been shown to cleave a minimum of 3 Sia residues on small molecules[52] and 5 Sia residues on proteins[53]. The enzyme was isolated from bacteriophage that bind to and degrade the K1 polySia capsule found on *Escherichia coli*[54]. EndoN forms a homotrimer that is capable of binding 3 polySia chains simultaneously[55]. This trimer conformation also provides the enzyme with structural stability and robustness giving it SDS-resistance[56]. It has since been cloned for easy and inexpensive production and has become a necessary control in polySia research. A catalytically inactive version of the enzyme has been developed through mutating 2 of the 3 essential amino acids in the active site[55]. This inactive EndoN retains binding activity to polySia without degrading it, giving rise to a polySia lectin with high specificity. It has also been expressed as a fusion with green fluorescent protein (GFP)[50] to act as a stable, readily available antibody substitute. In comparison to mAb735 with a K_D of 7×10^{-9} M, EndoN-GFP has a K_D of 1.9×10^{-10} M making its binding activity nearly equivalent to the antibody. It can be successfully used in staining cells and tissue for microscopy and immunoblot analysis of polysialylated proteins[50]. Thus, the active EndoN and $EndoN_{DM}$ lectins proved to be essential tools for studying polySia.

An ELISA using the $EndoN_{DM}$ lectin, polySia antibody and active EndoN was recently developed to probe polySia biology[57]. The GFP-fusion lectin can be adhered to the ELISA plate to capture polySia which is detected by a polySia antibody. Through this technique total polySia concentration in complex mixtures such as serum can be determined. This has been useful in comparing the amount of polySia in healthy versus disease serum from females and males. It can also be used to verify polysialylated proteins with a small modification of the primary antibody to detect a desired protein. Using this method, it was discovered that total serum polySia does not equate to total serum polySia-CD56, suggesting polySia is on other proteins in serum[57]. This modified ELISA is an additional technique used to validate polysialylated proteins. It combines all reliable the tools currently available to give further characterization to polySia in complex mixtures.

Immunoprecipitation is a common method used to isolate proteins from complex mixtures. While this technique does not require a lot of starting material, many protein hits come from mass spectrometry analysis and polySia immunoprecipitation magnifies non-specific hits due to proteins binding the negatively charge polySia. The samples cannot be washed vigorously with detergent to reduce the non-specifically binding proteins without denaturing the antibodies. With many proteins needing validation either some must be cherry picked or many antibodies must be purchased. In order to see progress in the field of polySia research, a more effective technique to isolate proteins from complex mixtures is needed.

1.10 Hypothesis

Based on the phenotypes shared between cancer and scleroderma, I hypothesize that polySia is dysregulated in patients with scleroderma and that this dysregulation may be tied to disease outcomes. I aim to characterize polySia in tissues from patients with scleroderma, with a particular focus on determining what proteins are modified with polySia, since these may provide insight into the role of polySia is disease. Because it is challenging to identify polysialylated proteins using traditional immunoprecipitation, we aim to develop a novel technique that will allow us to identify polysialylated proteins from complex mixtures. I hypothesize that this methodology will allow us to identify numerous previously unknown

polysialylated proteins, not just in scleroderma but in multiple tissues associated with health and disease.

Chapter 2. Methods and Materials

2.1 Biological Samples

NK-92 cells (ATCC; CRL-2407) were cultured in RPMI-1640 (Gibco) supplemented with 10% horse serum (Gibco), 10% fetal bovine serum (Company), 25 mM HEPES pH 7, 1% penicillin-streptomycin (Gibco), 200 U/mL IL-2 (company), and 1 mM hydrocortisone (Stemcell). MCF-7 cells (ATCC; HTB-22) were grown in DMEM (Gibco) supplemented with 10% FBS and 1% penicillin-streptomycin. $CD3^+$ T cells were isolated from human PBMCs (Stemcell) and grown in Immunocult (Stemcell). Fibroblasts from patients with scleroderma and healthy fibroblasts were harvested from skin biopsies by Dr. Mohammed Osman (Faculty of Medicine and Dentistry, University of Alberta) and grown by Lamia Khan in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin. Commercial foreskin fibroblasts (ATCC) grown in the same conditions as scleroderma fibroblasts. Off the clot serum from healthy human females and males (age 23 – 69) was obtained from Innovate Research Inc.Breast milk was received from NorthStar Milk Bank.

Antibodies used in these experiments were α-polySia mAb 735 (1:1000, BioAspect; BA-Ab00240-2.0), α-CD56 (1:500, R&D systems; mAb2408), α-beta-tubulin (1:4000, Abcam; ab6046), α-streptavidin-HRP (1:500, Biolegend; 405210), α-GPR15 (1:500 in 5% BSA, R&D Systems, mAb3654), goat α-mouse IgG-HRP (1:1000, R&D systems; HAF007), and goat α-rabbit IgG-HRP (1:10000, Abcam; ab6721).

2.2 ELISA

The ELISA was performed as previously described by Tajik et al[57]. Briefly, GFP-$EndoN_{DM}$ was adhered to the bottom of 96-well plates at a concentration of 10 μg/ml. Wells

were blocked with 5% BSA in PBS, 0.5% Tween (PBS-T). Serum was diluted 1:40 in 0.1M Tris-HCl pH8 and colominic acid (Sigma) standards were made from 0.1 ng/ml – 6.4 ng/ml. For negative controls, at least one sample was treated with active EndoN. After incubation with primary and secondary antibodies (room temperature 1 h), 3,3',5,5'-tetramethylbenzidine (TMB) was added and absorbance read at 625 nm for 30 minutes. For each plate raw absorbance measurements from healthy controls were averaged. All samples were then normalized to this value. For verifying polysialylation status, T cell lysate was used with α-GPR15 primary antibody.

MalE-EndoN was purified as previously described[52]. Briefly, MalE-EndoN was expressed in *E. coli* following overnight culture in LB media and 0.5 mM isopropyl 1-thio-β-D-galactopryranoside (IPTG) induction. After 24 hr incubation, the cells were harvested and lysed. The enzyme was purified with amylose resin in 20 mM Tris-HCl pH 8, 200 mM NaCl, 2 mM EDTA. MalE-EndoN was eluted with 10 mM maltose in binding buffer and purity assessed with SDS-PAGE followed by Coomassie staining.

2.3 Purification and immobilization of EndoN

The His_6-Avi-$EndoN_{DM}$ was constructed from a 3-way ligation by Samantha Rodriguez. The plasmid expressing GFP-$EndoN_{DM}$[58] was digested with *BamHI*-HF and *NheI* restriction enzymes to produce the *NheI-BamHI* vector fragment and the *BamHI-BamHI* $EndoN_{DM}$ insert. Primers corresponding to *NheI-BamHI* Avi tag were phosphorylated with the T4 polynucleotide kinase (NEB) and annealed. Equimolar amount of the Avi tag and $EndoN_{DM}$ fragments were ligated for 1 h before adding the vector fragment and allowing the ligation reaction and positive transformants were verified by sequencing. The protein was expressed and purified as previously described[57].

His-MalE-BirA was expressed and purified as previously described[59]. Briefly, BirA biotin ligase was harvested from *E. coli* following 0.5 mM IPTG induction. The enzyme was purified using Ni-NTA resin in 20 mM Tris-HCl pH8, 500 mM NaCl 10mM β-mercaptoethanol, 10% glycerol, 10 mM imidazole. BirA was eluted with 300 mM imidazole in binding buffer and purity assessed with SDS-PAGE followed by Coomassie staining.

His-Avi-EndoN$_{DM}$ was biotinylated using BirA at a 1:0.01 concentration in the presence of 0.3mM biotin and 5mM ATP in 25mM Tris-HCl pH 8 for 4 hours. Excess biotin was removed through filter centrifugation. The biotinylated lectin was added to prewashed streptavidin-agarose beads (Sigma) at a concentration of 0.5 mg/ml and incubated for 1 hour with gentle rotation. Beads were washed with 0.1% BSA in PBST followed by PBS washes. Beads were stored in PBS at 4°C until further use.

2.4 Isolation of polysialylated proteins

For isolation of polySia from cells, 10-20 million cells were harvest and lysed in 250 µL RIPA buffer containing 50 mM Tris-HCl pH8, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40, 1 mM phenylmethylsulfonyl fluoride and benzonase (EMD Millipore). All samples were prepared in quadruplicate. For each cell type, 4 biological replicates were each divided in two with one half left untreated and the other treated with active for 1 h as a negative control. SDS was added to a final concentration of 1% and then mixed with 35 µL of immobilized lectin bead slurry in PBS. The samples were incubated with beads for 1 h and then the beads were transferred to mini-columns (Millipore) and washed with 1% SDS in PBS, followed by PBS, and finally 50 mM NH_4HCO_3. Beads were transferred to clean tubes and the liquid was removed before sending for proteomic analysis.

For breast milk samples, frozen breast milk was thawed and centrifuged at 2000 × *g* for 10 minutes to separate the fat, which was removed. 4 mL per replicate was used and PBS was added to bring the pH to 7, after which the polysialylated proteins were immobilized on beads as described above. For serum samples, serum from multiple patients was combined to a volume of 1000 µL and used directly.

2.5 Immunoblotting

Protein concentrations in lysates were measured using a BCA assay (Pierce). To blot fibroblast samples, an equivalent amount of protein was boiled for 30 sec to preserve polySia. Proteins were separated via SDS-PAGE gel and transferred to a PVDF membrane at 60 V for 1 hr. The membrane was blocked with 5% BSA in PBST at room temperature. Following blocking the membrane was cut in half and incubated with α-polySia or α-CD56 at in PBS-T overnight at room temperature. The remaining half of the blot was incubated with 1:4000 α-beta-tubulin. The membranes were washed with PBS-T and incubated goat α-mouse IgG-HRP or goat α-rabbit IgG-HRP. Following washes, the membrane was developed using Pierce ECL Western Blotting Substrate and BioRad ChemiDoc. For samples demonstrating isolation of polysialylated proteins, equivalent volumes of samples were added to the wells. In the final step, the beads were boiled for 30 sec before blotting as above. α-beta-tubulin was used as a loading control.

2.6 Mass Spectrometry and Analysis

All mass spectrometry methods and analysis were performed at the discretion of Dr. Nichollas Scott (Department of Microbiology and Immunology, University of Melbourne). The following methods, 2.6.1-2.6.3, were authored by Dr. Nichollas Scott.

2.6.1 Sample preparation for proteomic analysis

Beads containing the polySia enriched proteins or mock enrichment controls were solubilized in 4% SDS, 100 mM Tris pH 8.5 by boiling them for 10 min at 95°C and the resulting lysate prepared for digestion using Micro S-traps (Protifi, USA) according to the manufacturer's instructions. Briefly, samples were reduced with 10m M DTT for 10 mins at 95°C and then alkylated with 40mM IAA in the dark for 1 hour. Reduced and alkylated samples, including the beads, were acidified to 1.2% phosphoric acid then diluted with seven volumes of S-trap wash

buffer (90% methanol, 100 mM Tetraethylammonium bromide pH 7.1) before being loaded onto the S-traps and washed 3 times with S-trap wash buffer. Samples were then digested with 2µg of Trypsin overnight before being collected by centrifugation with washes of 100 mM Tetraethylammonium bromide, followed by 0.2% formic acid followed by 0.2% formic acid / 50% acetonitrile. Samples were dried down and further cleaned up using C18 Stage[60,61] tips to ensure the removal of any particulate matter.

