Development and Application of Tools for Glycan Characterization

by

Nia Beckley

B.S., Chemical Biological Engineering (2007)

Massachusetts Institute of Technology

Submitted to the Department of Biological Engineering
in Partial Fulfillment of the Requirements for the Degree of Masters of Engineering on Biological Engineering

at the

Massachusetts Institute of Technology

June 2009

Signature of Autho ..... Department of Biological Engineering
May 15, 2009

Certified By..... Ram Sasisekharan
Thesis Advisor

Accepted By..... Darrell Irvine
BE M.Eng Program Co-Director

Accepted By..... Jay Han
BE M.Eng Program Co-Director

Development and Application of Tools for Glycan Characterization

By

Nia Beckley

Submitted to the Department of Biological Engineering
On May 15, 2009, in partial fulfillment of the requirements for the degree of Masters of Engineering in Biological Engineering

Abstract

Glycans are essential components of all living things because they function as key elements of cellular membranes and extracellular spaces by mediating cell-cell communication, transduction pathways, and cellular development, function, and survival. Because glycans are secondary gene products that depend on the availability of sub-cellular enzymes for synthesis, research on their structure, synthesis, and biological significance has lagged behind that of DNA and proteins due to both a lack of appreciation of their importance and the slow pace at which tools are being developed to study them. In this thesis, three projects focus on the development, application, and exploration of tools for glycan characterization. The first project resulted in the successful optimization of an analytical method to isolate and characterize O-linked glycans, on which relatively few research projects focus because of the limited availability of tools to isolate them and the need for specific analytical equipment to properly characterize them. Using this optimized method, the O-linked glycans of bovine mucin and fetuin were successfully profiled, and the analysis of the former provided the motivation for a second project focused on the significance of goblet cells and mucins in influenza infection. This project explored the potential benefits of a glycoprotein direct binding assay as a way to obtain quantitative information about lectin and influenza hemagglutinin specificities. Using mucins adsorbed to a polystyrene plate, it was possible to obtain quantitative binding constants for two commonly used lectins. The last project focused on the isolation and characterization of the cell surface N-linked glycans from chicken erythrocytes, turkey erythrocytes, and human tracheal epithelial (HTE) cells. Analysis of these cell types is warranted due to their importance as model systems to study influenza infection. The results of this project provide a context for future questions about the relevance of the erythrocyte model system for studying influenza binding specificities. All of these projects reiterate the importance of the study of glycobiology by showing how both the development and application of tools to study glycans can provide in new and interesting information about pathological processes related to human health and disease.

Thesis Supervisor: Ram Sasisekharan
Title: Underwood-Prescott Professor of Biological Engineering and Health Sciences and Technology

Acknowledgments

I feel privileged to have worked under Professor Ram Sasisekharan for four summers and nine academic terms. I began in his laboratory as an undergraduate research assistant who was unsure of what research meant and what career paths were available. Because my experience in the lab, I have discovered my passion for research in the field of health-related science, and I would like to thank Professor Sasisekharan for providing me with this opportunity to grow as a scientist.

Much of my training was possible because of the support and mentorship provided by Aravind Srinivasan, a past post-doc in the Sasisekharan Laboratory. He has guided me since my undergraduate years and has provided me with invaluable support and training. I owe much of my skills and success in glycobiology related research to him, and I am sure that the impact that he has had on my professional life will be felt for decades to come. I would also like to thank Karthik Viswanathan, Zachary Shriver, Rahul Raman, Carlos Bosques, Akila Jayaraman, Ganpan Gao, Pete Wishnok, and Li Li from the Department of Chemistry Instrumentation Facility for taking the time to patiently answer questions about experimental methods and research principles. They have all served as excellent mentors, and I have grown tremendously as a result of their influence. The rest of the Sasisekharan Laboratory was also instrumental to my development and growth, and I will always cherish their support. I would also like to thank the National Institutes of Health for their financial support during my graduate years, without which, none of my successes would have been possible.

I would especially like to thank my parents, whose constant encouragement, support, and love have always inspired me to do my best and face challenges with energy and determination. Even from 3,000 miles away, my mother still took interest in my research and my development as a graduate student by sharing common experiences, dispensing invaluable advice, and encouraging me to reach for the stars. My father's encouragement and words of affirmation, as well as the articles he sends to me about anything related to biological research, have kept me excited about my field, and determined to succeed. Without both of their influences, I would have never been able to reach this point in my graduate career. My brother's youth and energy has also kept me optimistic and balanced, which can be crucial at a place like MIT. Lastly, I would like to thank my boyfriend of 6 years, Michael, for his constant support and undying love, both of which have positively impacted me in every genre of my life.

Table of Contents

List of Figures

List of Tables

Chapter 1

1 Introduction

1.1 BACKGROUND

The majority of membrane–bound and cellular species are glycosylated, making glycosylation one of the most common types of post-translational modification. Glycoconjugates, or biological species having covalently linked carbohydrates called glycans, function as important components of cellular membranes and extracellular spaces by mediating cell-cell communication, cell transduction pathways, and cellular development, function, and survival [1]. Their non-carbohydrate portions (aglycones) are often transcribed under the genetic control of the cell, while their carbohydrate moieties are enzymatically created and later added via the concerted action of a series of glycosidases and glycosyltransferases.

Because their carbohydrate moieties are added via non-template synthesis, glycoconjugates have many different carbohydrate compositions and conformations, and thus have diverse molecular properties and functions. For example, the influenza virus recognizes and interacts with specific glycan residues on glycoproteins in the bronchial epithelium in order to infect these cells [1, 3, 4]. Variations in protein glycoforms can be strong indicators of a change in the virulence of a disease because cell surface binding can be increased or decreased. Thus, the isolation and characterization of glycoconjugates and their glycan moieties in their normal and altered forms could provide important information about the molecular mechanisms of disease pathogenesis.

Since the function of a glycoconjugate can sometimes be greatly attributed to its glycan moieties, many studies have focused on the characterization of glycan structures. Glycans are technically defined as linear or branched carbohydrates composed of monosaccharides linked by glycosidic bonds via a hydroxyl group. The most common monosaccharides found in animals are shown in Table 1.1 with their abbreviations, residue formulas and masses. These monosaccharide units can be neutral or negatively charged (acidic).

Table 1.1 Common Animal Monosaccharide Names, Abbreviations, Formulas, and Residue Masses

Monosaccharide	**Examples**	**Residue Formula**	**Residue Mass**[†]
Deoxyhexose	Fucose (Fuc)	$C_6H_{10}O_4$	146.078
Hexose	Glucose (Glu), Mannose, (Man), Galactose (Gal)	$C_6H_{20}O_5$	162.053
HexNac	GalNAc, GlcNAc	$C_8H_{13}NO_5$	203.079
N-acetyl neuraminic acid	NeuAc	$C_{11}H_{17}NO_8$	291.095
N-glycoyl neuraminic acid	NeuGc	$C_{11}H_{17}NO_9$	307.090

[†]Residue masses are .monoisotopic mass calculated based on C= 12.000000amu, H= 1.007825amu, N = 14.003074amu, O= 15.994915amu). These residue masses do not include water, which is typically lost in the condensation reaction involving two monosaccharides. Since the terminal monosaccharide is only involved in one condensation reaction, the mass of one water molecule must be added to this residue mass.

Glycans composed of at least 2 or more monosaccharide units are called oligosaccharides, while polysaccharides tend to be composed of many repeating oligosaccharide units. The 6 major classes of glycoconjugates include proteoglycans, glycolipids, lipopolysaccharides, peptidoglycans, and glycoproteins. Shown in Figure 1.1 are some of the major glycoconjugates, all of which play important roles in the functionality of a cell's surface and extracellular environments.