2.6.2 Reverse phase Liquid chromatography–mass spectrometry

Prepared purified peptides from each sample were re-suspended in Buffer A* (2% acetonitrile, 0.01% trifluoroacetic acid) and separated using a two-column chromatography setup composed of a PepMap100 C_{18} 20-mm by 75-µm trap and a PepMap C_{18} 500-mm by 75-µm analytical column (Thermo Fisher Scientific). Samples were concentrated onto the trap column at 5 µL/min for 5 min with Buffer A (0.1% formic acid, 2% DMSO) and then infused into an Orbitrap Fusion Lumos (Thermo Fisher Scientific) at 300 nL/minute via the analytical columns using a Dionex Ultimate 3000 UPLC (Thermo Fisher Scientific). Samples were separated using a 95-minute analytical gradient undertaken by altering the buffer composition from 2% Buffer B (0.1% formic acid, 77.9% acetonitrile, 2% DMSO) to 22% B over 65 min, then from 22% B to 40% B over 10 min, then from 40% B to 80% B over 5 min. The composition was held at 80% B for 5 min, and then dropped to 2% B over 2 min before being held at 2% B for another 8 min. The Orbitrap Lumos mass spectrometer was operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (300-1500 m/z and a resolution of 120k) and 3 seconds of Orbitrap MS/MS HCD scans of precursors (Stepped NCE of 35;30 and 35%, a maximal injection time of 80 ms, an Automatic Gain Control (AGC) value of 250% and a resolution of 30k).

2.6.3 Proteomic data analysis

Identification and LFQ analysis were accomplished using MaxQuant (v1.6.17.0)[62]. Data was searched against the human proteome (Uniprot: UP000005640) allowing for oxidation on Methionine. The LFQ and "Match Between Run" options were enabled to allow comparison between samples. The resulting data files were processed using Perseus (v1.4.0.6)[63] to compare the enrichment of proteins using student t-tests.

Chapter 3. Polysialic acid is dysregulated in skin and serum of patients with scleroderma

Previous work performed by our collaborator, Mohamed Osman, suggested that scleroderma shares a number of phenotypes with cancers, including high levels of DNA damage, increased proliferation, dysregulated metabolism, and immune evasion (unpublished data, Mohamed Osman). Given the strong connection between polySia and metastasis/poor prognosis in cancer, we hypothesized that polySia may also be dysregulated in scleroderma.

3.1 Fibroblasts lysates from patients with scleroderma express polysialic acid

To determine whether polySia was present in patients with scleroderma, we first examined *ex vivo* fibroblasts from patients with less progressive elSSc, more progressive edSSc, and healthy controls. Fibroblasts represent the main dysfunctional cell type in scleroderma as they transition to myofibroblasts, which are responsible for producing collagen and fibronectin leading to the thickening and stiffening of skin. Samples were collected, by the Osman lab, via biopsy and allowed to grow in the lab[64]. To examine polysialylation in these cells, they were harvested, lysed and protein concentration was measured using a BCA assay. All samples were additionally treated with and without active EndoN as a negative control. Proteins were then separated on SDS-PAGE and polySia was detected via immunoblot. A characteristic high

molecular weight polySia smear at ~250 kDa was observed in most scleroderma fibroblast samples (Fig 3.1 A). Additionally, there appeared to be a potential correlation with polySia and disease severity as early diffuse (edSSc) samples had the darkest polySia bands. Commercial fibroblasts did not produce any polySia (Fig 3.1 B). The high molecular weight band was demonstrated to be polySia due to its absence when the samples were treated with active EndoN. However, we do see a lot of variability between scleroderma samples in both early limited (elSSc) and early diffuse (edSSc). There is observable variation in even the most severe edSSc form of the disease (Fig 3.1), but we can conclusively say fibroblasts from most patients with scleroderma produce more polySia than healthy controls.

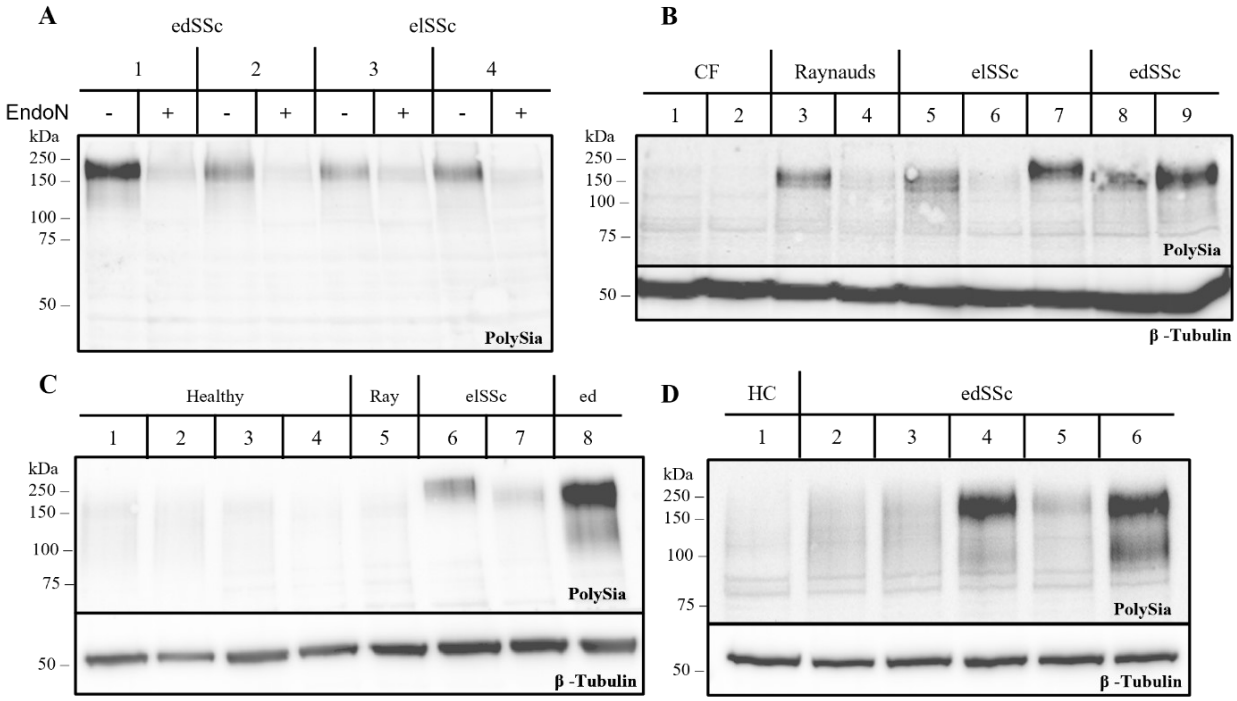

Figure 3.1: **Immunoblot analysis of scleroderma fibroblasts.** α-PolySia (1:1000) blots were used to detect polySia in early diffuse (edSSc, ed) and early limited (elSSc) scleroderma fibroblasts. **(A)** PolySia was confirmed with loss of signal after the addition of active EndoN to the lysate. **(B, C and D)** Commercial fibroblasts (CF) show no polySia expression. Raynauds (Ray), elSSc, and edSSc fibroblasts show variable amounts of polySia expression with the strongest signal present in edSSc fibroblasts. Fibroblasts collected from healthy individuals show negligible polySia expression. All edSSc fibroblasts express more polySia than healthy controls (HC). β – tubulin was used as a loading control in **B, C,** and **D**.

3.2 Polysialic acid in serum is increased in patients with scleroderma

Preliminary data in the Willis lab had shown that polySia was also dysregulated in the serum of patients with breast cancer (unpublished data, Carmanah Hunter). To determine if polySia was dysregulated in serum of patients with scleroderma, we performed a polySia specific ELISA[57]. Patient information is shown in Table 3.2. Due to scleroderma being a rare disease with higher incidence in women, we analyzed fewer male samples than female (Table 3.2). Both sexes showed significant polySia serum level increases with elSSc and edSSc compared to the age matched healthy controls (Fig 3.2). The difference between elSSc and edSSc in females had a significance of 0.001 compared to the 0.0001 significant difference of controls compared to scleroderma serum. However, still both are significant and provide promising potential for using polySia as a diagnostic marker for the most severe form of the disease. Even though the sample size for males was very small, there does not appear to be a significant increase in polySia for males with edSSc compared to elSSc.

	Total	Raynauds	elSSc	edSSc	Healthy
Patients (n)	124	11	45	19	49
Female (n)	91	8	41	11	31
Male (n)	33	3	4	8	18
Age (years), mean (SD)	47 (14)	38 (11)	52 (14)	48 (12)	46 (13)
Age (years), mean (SD), Female	48 (15)	36 (10)	53 (15)	43 (11)	45 (15)
Age (years), mean (SD), Male	48 (11)	44 (13)	47 (11)	54 (12)	47 (11)

Table 3.2: **The total number and average age of female and male serum samples categorized by disease state used to determined polySia and CD56 expression in serum.**

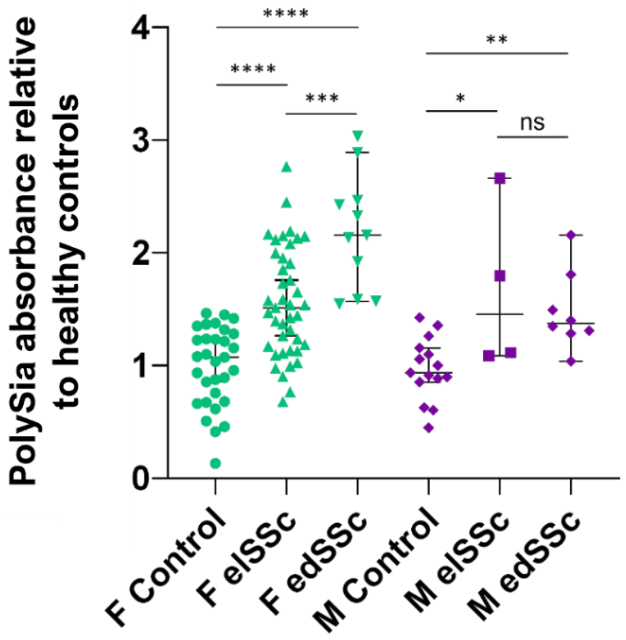

Figure 3.2: **Relative polySia concentration of females and males with scleroderma compared to healthy controls.** Significance was determined through a Mann-Whitney unpaired U test. (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, **** $p \leq 0.0001$, nonsignificant (ns)).

3.3 PolySia in scleroderma serum and fibroblasts is not associated with CD56

Due to the prevalence of polySia-CD56 across biology in both health and disease, we wanted to determine if polySia in scleroderma was a post-translational modification of CD56. We determined that the polySia expression observed in scleroderma fibroblasts was not attached to CD56 (Fig 3.3 A). As a positive control, we analyzed NK-92 cell lysate in parallel as it is known to make polySia-CD56. The Western blot shows significant CD56 signal around 200 kDa for the NK92 cell lysate (Fig 3.3 A). For all patient samples, there is no CD56 signal.

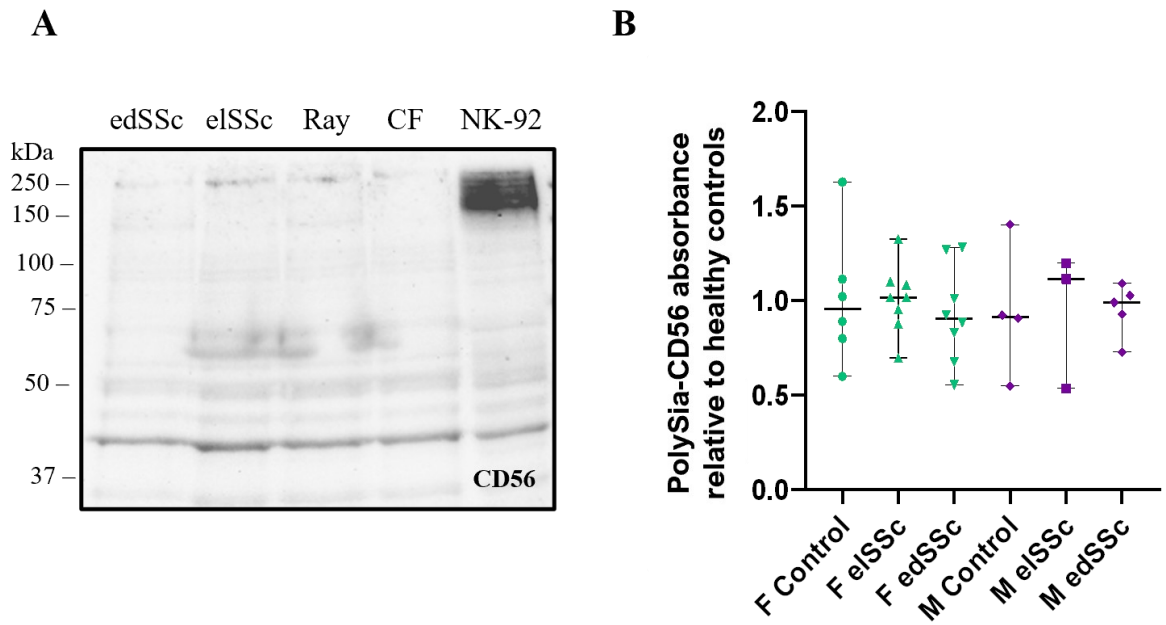

Figure 3.3: **Analysis of CD56 in Scleroderma. (A)** α-CD56 western blot analysis of lysate from early diffuse (edSSc), early limited (elSSc), Raynauds phenomenon (Ray) and commercial fibroblasts (CF) compared to NK-92 cell lysate. CD56 signal only appear in the NK-92 positive control. **(B)** ELISA analysis of polySia-CD56 absorbance in serum from patients with scleroderma compared to healthy controls.

We also performed the ELISA detecting polySia-CD56 specifically. We see that there is no trend or significant difference for polySia-CD56 based on disease severity or sex (Fig 3.3 B). Our observation that polySia is increased in fibroblasts and serum from patients with scleroderma but is not associated with CD56, suggests that there are unidentified proteins made and/or secreted in patients with scleroderma.

Chapter 4. New methodology to identify polysialylated proteins

The unique properties of polySia have added challenges to studying the glycan and isolating polysialylated proteins. PolySia is very labile, with self-cleavage occurring in water, degradation after prolonged freezing, hydrolysis in mild acidic conditions, and degradation with prolonged boiling[65]. Immunoprecipitation is a commonly used method for isolating proteins, however polySia reduces the success of such experiments. Proteomics experiments following enrichment of polysialylated proteins using the α-polySia mAb 735 yields dozens of hits with low sequence coverage. These hits then have to be further screened to select for those for which antibodies will be purchased to verify polysialylation of the protein by immunoblotting, a process that can introduce substantial bias into the results. The poor quality of the proteomics data is due in large part to the abundance of non-specific proteins which bind polySia, the α-polySia mAb, and the streptavidin-agarose beads. While proteins that bind to the α-polySia mAb and the streptavidin-agarose beads can be filtered out using proper controls, those that bind to polySia itself, including numerous nucleic acid-binding proteins, cannot. A new method for enriching polysialylated proteins that allows for removal of non-specifically bound proteins would substantially advance our understanding of polySia biology.

4.1 Immobilized inactive $EndoN_{DM}$ lectin can be used to isolate polysialylated proteins from complex mixtures

To identify novel polysialylated proteins from complex mixtures, we developed a technique that immobilizes a polySia specific lectin. We chose to use the polySia-binding $EndoN_{DM}$ lectin because of its high tolerance to strong detergents that would allow for vigorous washing to remove non-specifically bound proteins. We engineered the $EndoN_{DM}$ lectin to contain a HIS-tag and an AviTag which is a small 15 amino acid peptide that we specifically biotinylated using the *E. coli* BirA enzyme[59] (Fig 4.1 A). A substantial increase in streptavidin-HRP signal was observed in the presence of biotin compared to that of the negative control confirming biotinylation of the lectin. Some signal is present without the addition of biotin,

possibly due to the endogenous biotinylation in bacteria. Once the biotinylation was confirmed, the lectin was immobilized on streptavidin-agarose beads using a protocol established in the lab for other proteins.

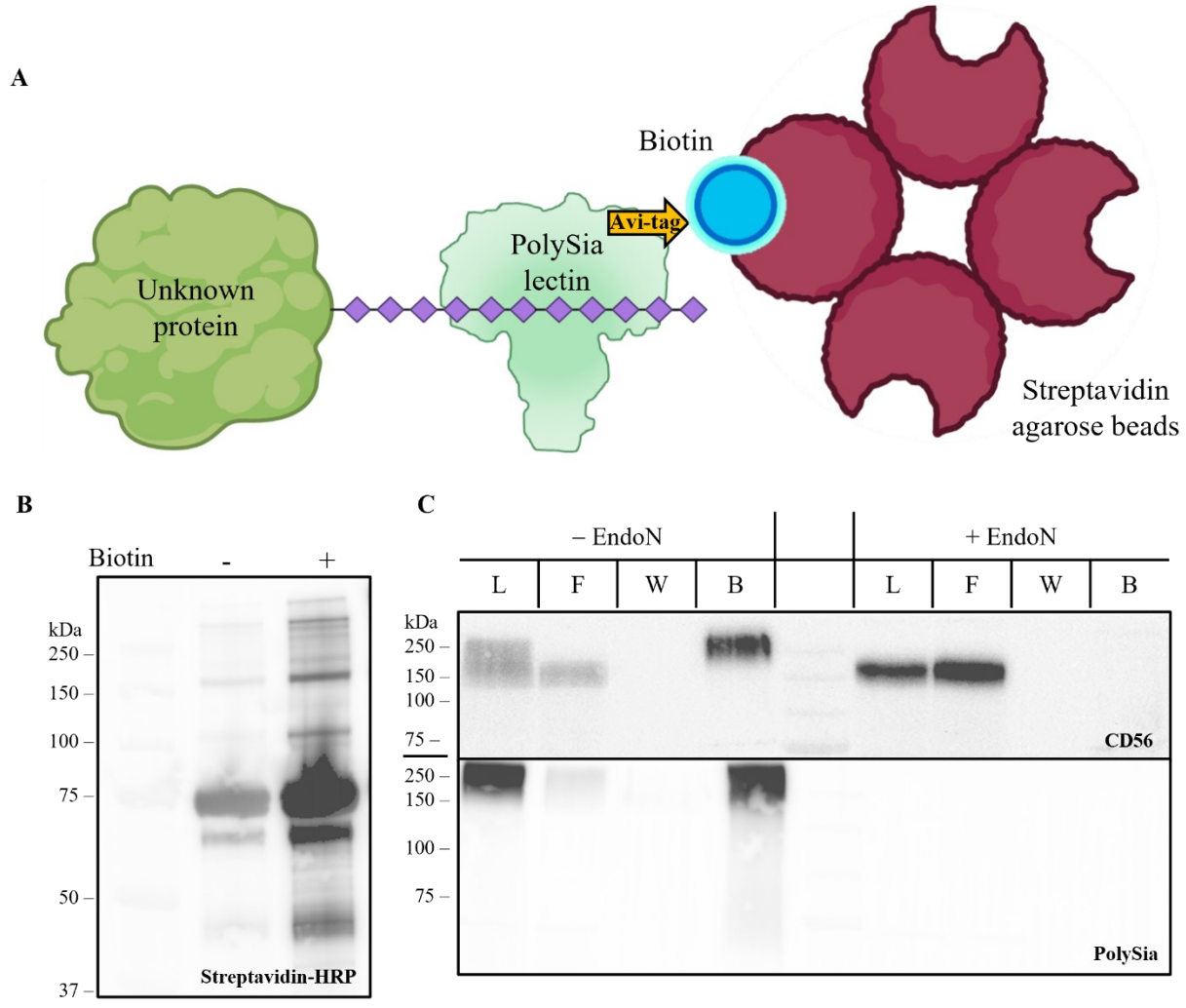

Figure 4.1: **Blot analysis of biotinylation of His-Avi-EndoN$_{DM}$ and affinity purification of NK92 polysialylated protein**. **(A)** Schematic of methodology for isolating polysialylated proteins. **(B)** His-Avi-EndoN$_{DM}$ and BirA mixture was left untreated or treated with biotin. The lectin is visible in the streptavidin-HRP blot (1:5000) at 75 kDa in both samples with substantial signal increase with biotin addition. **(C)** Immunoblot analysis of NK92 cell lysate samples treated without and with active EndoN before isolation on immobilized EndoN$_{DM}$. α-CD56 (1:500, top) and α-polySia (1:1000, bottom) blots showing lysate (L), flowthrough (F), wash

(W), and boiled beads (B) of samples. Immunoblots were performed to confirm presence and absence of polySia, uniform protein amounts were not loaded.

To determine if the immobilized lectin could isolate polysialylated proteins, affinity purification of polysialylated CD56 from NK92 cell lysate was performed and samples were analyzed through immunoblotting (Fig 4.1 B). In the CD56 Western blot, we see the expected smear of polySia-CD56 in lysate but a lower molecular weight band, likely corresponding to unpolysialylated CD56, in the flowthrough. The corresponding polySia immunoblot also shows polySia in the lysate but very little in the flowthrough confirming that polysialylated proteins are binding to the immobilized lectin. Additionally, both CD56 and polySia were observed to be associated with the beads. Pre-treated lysate with the active endosialidase to remove all polySia was used as a negative control. The CD56 blot shows unpolysialylated CD56 in the lysate and flowthrough will no signal present in the bead lane, as would be expected for non-polysialylated proteins. No signal shows up in lane of the polySia blot confirming hydrolysis by EndoN.