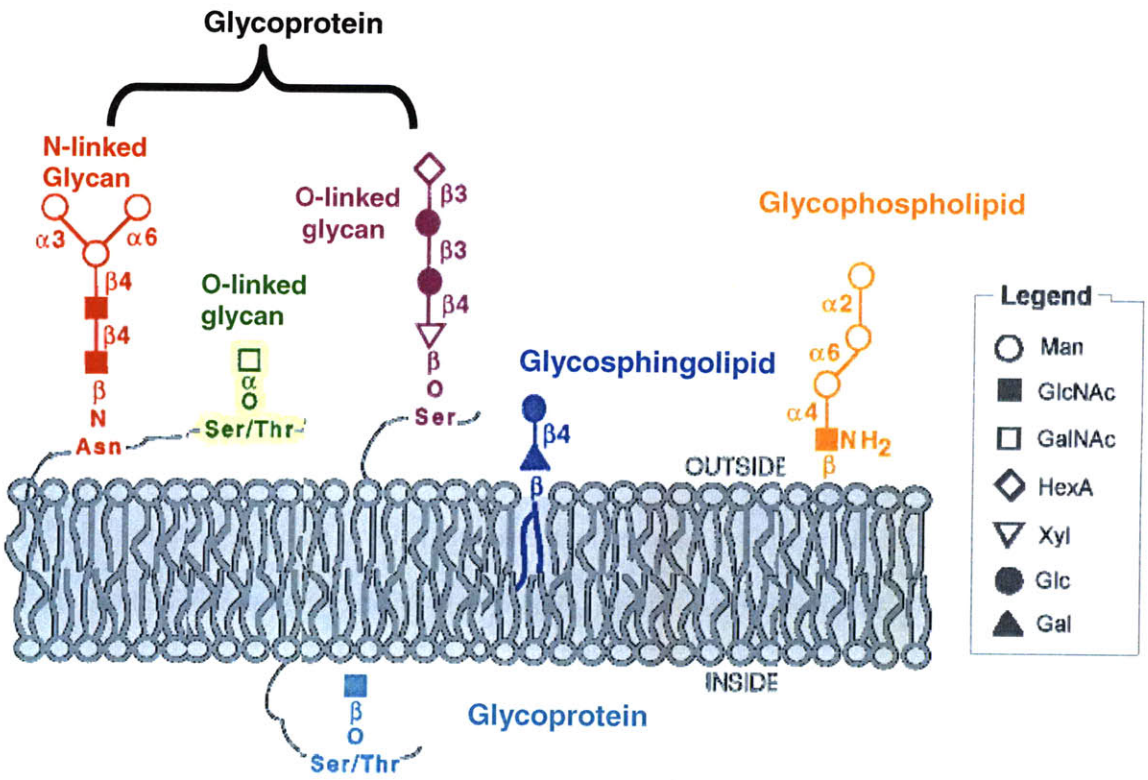


Figure 1.1 Common Glycoconjugates on the Cell Surface. Figure adapted from Varki et al 1999 [1].

This thesis focuses on secreted and membrane-bound glycoproteins, or proteins carrying one or more nitrogen-linked (N-linked via Asparagine) or oxygen-linked (O-linked via Serine or Threonine) glycan. Secreted glycoproteins possess a variety of important biological functions which are often mediated by their glycan moieties. They can serve as structural molecules (collagen), lubricants (mucins), transport molecules (ceruloplasmin and tranferrin), hormones (thyroid stimulating hormone), immunologic molecules, and enzymes (5). Cell membrane associated glycoproteins, on the other hand, are involved in cell-cell, virus-cell, bacterium-cell, and hormone-cell interactions (5). For both classes of molecules, much is still unknown with respect to the structures and functions of their glycan moieties.

Although they appear to differ only by their linkage to the protein backbone, N- and O-linked glycans differ tremendously in structure, synthesis, and biological significance. Because of their size and susceptibility to enzymatic removal, the majority of studies on glycans have focused on N-glycans and

their biological significance. However, there has been an increased interest in O-linked glycans because of the potential roles they play in several biological processes.

1.1.1 N-glycan Structure and Synthesis

The formation of a nitrogen-linked glycan (N-glycan) starts with the synthesis of an oligosaccharide precursor linked to a lipid Dol-P (Figure 1.2). This dolichol lipid structure, which resides in cellular compartments other than the endoplasmic reticulum (ER), is linked via a pyrophosphate linkage to an oligosaccharide containing 14 monosaccharides [1]. This procedure first involves the transfer of two GlcNAc residues to the Dol-P precursor, followed by the addition of five mannose residues, all of which takes place in the cytosol. From here the structure is flipped such that the oligosaccharide faces the lumen of the ER and is further modified with 4 mannose and 3 glucose residues [1]. This structure is then transferred to an asparagine residue on a nascently translated protein in a high energy reaction carried out by the oligosaccharyltransferase (OST) complex [1]. Typically, the oligosaccharide precursor is transferred to an asparagine residue that is a part of the consensus sequence Asn-X-Thr/Ser (where X is any amino acid except for proline) [1].

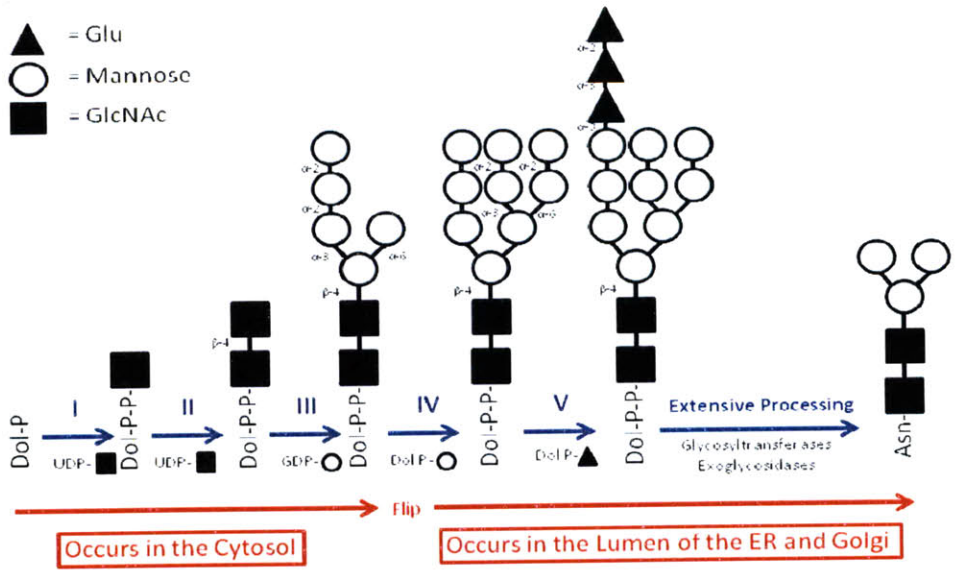


Figure 1.2 Precursor Molecules and Dol-P Glycosyltransferases involved in the Biosynthesis of the N-glycan Pentasaccharide Core. Figure adapted from Varke et al, 1999[1].

After being linked to an asparagine residue, the oligosaccharide precursor, which is high-mannose type because of the prevalence of unsubstituted terminal mannose residues, is then trimmed by α-mannosidases in the ER and Golgi until it is converted into the pentasaccharide core (consisting of two GlcNac and 3 Man residues) traditionally associated with N-glycans (Figure 1.2) [1]. This pentasaccharide core is the subject to diversification by way of exposure to several different glycosyltransferases which can result in the elongation, branching, or termination of the N-glycan chains.

Whether secreted or membrane-bound, N-glycosylated proteins have a diverse array of functions, many of which are dependent on either the local or systemic presence of N glycans in specific peptide regions [6]. According to the SWISS-PROT database, it is estimated that 90% of all of the reported glycoproteins, which make up half of the total number of reported eukaryotic proteins, are N-glycosylated [6]. Although the majority of these proteins have not been studied in the context of their glycosylation, it is thought that N-glycans play crucial roles in their protein signaling processes, stabilization, immune responses, orientation, rigidity, and protein turnover, since they are typically exposed on the surface and have flexible or elongated arms that can extend 3nm or further into the solvent [6].

The different functionalities of N-glycans can be broadly classified into three categories. The first category deals with the role of N-glycans in the protein folding and quality control processes that occur in the ER. These specific roles became evident after an N-glycosylation inhibitor called tunicamycin was applied to cell culture. The resultant studies showed that many peptides were retained in the ER due to improper folding, and were eventually targeted for degradation [7, 8]. Another correlation between N-glycosylation and proper protein folding was drawn with studies on the Aichi strain influenza virus hemagglutinin, which contains only 6 N-glycosylation sites. Without only one of the six glycans (N81), HA cannot form any disulfide bonds because of a destabilizing effect that prevents disulfide bond oxidation, and thus fails ER protein folding quality controls [9]. The second category deals with the role of N-glycans in the intracellular transport and targeting processes that occur in the ER, Golgi, and trans-Golgi compartments. The importance of this function is evident in the trafficking of acid hydrolase

precursors, which are N-glycosidically linked to mannose-6-phosphate and ultimately targeted to the trans-Golgi and lysosomal compartments [6,10]. Lastly, N-glycans play important roles in the function and stability of mature proteins that have been extensively processed in the Golgi. By far, this category has the largest number of specific N-glycan functions since every mature protein can have a different function. Regardless, studies have shown that glycosylated proteins are typically more stable, soluble, and resistant to proteases when compared to their non-glycosylated counterparts. It is hypothesized that N-glycans confer these properties to the stabilized folded protein by reducing the degrees of (mobile) freedom in the unfolded protein [6].