4.2 Proteins cannot be easily eluted from immobilized $EndoN_{DM}$ lectin

Next, we wanted to elute our isolated proteins from the lectin for further analysis. First, colominic acid, bacterial polySia, was used in an attempt to displace polySia from the lectin. However, even at concentrations up to 25 mg/ml colominic acid was not sufficient to elute polysialylated CD56 from the beads (Fig 4.2 A). We next tried high temperatures at low pH as polySia is heat- and acid-labile. Unfortunately, extensive boiling and low pH buffers were also insufficient to release the proteins from the lectin-bead complex (Fig 4.2 B & C). There is potentially a small amount of protein eluted post boil (Fig 4.2 B), but we see the majority of CD56 remains on the beads. Given the stability of the association between polySia and the immobilized $EndoN_{DM}$ lectin, we reasoned that elution of polysialylated proteins alone without lectin contamination was unlikely and further trials were not attempted. We decided to move forward with analysis and send non-eluted samples for mass spectrometry.

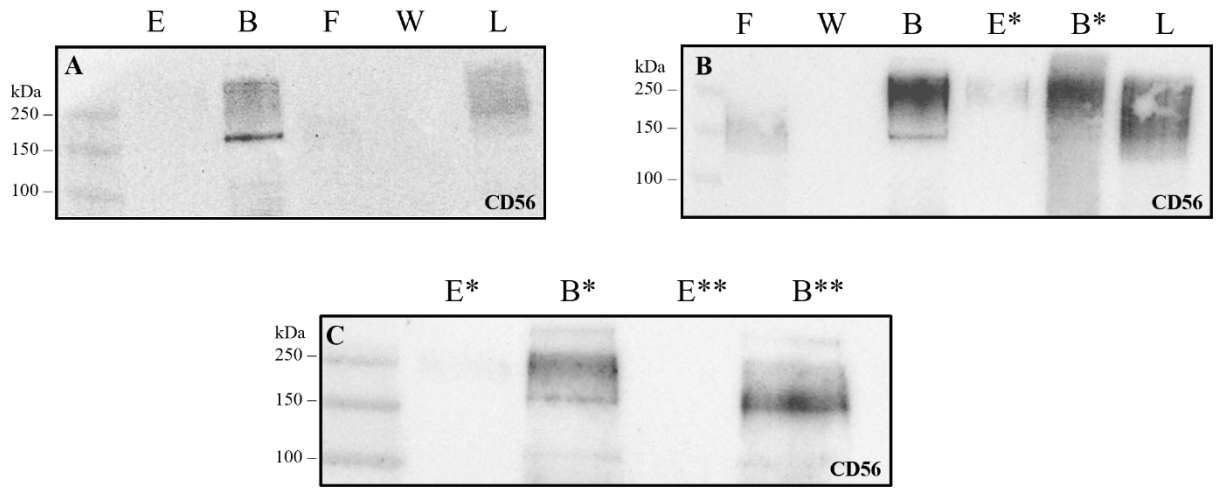

Figure 4.2: **Blot analysis of CD56 elution attempts.** NK92 lysate was incubated with immobilized EndoN and flowthrough (F), wash (W), beads (B) and elutions (E) were blotted for CD56 (1:500). Original lysate (L) was used as positive control. **(A)** Colominic acid (25 mg/mL) was used to elute proteins from the lectin, however no CD56 was eluted. **(B)** Beads were boiled to elute proteins (E*). CD56 signal remained present on the beads post boil (B*). **(C)** Beads were boiled to elute proteins and CD56 signal remained present on the beads post boil. Further elution was attempted with the addition of low pH buffer (pH4) (E**), however no CD56 signal is present in the eluate. CD56 signal remained on the bead after treatment with low pH buffer (B**).

4.3 NK-92 proteomics demonstrates successful proof of concept

To demonstrate that immobilized $EndoN_{DM}$ could be used to isolate polysialylated proteins for subsequent identification for proteomics, we piloted the methodology using NK-92 cell lysates. Replicates of untreated and EndoN treated samples were analyzed by mass spectrometry and visualized using a volcano plot. Unfortunately, >90% of peptides in the mass spectrum could be identified as streptavidin or $EndoN_{DM}$, suggesting that optimization of binding polysialylated proteins to the resin was needed.

In order to reduce the non-specific binding as well as increase the yield and confidence in the identified polysialylated proteins, we tested the ability of proteins to bind to the lectin in the

presence of 1% SDS. PolySia and CD56 retained their ability to adhere to the beads in presence of 1% SDS in the binding and washing steps (Fig 4.3.1).

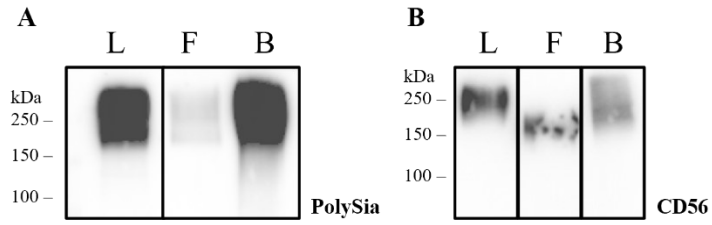

Figure 4.3.1: **Blot analysis of optimizing lectin binding solution.** NK-92 cell lysate (L) was used to determine if 1% SDS effected the ability of polysialylated proteins to bind to the immobilized lectin beads (B) or be present in the flowthrough (F). **(A)** α-PolySia (1:1000) immunoblot shows that very little polySia is lost to the flowthrough and most of it remains on the beads. **(B)** The α-CD56 (1:500) blot shows polysialylated CD56 in the lysate and on the beads, with unpolysialylated CD56 in the flowthrough.

To determine whether the addition of 1% SDS improved subsequent proteomics experiments, we analyzed the beads by mass spectrometry. CD56 was visibly identified as a polysialylated protein being our first hit and with a higher significance value and clear separation from other proteins (Fig 4.3.2). Also identified in the potentially polysialylated hits was ST8Sia4, another proven polysialylated protein[22]. ST8Sia4 has lower T-test difference and significance because it only showed up in 3 of 4 replicates. The majority of the other hits (red) are ribosomal proteins or nucleic acid binding proteins and therefore likely not polysialylated. PolySia-CD56 is the gold standard for polySia research. Successful enrichment of CD56 from NK-92 cells indicates that our new method is feasible for identifying polysialylated proteins.

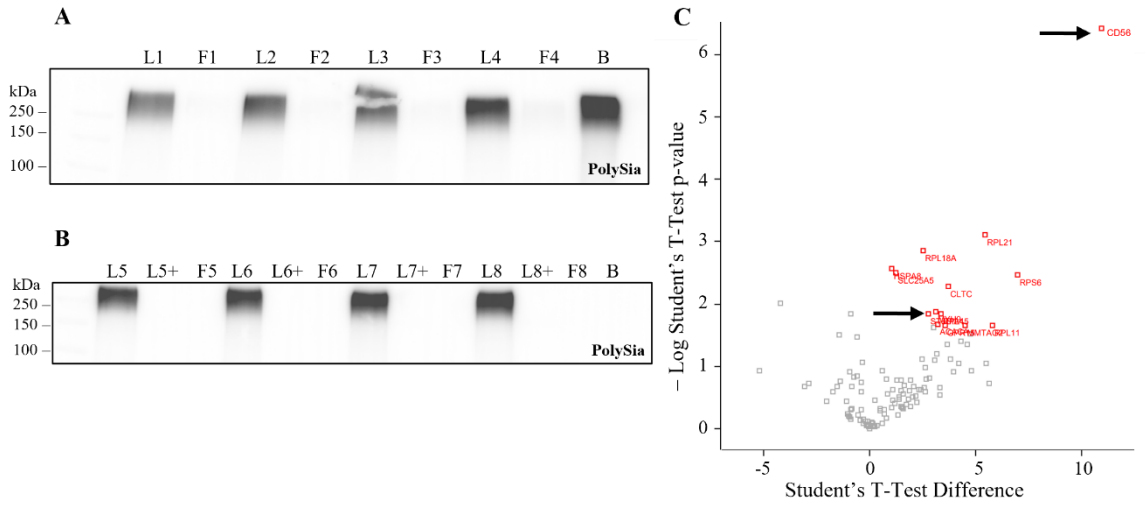

Figure 4.3.2: **Proteomic analysis of NK92 cells. (A)** α-PolySia (1:1000) immunoblot of NK92 lysate (L), flowthrough (F) and combined beads (B) from affinity purification with immobilized EndoN lectin. PolySia signal is only present in the lysate and on the beads. **(B)** PolySia immunoblot of NK92 untreated lysate (L), lysate treated with active EndoN (L+), flowthrough (F) and combined beads (B) from affinity purification with immobilized EndoN. Immunoblots were performed to confirm presence and absence of polySia, uniform protein amounts were not loaded. **(C)** Volcano plot of isolated proteins showing CD561 and ST8Sia4 (black arrows) being identified polysialylated proteins. Experiments were performed in quadruplicate.

Chapter 5. Identification of novel polysialylated proteins in biological samples

5.1 Polysialylated proteins identified in MCF-7 cells are consistent with previous results

MCF-7 is a breast cancer cell line known to make intracellular polySia[66], however the proteins it is attached to are unknown. Our group has recently investigated polysialylated proteins from MCF-7 cells using a complementary metabolic labelling method [67]. Using this

method, seven potential proteins were identified, three of which were confirmed to be polysialylated by ELISA (Fig 5.1 A). However, the metabolic labeling method described in this paper did not yield high quality proteomics data, so we decided to perform analysis of MCF-7 cells using our new immobilized $EndoN_{DM}$ technique. We successfully isolated polysialylated proteins from MCF-7 cells in quadruplicate and with active EndoN controls (Fig 5.1 B & C)

Proteomics results showed that four of the proteins identified in the initial metabolic labelling procedure (Fig 5.1 B) were also identified with the immobilized lectin (Fig 5.1 C). Quiescin sulfhydryl oxidase (QSOX2), Golgi integral membrane protein 4 (GOLIM4), alpha-1,6-mannosyl-glycoprotein-2-beta-N-acetylglucosaminyltransferase (MGAT2), and isoform 5 of 45 kDa calcium-binding protein (SDF4) were identified in both methods. QSOX2 and GOLIM4 were both validated as polysialylated via ELISA, with strong signal that disappeared when the samples were pretreated with active EndoN (Fig 5.1 A). However, we also identified new proteins, including keratinocyte- associated transmembrane protein 2 (KCT2), Golgi membrane protein 2 (GOLM2), alpha-mannosidase 2 (MAN2A1), endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase (MAN1B1), Golgi-associated kinase (GASK1B), and glycoprotein endo-alpha-1,2-mannosidase-like protein (MANEAL) (Fig 5.1 B). Aside from our previous metabolic labelling data, none of the identified proteins are known to be polysialylated.

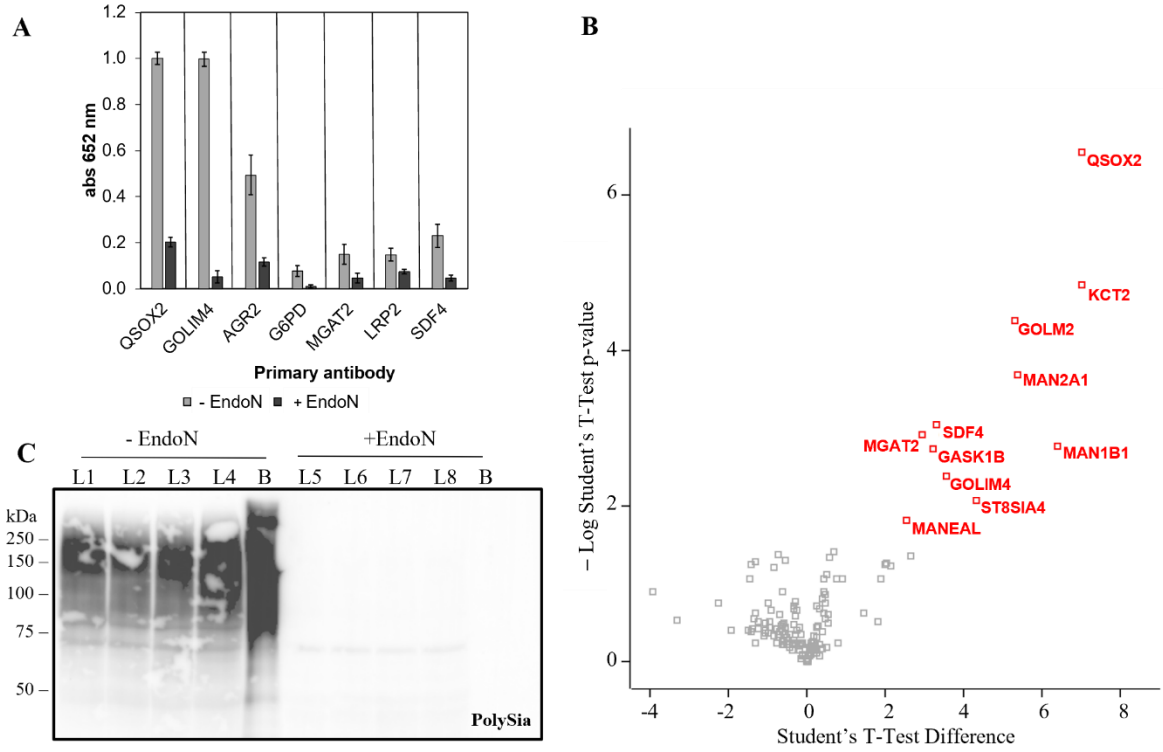

Figure 5.1: **Proteomic analysis of MCF-7 cells. (A)** Validation of polysialylated proteins identified with a previous metabolic labelling technique. ELISA absorbance values of proteins untreated or treated with active EndoN. QSOX2, GOLIM4, MGAT2, and SDF4 proteins were present in both proteomic techniques. Samples treated with active EndoN do not display any polySia signal[67]. **(B)** Volcano plot showing 11 proteins enriched for polySia. Immunoblot was performed to confirm presence and absence of polySia, uniform protein amounts were not loaded. Experiments were performed in quadruplicate. **(C)** α-PolySia (1:1000) immunoblot of MCF-7 lysate (L) and combined beads (B) from affinity purification with immobilized EndoN lectin.

5.2 Identification of polysialylated proteins in human serum

5.2.1 Healthy serum proteomics

The Willis lab previously showed that the concentration of polySia in serum was higher in males than females, but this difference was not attributed to increased polySia-CD56[57]. In

order to determine if the higher polySia values were due to polySia on other proteins we performed our novel protein isolation technique. Our initial protocol involved incubating the immobilized lectin directly in 100% serum. This led to a slight enrichment of polySia-CD56 (Fig 5.2.1 A). Following the optimization of our protocol, we treated serum with PBS buffer to bring the pH closer to neutral and added 1% SDS before incubating the mixture with the beads. This gave us a substantial increase in enrichment of polysialylated proteins (Fig 5.2.1 B). Both the Student's T-test difference and -log p-value were increased compared to our prior method, with a T-test difference of 3.67 to 6.47 and p-value from -log 1.98 to -log 5.46. This improved protocol not only isolated CD56, but ST8Sia4 as well. While it has been previously established that ST8Sia4 is found in serum (Human Protein Atlas[68]) this is the first documented instance of ST8Sia4 being polysialylated in serum.

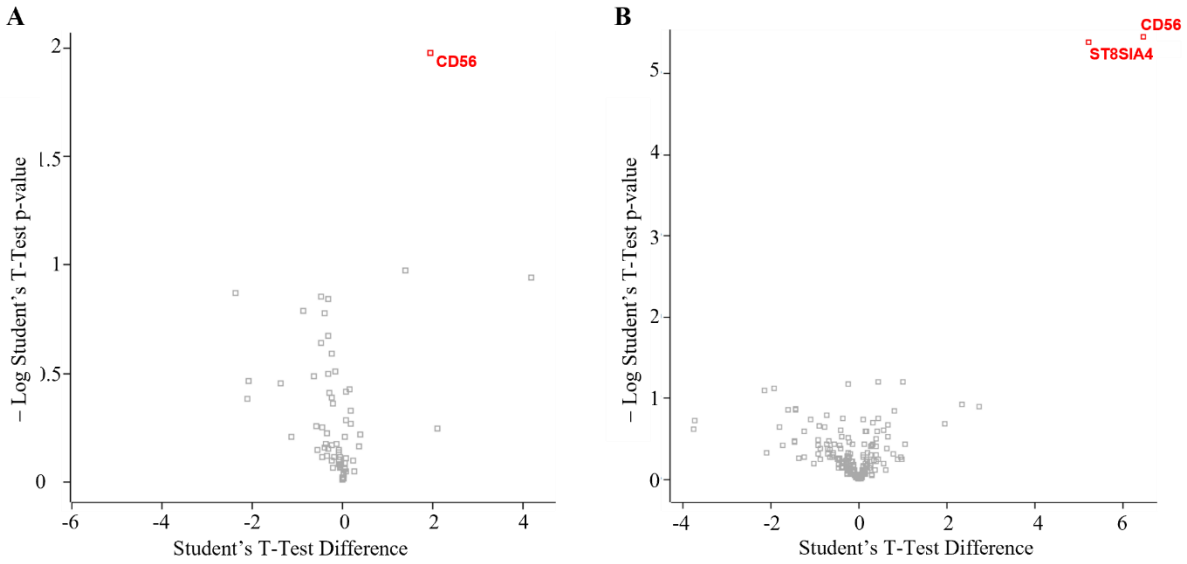

Figure 5.2.1: **Proteomic analysis of healthy serum**. **(A)** Volcano plot analysis using the initial binding protocol without serum pH adjustment and added SDS. CD561 was identified as a polysialylated protein **(B)** Volcano plot analysis using the final binding protocol with serum adjusted to a higher pH and in the presence of 1% SDS. ST8Sia4 and CD561 were identified as polysialylated proteins. Experiments were performed in quadruplicate.

5.2.2 Scleroderma serum proteomics

The original goal of developing the new methodology to identify polysialylated proteins was to determine which of them might be dysregulated in patients with scleroderma. We isolated polysialylated proteins from scleroderma serum and performed proteomic analysis to see what polysialylated proteins may be playing a role in disease progression. Similar to healthy serum, both CD56 and ST8Sia4 were the only polysialylated proteins identified (Fig 5.2.2).

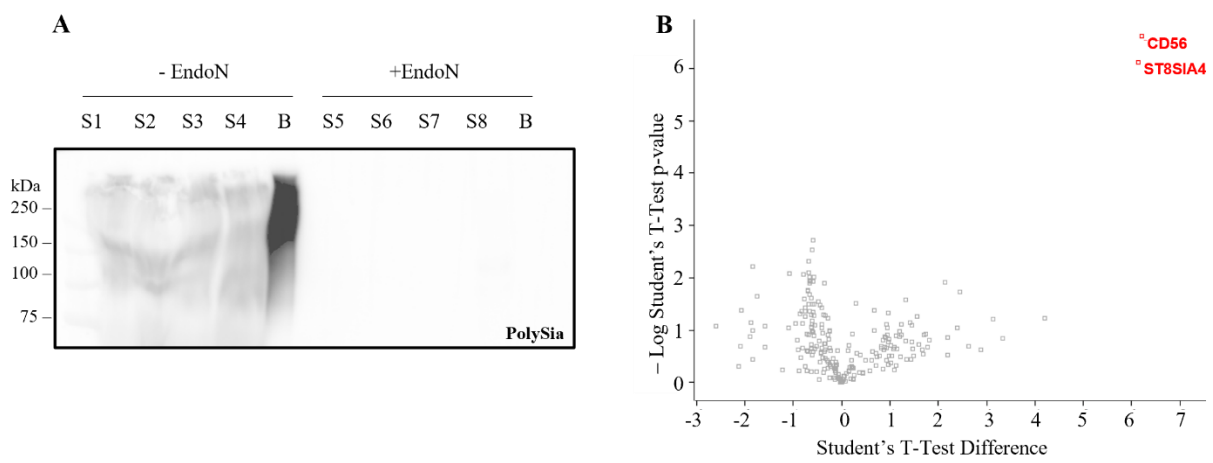

Figure 5.2.2: **Proteomic analysis of serum from patients with scleroderma**. **(A)** α-PolySia (1:1000) immunoblot showing polySia in serum replicates (S) and on beads (B). Samples treated with active EndoN do not show any polySia signal. Immunoblot was performed to confirm presence and absence of polySia, uniform protein amounts were not loaded. **(B)** Volcano plot analysis showing ST8Sia4 and CD561 as being the only polysialylated proteins detected. Experiments were performed in quadruplicate.

5.3 Identification of polysialylated proteins in human breast milk

PolySia has been previously identified in human breast milk, however the polysialylated protein identified, CD36, was not validated[18]. We observed weak polySia staining in human

breast milk, demonstrated by the immunoblot in the absence and presence of active EndoN (Fig 5.3 A & C). Our proteomics technique was applied to breast milk after removing some of the fats to decrease the mixture complexity and reduce non-specific binding. We started with a larger sample volume of breast milk, but similar protein concentration to the cell lines. PolySia was clearly isolated on the beads from the breast milk (Fig 5.3 A).

From the 8 replicates performed, proteomics results showed only two potential polysialylated proteins neither of which were CD36. ST8Sia4 was the top hit and has been identified in serum proteomics (Human Protein Atlas[68]), but not previously shown in breast milk. Further validation is needed to confirm ST8Sia4 as polysialylated in breast milk. Alpha-S1-casein (CSN1S1) was also a top hit. Casein is an abundant milk glycoprotein and further investigation into its polysialylation status is necessary before we can confirm its polysialylation.