1.1.2 O-glycan Structure and Synthesis

Oxygen-linked glycans (O-glycans) are formed through the addition of a monosaccharide residue, usually a GalNac, to either a serine or threonine residue via an α-O-glycosidic linkage [1]. Unlike N-glycan synthesis, O-glycan synthesis does not require a lipid-linked oligosaccharide precursor in the synthesis initiation event. Instead, a GalNac-transferase initiates synthesis in the Golgi Body by adding a GalNAc residue to a threonine/serine residue [1].

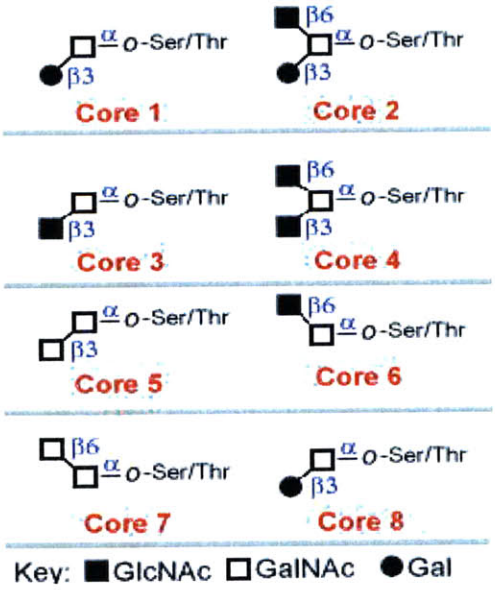


Figure 1.3 The Eight Core Structures of O-linked Glycans. Figure from the Consortium for Functional Glycomics [11].

From this point, the attachment of an additional GalNAc, GlcNAc, or Gal residue determines what is termed as the "core structure" of the O-glycan (Figure 1.3) [1]. In total, there are currently 8 core structures, 4 of which are largely prevalent (Cores 1-4). The existence of many core structures as well as the difference in amino acid linkage distinguish N- and O- glycans, since N-glycans are always linked to an asparagine residue and always contain a common pentasaccharide core region. Furthermore, unlike N-glycans, whose asparagine residue is always a part of the peptide consensus sequence Asn-X-Ser/Thr (where X is any amino acid except proline), O-glycans have no consensus sequence, even though proline residues seem to be abundant in the -1 and +3 positions in heavily O-glycosylated proteins [1].

After the formation of the O-glycan core structure, additional monosaccharides (GalNAc, GlcNAc, Gal, Fuc, or NeuAc) can be added by specific glycosyltransferase enzymes to elongate the glycan in a linear or branched fashion, although the latter is much less common [1]. The number of possible glycan structures, however, is limited by the rules of the O-glycan biosynthetic pathways (i.e. the selectivity of the glycosyltransferases); in other words, certain transferases will only elongate a glycan chain according to the presence of specific terminal monosaccharides linked in specific ways (Figure 1.4). On the whole, O-glycans mainly exist as short linear chains with sialic acid or fucose residues in the terminal position. However, longer O-glycans that contain repeating Gal-GlcNAc-Gal (polylactosamine) units do exist. For example, linear polylactosamines can be subjected to further modification in humans to form the ABO Blood Group Antigens, which essentially describe 3 different types of terminal oligosaccharide groups present on the glycoproteins on the surface of red blood cells [1, 8].

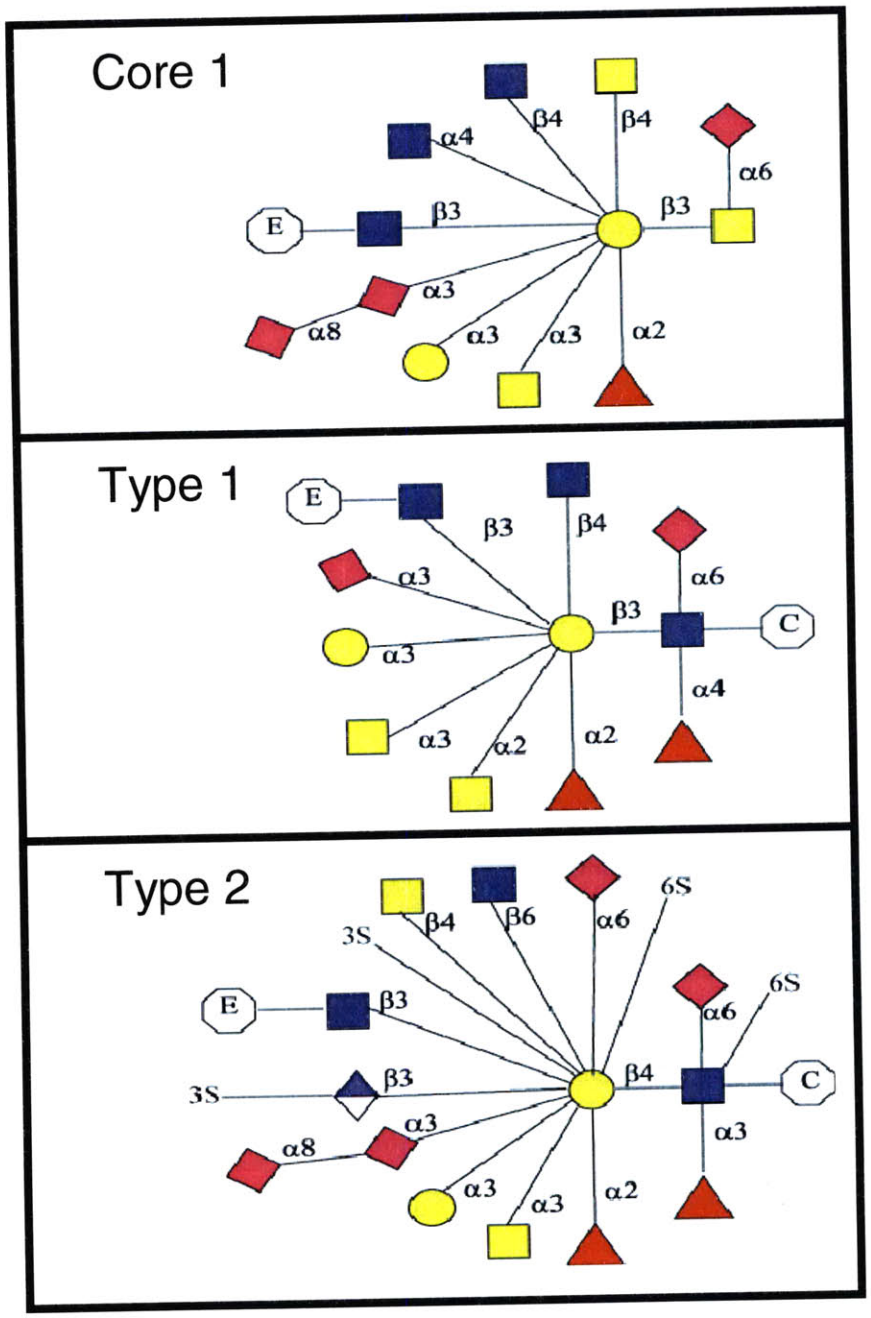


Figure 1.4. Types of Extensions for O-linked Glycan Arms. Core 1 extensions are unique only of Core 1 structures. They may be elongated in a Type 1 or 2 fashion only after the addition of the appropriate Core 1 extension. Type 1 and 2 extensions may be applied to any Core structure, and can include both monosaccharide and adduct (sulphate/phosphate) additions. (E) = Extension, (C) = Core, ◆ = NeuAc, ■ = GlcNAc, □ = GalNAc, ○ = Glucose, ▲ = Fucose, ▽ = GlcA, S= Sulphate (SO_4^-). Figures adapted from the Consortium of Functional Glycomics [11].