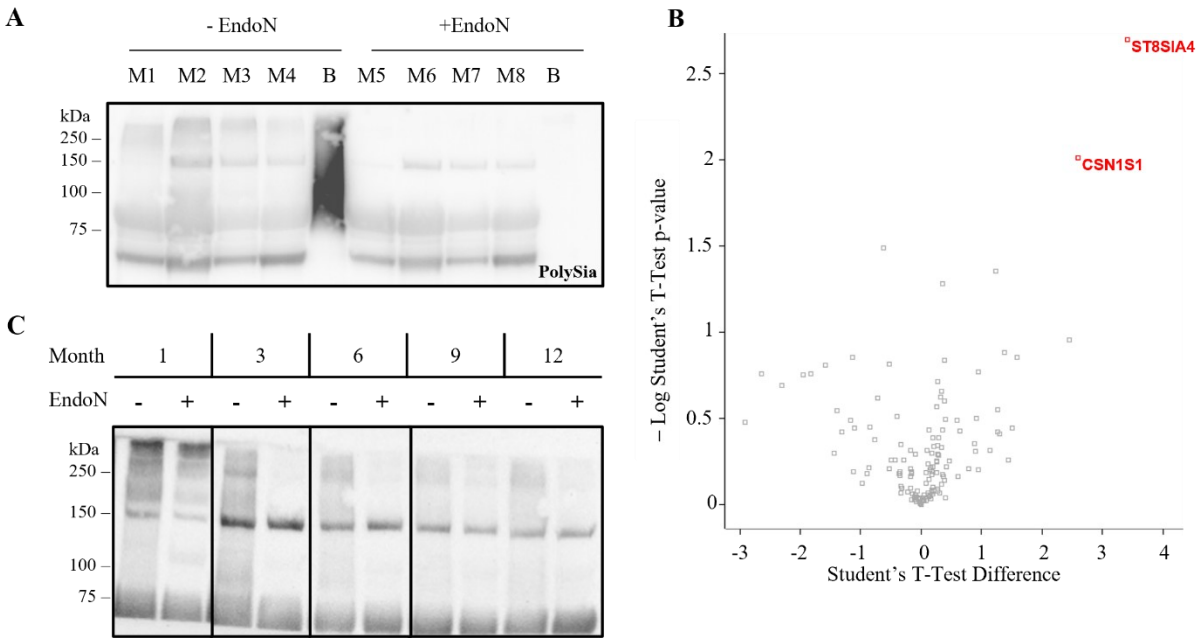

Figure 5.3: **Proteomic analysis of breast milk. (A)** α-PolySia (1:1000) immunoblot showing polySia in breast milk (M) replicates and on beads (B). Samples treated with active EndoN do not show any polySia signal. Immunoblot was performed to confirm presence and absence of polySia, uniform protein amounts were not loaded. **(B)** Volcano plot analysis showing ST8Sia4 and CSN1S1 as being the only identified polysialylated proteins. Experiments were performed in quadruplicate. **(C)** α-PolySia (1:1000) immunoblot showing 1-, 3-, 6-, 9-, and 12-month samples

of breast milk. PolySia was confirmed with absence of signal in the presence of active EndoN. Breast milk used for proteomics was a combination of samples from 1 – 12 months with each replicate being sourced from only one time period.

5.4 A newly identified polysialylated protein in primary human T cells

PolySia expression in T cells has been described[33] but not characterized. Given our success in identifying polysialylated proteins as described above, we were encouraged to look for unidentified polysialylated proteins in primary human T cells. Four replicates of activated primary T cells untreated or treated with active EndoN were analyzed via mass spectrometry. The top hits were ST8Sia4, CD56, G-protein coupled receptor 15 (GPR15), peptidyl-proplyl cis-trans isomerase A (PPIA), and C-C chemokine receptor 10 (CCR10) (Fig 5.4.1 B). We were surprised to observe CD56 in these experiments, since the T cells were purified by negative selection, which should remove all NK and NKT cells from the mixture. GPR15, CCR10 and PPIA are all novel putative polysialylated proteins.

GPR15 was the first novel highly enriched protein from T cell lysate to undergo polySia validation. A Willis lab postdoctoral fellow, Dr. Carmanah Hunter, confirmed its polysialylation through the polySia ELISA (Fig 5.4 C). Lysates treated with active EndoN have a visibly decreased signal compared to untreated lysates, confirming the polysialylation of GPR15.

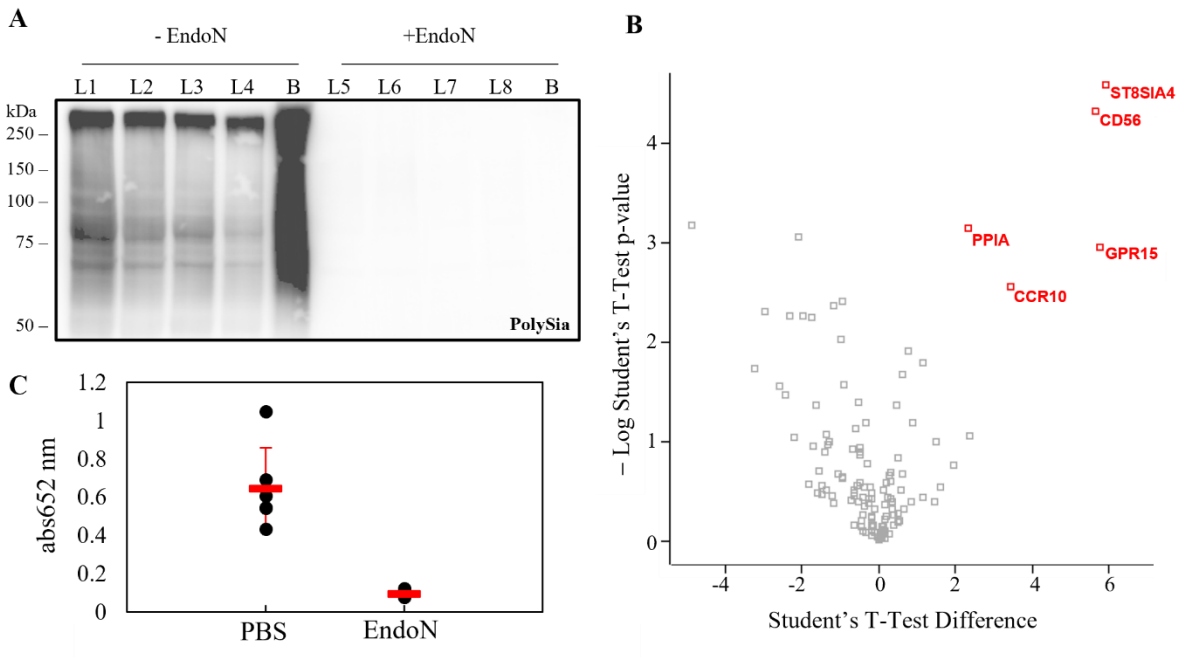

Figure 5.4: **Proteomic analysis of primary T cells**. **(A)** α-PolySia (1:1000) immunoblot showing polySia in lysate replicates (L) and on beads (B). Samples treated with active EndoN do not produce any polySia signal. Immunoblot was performed to confirm presence and absence of polySia, uniform protein amounts were not loaded. **(B)** Volcano plot analysis showing ST8Sia4, CD561, PPIA, GPR15, and CCR10 as being polysialylated protein candidates. Experiments were performed in quadruplicate. **(C)** ELISA probing for GPR15 on captured polysialylated proteins treated with or without active EndoN.

Chapter 6. Discussion and Conclusions

PolySia dysregulation has been characterized in a variety of diseases and is especially correlated with poor outcomes in cancer[38]. We demonstrated the dysregulation of polySia in fibroblasts and serum from patients with the autoimmune disease scleroderma. Our study is the first instance of polySia being dysregulated in an autoimmune disease. However, future evaluation for the presence of polySia in other autoimmune diseases, like systemic lupus erythematosus, will be needed to determine if polySia dysregulation is a general feature of autoimmune diseases as it is for cancer.

We were surprised to find sex-specific differences in polySia dysregulation with females having a larger increase in serum polySia than males, especially for the edSSc samples. Scleroderma has noticeable differences in incidence and disease progression based on sex[42]. While these differences need to be further explored and understood, it's important to note that both sex and gender should be considered. Sex refers to the biological attributes observed at birth while gender refers to the social roles and personal identity of an individual. It has been noted that behavioural differences and occupational hazards may contribute to the more severe disease expression in men[69]. We were limited to having only sex information for patients, but the role of gender remains largely unexplored. The two terms are often used interchangeably in the literature preventing clear disaggregation between the effects of sex and gender not only in scleroderma, but other diseases as well[70]. Our observation that polySia is significantly increased in females with scleroderma compared to healthy controls could be beneficial for developing a diagnostic assay. As it is harder to predict the progression of scleroderma to edSSc in females, a prognostic assay is imperative. Samples taken at time of diagnosis with a 5-year follow up will be required to determine the prognostic potential of polySia. Knowing who will progress to edSSc is critical for physicians to treat their patients since autologous bone marrow transplantation, the only current treatment for scleroderma, has poor success after fibrosis occurs.

Due to the frequent occurrence of polySia-CD56 in healthy and diseased tissues, we investigated whether it could be contributing to the observed polysialylation in scleroderma. We found that polySia-CD56 levels in serum do not increase with disease severity and scleroderma fibroblasts do not express CD56 or polySia-CD56. There are a number of cancers which also

express polySia and not CD56, but the modified proteins have yet to be identified[39,66]. To further understand the role polySia plays in scleroderma disease progression, as well as other pathophysiological conditions, it will be important to determine which proteins are polysialylated.

Method	Strengths	Weaknesses
Immunoprecipitation	• Less starting material required	• Many non-specific proteins identified • Cannot be washed with detergent
Chemically immobilized $EndoN_{DM}$	• Vigorous detergent washes can be performed	• More starting material required • Lectin and contaminating proteins will be chemically coupled at any lysine or cysteine side chain
Avi-tagged immobilized $EndoN_{DM}$	• All lectin oriented for optimal polySia binding • Only Avi-tagged proteins will be biotinylated and attached to beads • Vigorous detergent washes can be performed	• More starting material required

Table 6.1: Comparison of protein isolation methods used for polysialylated proteins.

To identify the unknown polysialylated proteins from complex mixtures related to health and disease, we developed an immobilized $EndoN_{DM}$ lectin technique. The immobilization of the inactive EndoN was not a novel idea; however, previous methods have chemically coupled the enzyme to tosyl-activated beads[12,17]. The coupling of proteins to tosyl-activated beads occurs at

any exposed lysine or cysteine side chain and thus may not be reproducible from batch to batch and in worst cases, may even block the binding sites. These methods have been used to identify some polysialylated proteins, including ST8Sia2[12]. However, our method improves upon previous efforts by immobilizing the enzyme in an optimal orientation which exposes the polySia binding sites (Table 6.1). Additionally, the AviTag only allows $EndoN_{DM}$ to be biotinylated preventing other proteins in poorly purified mixtures from also coupling to the beads. This increases our confidence in the proteins pulled down and helps eliminate some non-specific binding. The success of our method was confirmed by the incredible enrichment of CD56 from NK-92 cells along with the identification of ST8Sia4 in all cell lines studied. This internal control of ST8Sia4 demonstrates the increased effectiveness of this technique.

One of the great successes of this method comes from the proteins identified in MCF-7 cells. We had previously identified seven potential polysialylated proteins from MCF-7 cells using a variation of the chemical coupling of $EndoN_{DM}$ to beads but the quality was poor[67]. Of the seven proteins, four were also detected using our new site-specific immobilization methodology, including QSOX2, the most enriched polysialylated protein. Not only have we confirmed polySia expression on QSOX2, but also that it is secreted from MCF-7 cancer cells. While further validation is needed for the remaining proteins, we have full confidence in QSOX2 being a polysialylated protein. We are especially interested in QSOX2 polysialylation since QSOX2 has been shown to be secreted in lung cancers and, similar to polySia, is associated with poor prognosis[71]. Potential polysialylation of QSOX2 was not investigated in these studies so we are following up on the biological role of polySia-QSOX2 with respect to cancer progression.

An advantage of $EndoN_{DM}$ is its resistance to denaturation in strong detergents, allowing us to query samples that would be challenging to investigate by conventional approaches, such as serum. It was previously noted that polySia levels in serum differ for healthy females and males with females having less total polySia in their serum compared to males[57] and we are interested to know what contributes to this variation. One theory was that the difference is due to females having fewer NK cells[72], but the difference in polySia-CD56, the major polysialylated protein in NK cells, in serum between males and females was not significant[57]. This suggests that the variation in polySia comes from another source. When we analyzed polysialylated proteins in serum, we only identified ST8Sia4 and CD56 in healthy serum. Both proteins have already been

identified in serum, however, polysialylated ST8Sia4 has not been documented and could potentially explain the sex difference. We are now adapting our previous ELISA methodology to determine if ST8Sia4 displays any sex bias.

It is interesting to note that CD56 and ST8Sia4 were also the only identified proteins in serum from patients with scleroderma. We expected something novel to be contributing to the increase of polySia, however the changes may be attributed to increased chain length of polySia or increased amounts of polySia-ST8Sia4. It may also be possible that our method is not sufficiently sensitive to detect low abundance proteins in serum. Further investigation will be required, including identifying polysialylated proteins from scleroderma fibroblast lysate as they may be secreted into serum.

Our results from human breast milk were also unexpected, given previous reports of it containing polysialylated CD36[18]. However, CD36 did not appear in our proteomics results, and we have also not been able to visualize polySia-CD36 in breast milk. We did observe ST8Sia4 in the analysis, though it was not mentioned in the previous study on polySia in breast milk. Validating polySia-ST8Sia4 and its activity in breast milk is under investigation. Alpha-S1-casein (CSN1S1) was also a top hit in our breast milk analysis. Interestingly, alpha-S1-casein can be expressed outside the mammary gland in a few diseases. It has also been shown to influence monocyte differentiation into a macrophage-like phenotype and a proinflammatory response through Toll-like receptor 4 *in vitro*[73,74]. Surprisingly, QSOX2 detected in breast cancer cells, also did not show up in our results. This suggests that the polysialylation of QSOX2 in breast cancer is abnormal and could potentially be used for future diagnostics. PolySia in non-human breast milk is of interest[75] and our new methodology will likely contribute to understanding the roles of polySia in mammalian breast milk.

One of the most fascinating proteins to come out of this project was GPR15 in T cells. GPR15 is transmembrane G-protein-coupled receptor (GPCR) that was identified as being a viral co-receptor for HIV[76,77]. GPR15 has since been found on regulatory, memory and effector T cells, where it plays a role in T cell homing to the colon and inflamed mucosa[77]. With these roles and limited expression on T cells, GPR15 was a very likely candidate for carrying polySia. PolySia expression on GPR15 was validated with removal and addition of polySia. The N-terminal tail of GPR15 has previously been reported as O-glycosylated and capped with sialic

acid[77]. Removal of sialic acid appears to increase binding of the GPR15 ligand. Further validation using the ELISA and immunoprecipitation are in progress to fully confirm GPR15 is polysialylated. In addition to GPR15, CCR10 was identified as a potentially polysialylated protein in T cells. CCR10 is also a chemokine receptor involved in cell migration and we are working to confirm its polysialylation. The presence of two potential polysialylated chemokine receptors in T cells suggests that polySia might play a role in T cell migration, as it does in the nervous and immune systems.

This idea of using inactive glycoside hydrolases to isolate substrates is not novel. Dispersin B, an enzyme responsible for the hydrolysis of the exopolysaccharide poly-*N*-acetylglycosamine (PNAG), has been mutated to probe for PNAG during biofilm formation of medically important pathogens[78]. This protein successfully binds high density PNAG making it specific for visualizing biofilm formation and was fused with GFP for use in microscopy, similar to GFP-EndoN$_{DM}$. Also very recently, the O-glycoprotease StcE was mutated to inhibit hydrolase activity but retain binding ability[79]. This modified enzyme was immobilized using chemical methods and then used to enrich for mucin-domain glycoproteins from complex mixtures allowing for the characterization of known and new mucin domain proteins, including a GPCR[79]. This concept of using immobilized inactive enzymes is becoming increasingly more useful and the need to characterize glycan specific enzymes arising.

In summary, we have characterized polySia in scleroderma for the first time and have developed an effective method for isolation polysialylated proteins from a variety of complex mixtures. We have identified known polysialylated proteins from new sources and have characterized two new polysialylated proteins, with more to be validated in the future. Our technique has great potential to be used in a many human and non-human sources for identifying polysialylated proteins.

References

1. Varki, A. Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution. *Yearb Phys Anthropol* **44**, 54–69 (2001).

2. Ghosh, S. Sialic acid and biology of life: An introduction. *Sialic Acids and Sialoglycoconjugates in the Biology of Life, Health and Disease* 1 (2020) doi:10.1016/B978-0-12-816126-5.00001-9.

3. Varki, A. Sialic acids in human health and disease. *Trends Mol Med* **14**, 351–360 (2008).

4. Ereño-Orbea, J. *et al.* Molecular basis of human CD22 function and therapeutic targeting. *Nature Communications 2017 8:1* **8**, 1–11 (2017).

5. Grewal, P. K. The Ashwell–Morell Receptor. *Methods Enzymol* **479**, 223–241 (2010).

6. Büll, C., den Brok, M. H. & Adema, G. J. Sweet escape: Sialic acids in tumor immune evasion. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1846**, 238–246 (2014).

7. Colley, K. J., Kitajima, K. & Sato, C. Critical Reviews in Biochemistry and Molecular Biology Polysialic acid: Biosynthesis, novel functions and applications. (2014) doi:10.3109/10409238.2014.976606.

8. Sato, C. & Kitajima, K. Polysialylation and disease. *Mol Aspects Med* **79**, (2021).

9. Bonfanti, L. PSA-NCAM in mammalian structural plasticity and neurogenesis. *Prog Neurobiol* **80**, 129–164 (2006).

10. Rutishauser, U. Polysialic acid in the plasticity of the developing and adult vertebrate nervous system. *Nat Rev Neurosci* **9**, 26–35 (2008).

11. Villanueva-Cabello, T. M., Gutiérrez-Valenzuela, L. D., Salinas-Marín, R., López-Guerrero, D. V. & Martínez-Duncker, I. Polysialic Acid in the Immune System. *Front Immunol* **12**, (2021).

12. Simon, P. *et al.* Polysialic acid is present in mammalian semen as a post-translational modification of the neural cell adhesion molecule NCAM and the polysialyltransferase ST8SiaII. *Journal of Biological Chemistry* **288**, 18825–18833 (2013).

13. Galuska, S. P. *et al.* Synaptic cell adhesion molecule SynCAM 1 is a target for polysialylation in postnatal mouse brain. *Proc Natl Acad Sci U S A* **107**, 10250–10255 (2010).

14. Curreli, S., Arany, Z., Gerardy-Schahn, R., Mann, D. & Stamatos, N. M. Polysialylated Neuropilin-2 Is Expressed on the Surface of Human Dendritic Cells and Modulates Dendritic Cell-T Lymphocyte Interactions. *Journal of Biological Chemistry* **282**, 30346–30356 (2007).

15. Werneburg, S. *et al.* Polysialylation and lipopolysaccharide-induced shedding of E-selectin ligand-1 and neuropilin-2 by microglia and THP-1 macrophages. *Glia* **64**, 1314–1330 (2016).

16. Rey-Gallardo, A. *et al.* Polysialylated neuropilin-2 enhances human dendritic cell migration through the basic C-terminal region of CCL21. *Glycobiology* **20**, 1139–1146 (2010).

17. Kiermaier, E. *et al.* Polysialylation controls dendritic cell trafficking by regulating chemokine recognition. *Science (1979)* **351**, 186–190 (2016).

18. Yabe, U., Sato, C., Matsuda, T. & Kitajima, K. Polysialic Acid in Human Milk. *Journal of Biological Chemistry* **278**, 13875–13880 (2003).

19. James, W. M. & Agnew, W. S. α-(2 → 8)-Polysialic acid immunoreactivity in voltage-sensitive sodium channel of eel electric organ. *Proceedings of the Royal Society of London. B. Biological Sciences* **237**, 233–245 (1989).

20. Zuber, C., Lackie, P. M., Catterall, W. A. & Roth, J. Polysialic acid is associated with sodium channels and the neural cell adhesion molecule N-CAM in adult rat brain. *Journal of Biological Chemistry* **267**, 9965–9971 (1992).

21. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (1979)* **347**, (2015).

22. Bhide, G. P., Zapater, J. L. & Colley, K. J. Autopolysialylation of polysialyltransferases is required for polysialylation and polysialic acid chain elongation on select glycoprotein substrates. *Journal of Biological Chemistry* **293**, 701–716 (2018).

23. Stamatos, N. M. *et al.* Changes in polysialic acid expression on myeloid cells during differentiation and recruitment to sites of inflammation: Role in phagocytosis. (2014) doi:10.1093/glycob/cwu050.

24. Drake, P. M. *et al.* Polysialic Acid, a Glycan with Highly Restricted Expression, Is Found on Human and Murine Leukocytes and Modulates Immune Responses. *The Journal of Immunology* **181**, 6850–6858 (2008).

25. Hromatka, B. S. *et al.* Polysialic acid enhances the migration and invasion of human cytotrophoblasts. (2013) doi:10.1093/glycob/cws162.

26. Zuurveld, M. *et al.* Immunomodulation by Human Milk Oligosaccharides: The Potential Role in Prevention of Allergic Diseases. *Front Immunol* **11**, 801 (2020).