Although less understood than N-glycosylation, O-glycosylation has become equally important to our understanding of tissue-specific regulation and disease progression. A number of studies have helped to better elucidate potential roles of O-glycans, mostly in the context of mucins, a family of large and heavily O-glycosylated secreted and cell surface proteins primarily implicated in the maintenance of mucous membranes (Figure 1.5). Since these membranes help to regulate epithelial absorption and secretion, it is thought that mucins, which are usually larger than 200kDa and rich in serine/threonine repeats, regulate these processes through their dense water-retaining O-glycan coatings and form a viscous gelatinous material via mucin disulfide bond cross-linkages and other non-covalent interactions (1,12). These O-glycan coatings tend to look different in various tissues in the body, and since glycosylation is not template driven, the same peptide backbone can have drastically different functions in different tissues.

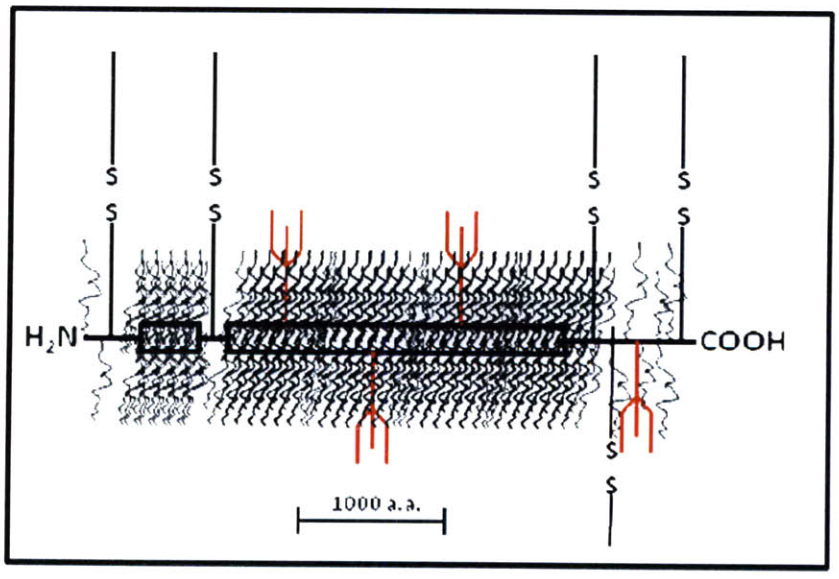


Figure 1.5. Structure of a Typical Secreted Mucin. The apomucin peptide backbone is rich in serine/threonine repeats which give rise to dense regions of O-glycans. Mucins form their thick gelatinous structure through non-covalent interactions as well as disulfide bonds. Note that a small number of N-linked glycans (red) can also be present on the apomucin backbone Figure adapted from Toribara et al 2001[13].

Of the two classes of mucins, secreted and membrane bound, the former seems to be implicated mostly in the protective function of viscous mucus gel, while the latter does not form oligomeric complexes and is implicated in a variety of functions ranging from cellular adhesion to scaffold formation [14]. The broadest definition of a mucin is a protein comprised of at least 50% O-glycans by mass, which leads to the inclusion of rarely studied non-epithelial mucins, which might serve as ligands for selectins and facilitate lymphocyte trafficking [15].

Even though the bulk of mucin structure is dominated by O-glycans, the protein backbone (apomucin), which is encoded by at least 19 human genes (MUC1-20), has gained just as much attention because it is variably expressed in different epithelial tissues throughout the body. All mucins have dense O-glycan coatings, but their protein backbones differ in the number of tandem repeats of a defined number of nucleotides [14], which allows for a large variability in mucin size. Overall, it is thought that the size and length of the mucins are not as important as the generic protein scaffold that they provide for oligosaccharide attachment, which can dramatically improve substrate binding affinities due to the multivalent nature of their dense glycan coating [14].

Interestingly, diseased tissues are noted for changes in both the level of mucin expression and the structural characteristics of the attached glycans. These mucins may express O-acetylated, sulphated, truncated, elongated or more sialylated glycan structures in addition to being over- or under- expressed themselves. Studies of mucins from diseased tissues have offered much insight on their pathological and physiological roles. For example, MUC1, a membrane-associated phospho-mucin that protects cells by binding to pathogens, is aberrantly expressed in many carcinomas and metastatic lesions [1]. It is thought to play a role in tumor progression and metastasis by destabilizing cell-cell and cell-matrix interactions, thus enabling cancerous cells to escape from the epithelial layer and into surrounding tissues [1]. Furthermore, tumor cells that overexpress MUC1 are more likely to be resistant to natural killer and cytotoxic T cell mediated death [1]. On the other hand, mucins from airways affected by bronchitis, asthma, or cystic fibrosis are constitutively expressed, and are produced in excess due to goblet cell

hyperplasia and submucosal gland hypertrophy. Even though it is clear that these mucins express glycans that are short, branched, and sulfated, we do not completely understand if these structures contribute to the pathology of the disease [1].

Still, the study of mucins, glycans, and their biological roles face many challenges. Firstly, it is important to understand that the analysis of the structural and functional role of mucins and their glycans in diseased tissues is hindered by a lack of analogous mucin from normal tissues. Therefore, proof of speculations about the unique structures of glycans and their specific role in disease pathophysiology has been difficult to obtain. Studies aimed at identifying and characterizing both sets of samples will ultimately provide more insight into the molecular basis for disease pathogenesis. Secondly, because glycans make up the majority of the structure of a mucin and critically determine its function, it is necessary to be able to characterize these structures in a way that links form to function. This means that studies of phenomena like cell binding, mucin aggregation, and lymphocyte activation will be best supported with supplemental structural analysis of the glycans of the species being analyzed. Currently, most of these studies are done separately, such that a full understanding of mucin form and function is not achieved. Lastly, unlike most glycosylated proteins, the study of mucins from diseased tissues is limited to a few organs. For example, currently the membrane bound MUC1 has only been characterized and described in lung, breast, and renal cancer [16]. The study of other tissues, diseased or normal, will greatly expand our understanding of how the same mucin functions in different environments. Furthermore, these studies could identify new potential therapeutic targets for many diseased tissues, since it is suspected that most tissues express at least one type of mucin [16].

Even though definitive comparisons of oligosaccharide structures from normal and diseased tissues are not currently possible, some studies do shed light on the physiological roles of specific mucins in various tissues. Salivary mucins, for example, both adhere to and protect tooth enamel demineralization caused by organic acids, as well as defend against oral microbes in the local environment of epithelial cells [17]. Other studies have suggested that mucins on the surface of leukocytes facilitate cell adhesion

events [4]. More specifically, the overexpression of core 2 branched O-glycans correlated to a reduction of interactions between leukocytes and antigen-presenting cells [7], which are key to the proper function of the immune system. Conversely, the elimination of a particular mucin gene leukosialin from the mouse genetic code resulted in hyperimmune responses, since the proteins and their O-glycan coatings are absent [6]. Pulmonary mucins, on the other hand, line the region spanning from the nasal passages to the respiratory bronchioles, and are thought to encapsulate particulate matter and pathogens in order to protect the lower airways and alveoli from obstruction [18].

1.1.3 The Role of Glycans in Influenza Infection

1.1.3.1 Influenza A

The potential pandemic threat of the Influenza A virus, which is a major cause of the yearly flu epidemic, has instigated the investment of billions of dollars in research surrounding its pathogenesis [1]. This virus, which is categorized based on its surface antigens hemagglutinin (HA) and neuraminidase (NA), affects approximately 15% of the U.S. population each year, and results in about 36,000 deaths and more than 200,000 hospitalizations [19]. Even though Influenza A subtypes mainly infect and transmit between wild aquatic birds, two human adapted strains of the virus named H1N1 and H3N2 are the primary subtypes circulating in the human population, and are thus responsible for the yearly influenza outbreaks and epidemics. The real threat from this virus comes from its potential to mutate and become highly pathogenic and virulent, such that its spread would cause serious global health concerns. These pathogenic Influenza A viruses, which either acquire mutations in key viral proteins or obtain new gene segments through reassortment, efficiently transmit from human-to-human, and cause a high mortality rate. In the last century, three major pandemics were caused by influenza viruses: the 'Spanish flu' of 1918 caused by the H1N1 subtype, the 'Asian flu' of 1957 caused by the H2N2 subtype, and the 'Hong Kong flu' of 1968 caused by the H3N2 subtype [20-25]. Based on the estimated mortality rate of the 1918 Spanish flu pandemic, epidemiological studies have projected that a pandemic in the present time would affect roughly 75 million people worldwide [19]. Furthermore, the H5N1 strain of Influenza A, otherwise known as bird flu, is known to have the

highest mortality rate, with greater than 60% of infected individuals succumbing to death [20-25]. However, other Influenza A subtypes, such as H2N2, H5N2, H7N7, and H9N2, have all adapted to infect humans and also present the realistic threat of a new flu pandemic equal in impact, or even greater than, that of the H5N1 Spanish influenza. Fortunately, many of these virus subtypes have not acquired the ability to efficiently transmit between humans; as a result, influenza outbreaks have been relatively localized around the world [25,26]. However, because a serious potential pandemic threat still exists, studies on influenza infectivity and pathogenesis represent the first steps in global preparedness to an outbreak of virulent subtypes.

On a molecular level, Influenza A is a virus of the Orthomyxoviridae family of viruses and contains 8 single-stranded negative-sense RNA strands that encode for 11 viral proteins [27]. Three of these proteins, hemagglutinin (HA), neuraminidase (NA), and matrix protein (M1), are present on the surface of influenza viruses. The other encoded proteins include RNA polymerase subunits (PB1, PB2, PA), nucleoprotein (NP), nonstructural proteins (NS1 and NS2), a proton-selective ion channel protein (M2) and a pro-apoptotic PB1-F2 protein. However, only three of the RNA strands contain the genetic information encoding the three proteins that are key to the first step in the viral infection (HA, NA, and PB1 polymerase). HA is mediator of the first step of viral infection and is responsible for binding to terminal sialylated glycans projected from human respiratory epithelia. HA also contributes to the intracellular release of virion particles in the host cell by way of fusion of the virus with the host cell membrane [28]. Inside the cell, host proteases cleave the viral HA precursor into two subunits linked by a disulfide bond) at a specific amino acid residue (which go on to play a critical role in infection [28]. Neuraminidase, on the other hand, is responsible for cleaving viral cell surface glycan sialic acids, and also has a role in viral entry and the release of mature virions [28]. Each Influenza subtype is characterized by the type of HA and NA present on the viral coating. For Influenza A in particular, 16 types of HA and 9 types of NA have been identified.

Antigenic shift and drift in Influenza viruses are responsible for the yearly appearance of new and more resistant strains of the virus. Antigenic shift occurs when two or more different strains of virus infect the same cell and are susceptible to genomic segmentation. Since different strains may contain drastically different genetic information, antigenic shift can result in major genetic changes. Antigenic drift, on the other hand, is caused by the error prone virus-encoded RNA-dependent RNA polymerases, which frequently make minor genetic errors in the form of amino acid substitutions ($1/10^4$ bases per replication cycle). Overall, while antigenic drift is a continuous process and leads to gradual changes in surface antigens, antigenic shift is an occasional occurrence leading to significant changes in antigen presentation on the virus surface. Since HA and NA are crucial to the virus' ability to initially infect the host cell, genetic changes in these proteins have a large impact on the efficiency and malevolence of the virus. In particular, changes in the HA and NA subtype due to antigenic shift and drift may result in the increased ability of the virus to escape vaccination or neutralizing antibody produced by the host immune system [28]. Some studies have also suggested that mutations in these proteins can lead to human adaptation, since the HA-host cell interactions are purely governed by the properties of the protein-glycan binding sites [3]. More specifically, the crossover from birds to humans is believed to be associated specifically with the ability of HA to switch its binding preference from α2-3 to α2-6 sialylated glycan receptors [29-36].

Even though our understanding of the virus' pathogenecity has improved, we have yet to fully grasp how the virus could acquire the ability to effectively transmit between humans. Understanding this process is key to our ability to develop improved vaccine strategies that rely less on a specific antigen, which often changes rapidly due to antigenic drift, and capitalizes on the host's acquired immune responses, which can most effectively remove infected cells after the initial assault [28]. Furthermore, we cannot begin to address the issue of viral transmission efficiency if we do not have a firm grasp on the molecular basis of the initial virus-host cell binding events prior to infection.

Thus far, we understand that there are three stages in the influenza infection cycle. The first step is the attachment of HA to α2-3 sialylated and α2-6 sialylated glycan receptors of the host cells for birds and humans, respectively [37-39]. This initial binding event is followed by the endocytosis of the virus and the structural modification of HA, which aids in the fusion of the viral membrane with the endosomal membrane. From here, the ion channel activity of M1 is activated, and viral RNA is transported into the host nucleus, where it is replicated and transcribed. After glycosidic modification in the Golgi Bodi, the newly synthesized HA and NA proteins are free associate with newly synthesized viral RNA transcripts to form the virion. These virions then present themselves on the outer surface of the cell, and are released from the cell surface via the action of NA. At this point, these progeny virions can infect other host cells and repeat the entire process.

1.1.3.2 Hemagglutinin and Host Cell Surface N-Glycans

Since the hemagglutinin protein is essential to the initial viral infection events, many researchers have focused on this protein as a means of both understanding the molecular mechanism of viral infection and identifying novel influenza vaccine targets. More specifically, with respect to Influenza A, researchers have focused on the genetic changes in HA that would enable H5N1, which normally infects birds, to infect and transmit efficiently between humans. Genetic changes in the hemagglutinin binding site in particular can lead to a change in the sialic acid binding specificity, and ultimately improve the efficiency of virus infection and transmission, since this binding event appears to be the key determinant of whether or not a particular virus can efficiently infect humans [3,28]. For example, avian influenza viruses of the H5N1 subtype only bind to 2-3 linked sialic acids bound to a penultimate galactose residue (on N-glycans), while their human counterparts bind to 2-6 linked sialic acids linked to a penultimate galactose residue [28]. This difference explains the differences in viral infectivity of birds and humans, since human and bird tracheal epithelial tissues mainly express glycans that have terminal α2-6 and α2-3 linked sialic acids, respectively [28]. Through H5N1 HA binding assays on glycan with different sialic acid linkage types and lengths, research has shown that α2-6 linked sialic acids preceded by repeating lactosamine units (Gal-GlcNAc) provide a larger 3-D space over which the glycan can move after it is bound to the

viral HA [3]. Analyses such as these will become important in accurately describing the binding specificities of different HA subtypes, and ultimately identifying therapeutic or vaccine targets that will actually confer protection against specific virus subtypes.

Besides understanding HA's binding specificity, researchers have also attempted to characterize the glycans on the relevant host cell surfaces so that the binding of a particular HA to a particular host cell type can be predicted [3]. This piece of the puzzle may prove to be the most important, since the HA binding pocket of an influenza virus that effectively transmits between humans will have had to mutate to specifically fit the glycan topologies present on the human respiratory tract cell surfaces. Previous research has shown that the cell surface N-glycans of human bronchial epithelial cells (HBE) are mostly composed of N-linked glycans which contain long repeating lactosamine units capped with a α2-6 linked sialic acid [3]. Therefore, it is postulated that the binding pocket of avian H5N1 strains must mutate to be able to bind glycans of this topology in order to infect human bronchial epithelial cells.

1.1.4 Tools to study HA-Glycan Interactions

1.1.4.1 HA Agglutination and Inhibition Assays

Avian influenza viruses are able to agglutinate erythrocytes (red blood cells, RBCs) through the binding of HA to receptors, and therefore provide an experimental system with which researchers have classified wild-type and mutant influenza A viruses, screened potential HA inhibitors, and identified new vaccine targets [3]. In this agglutination assay, if viral particles are present at a high enough concentration, a diffuse lattice of linked erythrocytes will form, preventing the erythrocytes from precipitating as a small pellet at the bottom of a well in a microtiter plate. RBCs from species such as chicken, turkey, equine, and guinea pigs are typically used in this assay, and because these erythrocytes have a mixture of α2-6 and α2-3 linked sialic acids, their respective cell surface glycans are de-sialylated using sialidase enzymes, and then re-sialylated using specific α2-6sialyl-tranferases [40].