27. Miyata, S. ;, Sato, C. ; & Kitajima, K. Glycobiology of Polysialic Acids on Sea Urchin Gametes. *Trends in Glycoscience and Glycotechnology* **19**, 85–98 (2007).

28. Lackie, P. M., Zuber, C. & Roth, J. Polysialic acid of the neural cell adhesion molecule (N-CAM) is widely expressed during organogenesis in mesodermal and endodermal derivatives. *Differentiation* **57**, 119–131 (1994).

29. Weinhold, B. *et al.* Genetic ablation of polysialic acid causes severe neurodevelopmental defects rescued by deletion of the neural cell adhesion molecule. *Journal of Biological Chemistry* **280**, 42971–42977 (2005).

30. Rey-Gallardo, A., Delgado-Martín, C., Gerardy-Schahn, R., Rodríguez-Fernández, J. L. & Vega, M. A. Polysialic acid is required for neuropilin-2a/b-mediated control of CCL21-driven chemotaxis of mature dendritic cells and for their migration in vivo. *Glycobiology* **21**, 655–662 (2011).

31. Kühnle, A. *et al.* Polysialic acid interacts with lactoferrin and supports its activity to inhibit the release of neutrophil extracellular traps. *Carbohydr Polym* **208**, 32–41 (2019).

32. Drake, P. M. *et al.* Polysialic acid governs T-cell development by regulating progenitor access to the thymus. *Proc Natl Acad Sci U S A* **106**, 11995–12000 (2009).

33. Villanueva-Cabello, T. M. *et al.* Polysialic acid is expressed in human naïve CD4+ T cells and is involved in modulating activation. *Glycobiology* **29**, 557–564 (2019).

34. Hane, M., Chen, D. Y. & Varki, A. Human-specific microglial Siglec-11 transcript variant has the potential to affect polysialic acid-mediated brain functions at a distance. *Glycobiology* **31**, 231–242 (2021).

35. MacAuley, M. S., Crocker, P. R. & Paulson, J. C. Siglec-mediated regulation of immune cell function in disease. *Nature Reviews Immunology 2014 14:10* **14**, 653–666 (2014).

36. Charles, P. *et al.* Re-expression of PSA-NCAM by demyelinated axons: an inhibitor of remyelination in multiple sclerosis? *Brain* **125**, 1972–1979 (2002).

37. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).

38. Falconer, R. A., Errington, R. J., Shnyder, S. D., Smith, P. J. & Patterson, L. H. Polysialyltransferase: A New Target in Metastatic Cancer. *Curr Cancer Drug Targets* **12**, 925–939 (2012).

39. Tanaka, F. & Kawano, Y. Expression of Polysialic Acid and STX, a Human Polysialyltransferase, Is Correlated with Tumor Progression in Non-Small Cell Lung Cancer1. (2000).

40. Glüer, S., Schelp, G., Gerardy-Schahn, R. & Von Schweinitz, D. Polysialylated neural cell adhesion molecule as a marker for differential diagnosis in pediatric tumors. *J Pediatr Surg* **33**, 1516–1520 (1998).

41. Glüer, S. *et al.* Serum polysialylated neural cell adhesion molecule in childhood neuroblastoma. *British Journal of Cancer 1998 78:1* **78**, 106–110 (1998).

42. Allanore, Y. *et al.* Systemic sclerosis. *Nature Reviews Disease Primers 2015 1:1* **1**, 1–21 (2015).

43. Van Caam, A., Vonk, M., Van Den Hoogen, F., Van Lent, P. & Van Der Kraan, P. Unraveling SSc Pathophysiology; The Myofibroblast. *Front Immunol* **9**, 2452 (2018).

44. Guilpain, P. *et al.* natural Killer cells exhibit a Peculiar Phenotypic Profile in systemic sclerosis and are Potent inducers of endothelial Microparticles release. **9**, (2018).

45. Horikawa, M. *et al.* Abnormal Natural Killer Cell Function in Systemic Sclerosis: Altered Cytokine Production and Defective Killing Activity. (2005) doi:10.1111/j.0022-202X.2005.23767.x.

46. Chizzolini, C., Brembilla, N. C., Montanari, E. & Truchetet, M. E. Fibrosis and immune dysregulation in systemic sclerosis. *Autoimmun Rev* **10**, 276–281 (2011).

47. Maria, A. T. J. *et al.* Intriguing relationships between cancer and systemic sclerosis: Role of the immune system and other contributors. *Front Immunol* **10**, 3112 (2019).

48. Bitter-Suermann, D. & Roth, J. Competition Heinz Karger Memorial Foundation Monoclonal Antibodies for Diagnosis and Research in Pathology Monoclonal Antibodies to Polysialic Acid Reveal Epitope Sharing between Invasive Pathogenic Bacteria, Differentiating Cells and Tumor Cells. *Immunol. Res* **6**, 225–237 (1987).

49. Sato, C. *et al.* Characterization of the antigenic specificity of four different anti- (α2→8-linked polysialic acid) antibodies using lipid-conjugated oligo/polysialic acids. *Journal of Biological Chemistry* **270**, 18923–18928 (1995).

50. Jokilammi, A. *et al.* Construction of antibody mimics from a noncatalytic enzyme–detection of polysialic acid. *J Immunol Methods* **295**, 149–160 (2004).

51. Willis, L. M. *et al.* Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporterdependent pathways in Gram-negative pathogens. *Proc Natl Acad Sci U S A* **110**, 7868–7873 (2013).

52. Morley, T. J., Willis, L. M., Whitfield, C., Wakarchuk, W. W. & Withers, S. G. A new sialidase mechanism: Bacteriophage K1F endosialidase is an inverting glycosidase. *Journal of Biological Chemistry* **284**, 17404–17410 (2009).

53. Schulz, E. C. *et al.* Structural Basis for the Recognition and Cleavage of Polysialic Acid by the Bacteriophage K1F Tailspike Protein EndoNF. *J Mol Biol* **397**, 341–351 (2010).

54. Pelkonen, S., Aalto, J. & Finne, J. Differential activities of bacteriophage depolymerase on bacterial polysaccharide: binding is essential but degradation is inhibitory in phage infection of K1-defective Escherichia coli. *J Bacteriol* **174**, 7757–7761 (1992).

55. Stummeyer, K., Dickmanns, A., Mühlenhoff, M., Gerardy-Schahn, R. & Ficner, R. Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F. *Nat Struct Mol Biol* **12**, (2005).

56. Mühlenhoff, M., Stummeyer, K., Grove, M., Sauerborn, M. & Gerardy-Schahn, R. Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. *J Biol Chem* **278**, 12634–12644 (2003).

57. Tajik, A., Phillips, K. L., Nitz, M. & Willis, L. M. A new ELISA assay demonstrates sex differences in the concentration of serum polysialic acid. *Anal Biochem* **600**, 113743 (2020).

58. Yu, C. C. *et al.* A glyco-gold nanoparticle based assay for α-2,8-polysialyltransferase from Neisseria meningitidis. *Chemical Communications* **49**, 10166–10168 (2013).

59. Li, Y. & Sousa, R. Expression and purification of E. coli BirA biotin ligase for in vitro biotinylation. *Protein Expr Purif* **82**, 162–167 (2012).

60. Rappsilber, J., Ishihama, Y. & Mann, M. Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663–670 (2003).

61. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols 2007 2:8* **2**, 1896–1906 (2007).

62. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology 2008 26:12* **26**, 1367–1372 (2008).

63. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods 2016 13:9* **13**, 731–740 (2016).

64. Henrot, P. *et al.* A Method for Isolating and Culturing Skin Cells: Application to Endothelial Cells, Fibroblasts, Keratinocytes, and Melanocytes From Punch Biopsies in Systemic Sclerosis Skin. *Front Immunol* **11**, 2408 (2020).

65. Manzi, A. E., Higas, H. H., Diaz, S. & Varkip, &it. Intramolecular Self-cleavage of Polysialic Acid*. *Journal of Biological Chemistry* **269**, 23617–23624 (1994).

66. Martersteck, C. M., Kedersha, N. L., Drapp, D. A., Tsui, T. G. & Colley, K. J. Unique α2, 8-polysialylated glycoproteins in breast cancer and leukemia cells. *Glycobiology* **6**, 289–301 (1996).

67. Hunter, C. *et al.* A new strategy for identifying polysialylated proteins reveals they are secreted from cancer cells as soluble proteins and as part of extracellular vesicles. *bioRxiv* 2022.09.01.506237 (2022) doi:10.1101/2022.09.01.506237.

68. Uhlén, M. *et al.* The human secretome. *Sci Signal* **12**, (2019).

69. Hughes, M. *et al.* Gender-related differences in systemic sclerosis. *Autoimmun Rev* **19**, 102494 (2020).

70. Hunter, C. D., Morris, K. M., Derksen, T. & Willis, L. M. The sweet side of sex as a biological variable. *Glycobiology* (2023) doi:10.1093/GLYCOB/CWAD030.

71. Li, Y. *et al.* QSOX2 Is an E2F1 Target Gene and a Novel Serum Biomarker for Monitoring Tumor Growth and Predicting Survival in Advanced NSCLC. *Front Cell Dev Biol* **9**, 1971 (2021).

72. Phan, M. T. *et al.* Natural killer cell subsets and receptor expression in peripheral blood mononuclear cells of a healthy Korean population: Reference range, influence of age and sex, and correlation between NK cell receptors and cytotoxicity. *Hum Immunol* **78**, 103–112 (2017).

73. Vordenbäumen, S. *et al.* Human casein alpha s1 (CSN1S1) skews in vitro differentiation of monocytes towards macrophages. *BMC Immunol* **14**, 1–11 (2013).

74. Vordenbäumen, S. *et al.* Human casein alpha s1 induces proinflammatory cytokine expression in monocytic cells by TLR4 signaling. *Mol Nutr Food Res* **60**, 1079–1089 (2016).

75. Hinterseher, J. *et al.* Milk Polysialic Acid Levels Rapidly Decrease in Line with the N-Acetylneuraminic Acid Concentrations during Early Lactation in Dairy Cows. *Biology (Basel)* **12**, 5 (2023).

76. Deng, H., Unutmaz, D., Kewalramani, V. N. & Littman, D. R. Expression cloning of new receptors used by simian and human immunodeficiency viruses. *Nature 1997 388:6639* **388**, 296–300 (1997).

77. Okamoto, Y. & Shikano, S. Tyrosine sulfation and O-glycosylation of chemoattractant receptor GPR15 differentially regulate interaction with GPR15L. *J Cell Sci* **134**, (2021).

78. Eddenden, A., Kitova, E. N., Klassen, J. S. & Nitz, M. An Inactive Dispersin B Probe for Monitoring PNAG Production in Biofilm Formation. *ACS Chem Biol* **15**, 1204–1211 (2020).

79. Malaker, S. A. *et al.* Revealing the human mucinome. *Nat Commun* **13**, (2022).