The agglutination assay is the primary tool used to classify HA subtypes by sialic acid specificities, screen potential HA inhibitors, and identify potential vaccine candidates because of the ease with which it can be performed and the low cost of acquiring all of the necessary materials [40]. However, some researchers have postulated that these assays may not be the best way to study HA-glycan interactions because they might not describe the HA-glycan interactions that are relevant in the influenza infection process [3]. It is possible that the glycans modified by simple de-sialylation and re-sialylation of erythrocyte cell surface glycans do not represent the natural glycan ligands present on the cell surfaces of human tracheal and bronchial epithelial cells. Ultimately, the success of this tool depends on the overlap between erythrocyte and human upper airway cell surface glycans, both of which are still relatively unexplored. It will therefore be necessary to compare and contrast the glycan topologies present on both the surfaces of commonly used erythrocytes and human tracheal and bronchial epithelial cells in order to begin to draw conclusions about HA binding specificities in humans.

1.1.4.2 HA-Glycan Tissue Binding Studies

HA-glycan tissue binding studies have helped to reveal some information about the glycans present on the surface of the human tracheal and bronchial epithelia. Typically, de-paraffinized and re-hydrated human upper airway tissue sections are incubated with labeled lectins, each of which bind to specific terminal glycan residues and can ultimately be detected by fluorescence microscopy after the addition of the appropriate reagent, such as Alexa fluor 546 streptavidin [3] (Figure 1.6). These studies have revealed that human tracheal and bronchial epithelia predominantly express α2-6 sialylated glycans; these glycans are omnipresent in the nasal mucosa, on epithelial cells of paranasal sinuses and the pharynx, and in the trachea and bronchi [43]. Knowledge of the glycans present on these surfaces is specifically important in the context of influenza infections because they are the first to be exposed to the free virion. On the other hand, glycans in the human deep lung alveolar cells, as well as those present at the infection site in birds, are mainly α2-3 sialylated [43]. Knowledge of the sialic acid residues present on these surfaces is beneficial to our understanding of how non-adapted avian influenza subtypes infect humans; for

individuals who breathe in large amounts of virion will be susceptible to virion aggregation, virion descent into the lower airways, and infection of these surfaces.

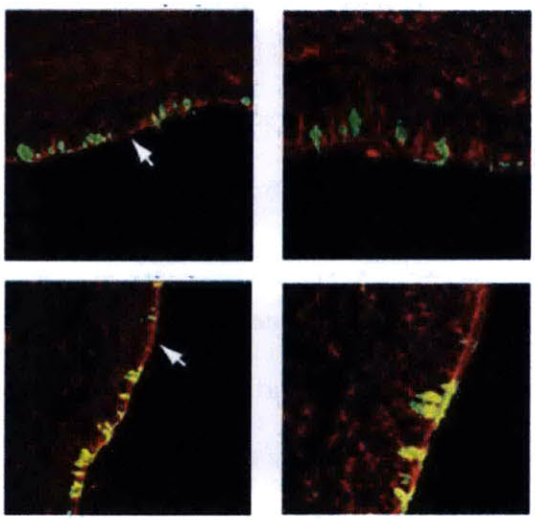


Figure 1.6. Co-staining of Tracheal Tissue Sections with Con A (red)/Jacalin (green) *[top two panels]* and SNA-I (red)/Jacalin (green) *[bottom two panels]*. For both the top and bottom panels, the figures on the left are at 25x magnification, while the figures on the right are shown a 63x magnification. Con A binds branched α -mannosidic structures, which includes high-mannose type, hybrid type and biantennary complex type N-linked glycans. SNA binds α2-6 linked Neu5Ac residues that are linked to a penultimate Gal residue. Jacalin binds to a number of residues containing GalNAc, with the strongest binding occurring to the motif (Sialic Acid) Galβ1-3GalNAcα1-Serine/Threonine. In the top Figure, Jacalin binding is distinct and localized to the goblet cells expressing O-linked glycans, while the regions of Con A binding correspond to ciliated cells expressing N-linked glycans (white arrow). SNA, however, extensively binds to both the goblet cells (co-stain with Jacalin in yellow) and the ciliated cells expressing N-glycans (red) [3].

While these studies have helped to reveal the global distribution of the terminal sialic acid residues present on the surface of human respiratory tract tissues, they still do not help to elucidate the types of glycan compositions present on these surfaces. Ultimately, in order to fully and specifically describe HA-glycan interactions, it is necessary to analyze the structural compositions of the glycans present on these cells so that our understanding of HA-glycan specificities can be improved.

1.1.4.3 HA-Glycan Array Studies

While the tissue binding studies can provide useful information about the terminal glycan moieties present on cell surfaces, other more specific arrays have helped to further describe HA-glycan specificities. An ELISA-like assay using biotinylated α2-3 and α2-6 sialylated glycans bound to a streptavidin coated plate have been used to determine virus receptor specificity (Figure 1.7). These assays are especially relevant to the study of HA binding to cell surface N-glycans because they theoretically present glycan moieties in a manner that is similar to the way in which cell surface glycoproteins present these same moieties. In these assays, recombinantly expressed HA viruses in their natural and mutant forms from H1, H3, and H5 subtypes have already been classified according to their binding preferences to α2-3 and α2-6 sialylated glycans of different lengths [3]. These studies have specifically shown that H1N1, for examples, binds strongly to the 6'SLN-LN glycan, but not to 3'SLN, 6'SLN, 3'SLN-LN, or 3'SLN-LN-LN [3]. This means that this HA subtype specifically requires long α2-6 sialylated glycan arms for binding, and that the glycan topology or composition, not just the sialic acid linkage type, is important in the study of HA-glycan interactions. Furthermore, the effects of sulfation and fucosylation on HA binding have also been determined using these assays. This assay has also been used to quantify the multivalent HA–glycan interactions, and explain how subtle mutations in HA amino acid sequences led to quantitative differences in glycan binding. The results of studies such as these ultimately show how mutations in HA proteins could lead to human adaptation of the virus by changing HA-glycan binding specificities.

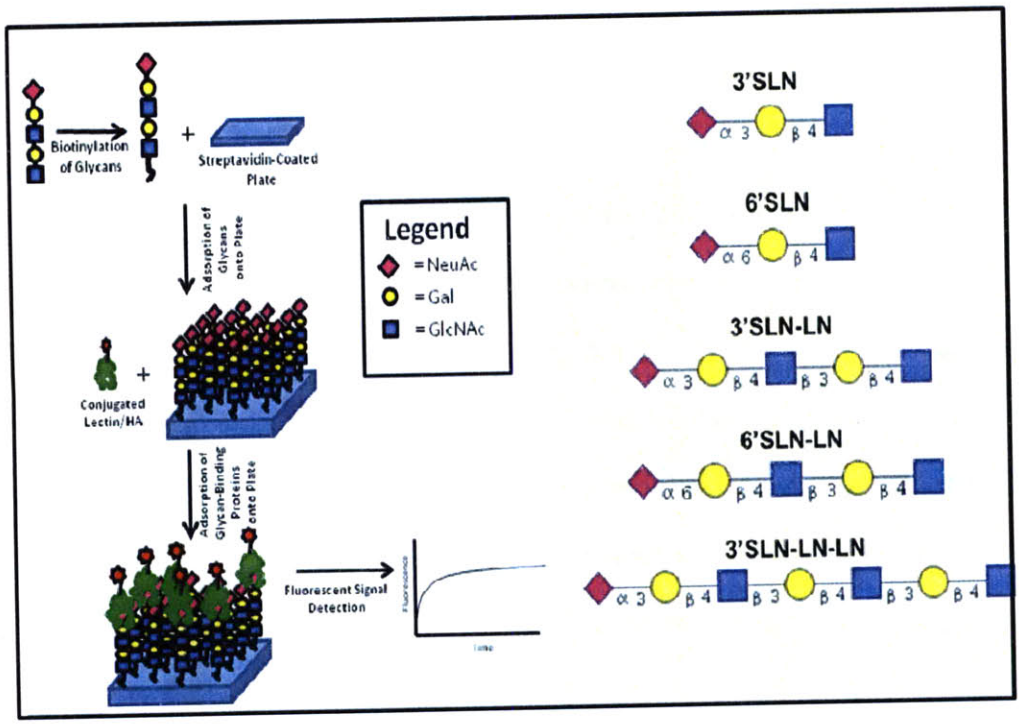


Figure 1.7. Conceptual Diagram of the Biotinylated Glycan Array. The five types of synthetic biotinylated glycans (left) were absorbed into separate wells on a streptavidin coated plate, and exposed to a glycan-binding protein labeled with a tag that would allow for its detection by fluorescent signal (either directly or indirectly) (right).

The direct binding assay can provide important quantitative binding affinities, such as the dissociation constant, K_d. There are several ways to conduct the assay by using different reporter proteins and labeled lectins, but for the work performed in this thesis, a FITC-conjugated protein of interest (lectin or hemagglutinin) is added to the array at various concentrations in a dose-dependent manner. The wells are then incubated with HRP-conjugated anti-FITC antibody, and then exposed to hydrogen peroxide (H_2O_2) and Amplex reagent so that a brightly fluorescent product is created [41]. The experimental rationale behind the calculation of these binding constants is that the fluorescent signal for each well is a direct readout for the number of lectins bound to the glycans. Since both the fluorescent signal and the rate of HRP reaction decrease as the substrate is used up, the most accurate indication of the number of lectins in each well is the initial slope of the fluorescent curve (in units of fluorescence per time). For each lectin concentration, a unique initial slope can be obtained, and the fractional saturation of lectins on glycans

(the concentration-specific initial slope divided by the maximum initial slope) can be calculated. From here, the Hill equation, which describes multivalent kinetic interactions, can be used to calculate the apparent dissociation constant, K_d.

$$y = \frac{[HA]^n}{[HA]^n + K_d'} \qquad [1]$$

In this equation, y is the fractional saturation, and n is the cooperativity factor. The linear form of this equation,

$$\log\left(\frac{y}{1-y}\right) = n * \log([HA]) - \log(K_d'), \qquad [2]$$

can be used to calculate K_d by plotting the log of the lectin or HA concentration against the log(y/1-y). The negative of the y-intercept is the log (K_d), while the slope is the cooperativity factor. To obtain accurate quantitative numbers, it is essential that enough data points be collected so that the trendline can be realistically determined (R^2> .90). While the cooperativity factor is important in understanding the strength of polyvalent interactions, an analysis of the cooperativity factors of different lectins is beyond the scope of this project because other factors complicate this analysis (glycan spacing, lectin conformation, etc).

Overall, these studies have helped to identify specific types of glycan moieties present on N-glycan arms that are important in HA binding. The combination of the glycan array and tissue binding studies therefore provide a strong experimental system for exploring HA-glycan interactions, and would be even better complimented with additional analytical studies on the glycan structures themselves.

1.1.5 HA and O-linked Glycans

Although less understood than N-glycosylation, O-glycosylation has become equally important to our understanding of tissue-specific regulation and disease progression. Of recent interest is the

pathophysiological role of mucins, if any, in influenza infection. Mucins play a critical role in the integrity of the human tracheal and bronchial epithelia because the viscous gel layer they create protects against shear stress and chemical damage, and traps particulate matter and microorganisms that are eventually removed by the "sweeping" abilities of the ciliated epithelium [14-18]. In the human upper respiratory tract, mucins are secreted from goblet cells, which are cylindrical "goblet"-shaped cells found scattered in between ciliated epithelial cells (Figure 1.8). Their nuclei and cellular organelles all sit on the basal side of the epithelia, while their apical side extends into the lumen and releases membrane-bound secretory granules filled with mucus. The secretion of mucins occurs via two pathways, constitutive and stimulated secretion, the latter of which is typically enhanced by irritating stimuli like dust, smoke, or microorganisms [14-18]. Goblet cells also have an abundance of mucins on the surface of their membranes that are mediators of membrane binding events (adhesion/cell-cell communication). In addition to their muco-protective and signal transducing functions, these cells also have the ability to differentiate into other cell types, such as ciliated epithelial cells, giving them a "stem cell-like" quality [14-18].

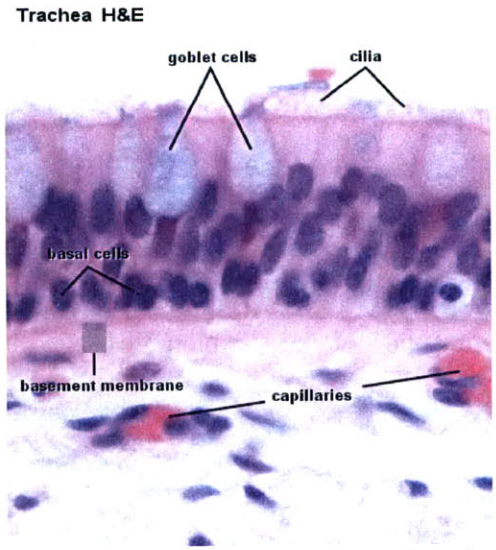


Figure 1.8. Hematoxylin and Eosin Stain of a Human Pseudostratified Columnar Epithelial Lining of the Trachea. Goblet cells filled with secretory granules are interspersed between ciliated epithelia cells, which sit above a basement membrane nestled by basal cells [42].

Since we still don't fully understand the events that take place or know the cell participants in the influenza infection process, the study of goblet cells and their mucins could shed light on these events, as well as help researchers elucidate some pathophysiological roles of mucins. Firstly, previous tissue binding experiments have shown that some humanized HA subtypes preferentially bind goblet cells over ciliated epithelial cells, suggesting that goblet cells may play a stronger role in viral infectivity than previously thought (Figure 1.9) [3]. These results have led to the hypothesis that goblet cell mucins are key mediators of viral binding, since they represent a biological "template" onto which many "ligands" (glycans) can be attached. This "template" ultimately allows for strong multivalent interactions, which can enhance viral infectivity.

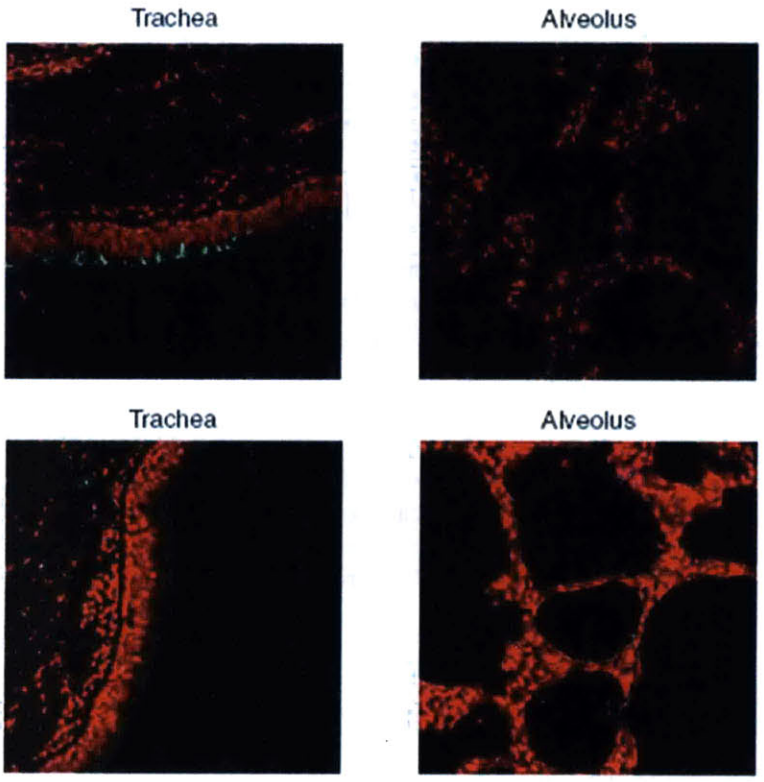


Figure 1.9. Costaining of Tracheal and Alveolar Tissue Sections with H1N1 A/South Caroline/1/18 HA (top) and H3N2 A/Moscow/10/99 HA (bottom). Both HAs stain green against the propidium iodide staining in red. The H1N1 ha seems to only stain goblet cells in the tracheal tissue, while the H3N2 HA seems to stain the entire apical side of the membrane. Neither HA heavily stains the alveolar tissues.

Lastly, very little is known about the protective functions of mucins in the context of influenza infection. It is possible that like many other diseased tissues, influenza-infected tissues overexpress specific kinds of mucin, either on the cell surface or in secretory granules, so that influenza binding to the host cell surface is impeded. Conversely, just like cancer cells, infected host cells might also aberrantly express certain mucins in order to impede protective immune responses. Confirmation of all of these potential roles of mucins requires an understanding of both the mucin glycan structures and the mucin-HA binding interactions. These two components allow researchers to relate structure to function, and could potentially highlight novel vaccine or therapeutic targets.

1.2 THESIS WORK

The study of the role of N- and O-linked glycans in influenza infection and pathogenesis is still young and presents exciting new areas of experimental exploration. This thesis describes design, optimization and application work that aims to further elucidate the roles of these glycans in a manner than acknowledges the significance of their structures or specific compositions. The following specific objectives were met:

1. The development, optimization, and application of a method for releasing, purifying, and analyzing O-linked glycan structures on glycoproteins. The structural information obtained from the above analysis will serve as the foundation on which conclusions can be made about the role of glycans in disease pathogenesis. More specifically, the knowledge of the structural motifs of specific mucins provided the motivation for the exploration of a direct binding assay that utilizes these mucins to obtain quantitative information on the O-glycan binding specificities of biologically significant proteins like lectins or HA.

2. The isolation and characterization of chicken erythrocyte, turkey erythrocyte, and human tracheal cell surface N-glycans as a means for scrutinizing the ability of the RBC agglutination assay to capture the HA-glycan interactions that are key determinants of virus adaptation.

The first objective was broken two major projects; one project focused on the design and optimization of an O-glycan analytical method, while the other focused on the use the information obtained in the former to design a direct binding assay that yields quantitative information on lectin or HA O-glycan binding specificities. The O-glycan analytical method consisted of three mini-projects (O-glycan MS analysis optimization, purification optimization, and release optimization). This subdivision was necessary in order to effectively explore all potential options for each process, and evaluate the processes that were effective in attaining the desired goal of maximal O-glycan detection. The completed analytical method was then utilized to study the types of terminal and core sialic acid linkages present on mucin O-linked glycans.

The direct binding assay exploration, on the other hand, sought to corroborate the information obtained from synthetic glycan arrays and tissue binding assays. Previous research has suggested that certain HA isotypes preferentially bind to the mucin-rich goblet cells in the trachea and bronchia instead of the N-glycan rich ciliated epithelial cells. The potential significance of this preference is completely unknown, but it is hypothesized that this binding preference is due to a high affinity or multivalent interaction between the O-glycan mucin coatings on the goblet cells and the HA isotype. In the context of the study of the interactions between O-linked glycans and HA, this assay would be more specific than the tissue binding assay because of the presence of characterized glycans, and also more relevant than the glycan array, since O-glycans are rarely presented in a biological setting as single isolated species. Therefore, the use of intact mucins with characterized glycan structures would enable researchers to capture the natural spacing of O-linked glycans and its potential effect on multivalent binding.

The second objective of this thesis broadly aimed to compare and contrast the glycan species present on the cell surfaces of chicken/turkey erythrocytes and human tracheal epithelial cells in order to draw conclusions about the ability of the agglutination assay to capture the HA-glycan interactions that are key determinants of virus adaptation. Since erythrocytes represent the only model system for influenza binding and inhibition studies, an analysis of the glycan moieties on the surfaces of these cells would highlight the relevance of these model systems in the context of human influenza infection. A previously established cell surface glycan extraction protocol was utilized to maximally isolate N-glycans from chicken red blood cells (cRBCs), turkey red blood cells (tRBCs), and human tracheal epithelial (HTE) cells. Glycan characterization consisted of MALDI-TOF MS analysis and sialic acid linkage analysis. These characterization methods were used to provide a preliminary profile of the relevant cell surface N-glycans, since they are thought to be integral in the process of influenza virus binding.

1.3 REFERENCES

1.) Varki, Ajit et.al. Essentials of Glycobiology. New York: Cold Spring Harbor Laboratory Press., 1999.

2.) Brett Garner, Anthony H. Merry, Louise Royle, David J. Harvey, Paμline M. Rudd, and Joëlle Thillet Structural Elucidation of the *N*- and *O*-Glycans of Human Apolipoprotein(a). ROLE OF *O*-GLYCANS IN CONFERRING PROTEASE RESISTANCE J. Biol. Chem. 276: 22200-22208.

3.) Chandrasekaran, A., A. Srinivasan, et al. (2008). "Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin." Nat Biotechnol **26**(1): 107-13.

4.) Kasson, P. M. and V. S. Pande (2008). "Structural basis for influence of viral glycans on ligand binding by influenza hemagglutinin." Biophys J **95**(7): L48-50.

5.) " Glycoprotein." wikipedia.com. 2000. Wikimedia Foundation.. 9 May 2009. <http://en.wikipedia.org/wiki/Glycoprotein>.

6.) Helenius, A. and M. Aebi (2004). "Roles of N-linked glycans in the endoplasmic reticulum." Annu Rev Biochem **73**: 1019-49.

7.) Hurtley, S. M. and A. Helenius (1989). "Protein oligomerization in the endoplasmic reticulum." Annu Rev Cell Biol **5**: 277-307.

8.) Klausner, R. D. and R. Sitia (1990). "Protein degradation in the endoplasmic reticulum." Cell **62**(4): 611-4.

9.) Hebert, D. N., J. X. Zhang, et al. (1997). "The number and location of glycans on influenza hemagglutinin determine folding and association with calnexin and calreticulin." J Cell Biol **139**(3): 613-23.

10.) " Mannose-6-phosphate" wikipedia.com. 2009. Wikimedia Foundation.. 9 May 2009. <http://en.wikipedia.org/wiki/Mannose-6-phosphate >.

11.) "Glyco Enzymes" www.functionalglycomics.org 2009 The Consortium for Functional Glycomics. 9 May 2009. <http://www.functionalglycomics.org/glycomics/molecule/jsp/glycoEnzyme/geMolecule.jsp>.

12.) Shimamura, M., Y. Inoue, et al. (1984). "Reductive cleavage of Xaa-proline peptide bonds by mild alkaline borohydride treatment employed to release O-glycosidically linked carbohydrate units of glycoproteins." Arch Biochem Biophys **232**(2): 699-706.

13.) Toribara, N. W., J. R. Gum, Jr., et al. (1991). "MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorphism." J Clin Invest **88**(3): 1005-13.

14.) Gendler, S. J. and A. P. Spicer (1995). "Epithelial mucin genes." Annu Rev Physiol **57**: 607-34.

15.) Shimizu, Y. and S. Shaw (1993). "Cell adhesion. Mucins in the mainstream." Nature **366**(6456): 630-1.

16.) Hanisch, F. G. and S. Muller (2000). "MUC1: the polymorphic appearance of a human mucin." Glycobiology **10**(5): 439-49.

17.) Amerongen, A. V., J. G. Bolscher, et al. (1995). "Salivary mucins: protective functions in relation to their diversity." Glycobiology **5**(8): 733-40.

18.) Rose, M. C. (1992). "Mucins: structure, function, and role in pulmonary diseases." Am J Physiol **263**(4 Pt 1): L413-29.

19.)"Flu" www.cdc.gov/flu 2009 The Centers for Disease Control. 9 May 2009. <http://www.cdc.gov/flu>.

20.) Tumpey, T. M., D. L. Suarez, et al. (2003). "Evaluation of a high-pathogenicity H5N1 avian influenza A virus isolated from duck meat." Avian Dis **47**(3 Suppl): 951-5.

21.) Tran, T. H., T. L. Nguyen, et al. (2004). "Avian influenza A (H5N1) in 10 patients in Vietnam." N Engl J Med **350**(12): 1179-88.

22.) Lee, C. W., D. L. Suarez, et al. (2005). "Characterization of highly pathogenic H5N1 avian influenza A viruses isolated from South Korea." J Virol **79**(6): 3692-702.

23.) Maines, T. R., X. H. Lu, et al. (2005). "Avian influenza (H5N1) viruses isolated from humans in Asia in 2004 exhibit increased virulence in mammals." J Virol **79**(18): 11788-800.

24.) de Jong, M. D., C. P. Simmons, et al. (2006). "Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia." Nat Med **12**(10): 1203-7.

25.) Maines, T. R., L. M. Chen, et al. (2006). "Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model." Proc Natl Acad Sci U S A **103**(32): 12121-6

26.) Yen, H. L., A. S. Lipatov, et al. (2007). "Inefficient transmission of H5N1 influenza viruses in a ferret contact model." J Virol **81**(13): 6890-8.

27.)"Influenza A Virus" 2009. www.wikipedia.com. Wikimedia Foundation. 9 May 2009. <http://en.wikipedia.org/wiki/Influenzavirus_A>

28.) Klausner, R. D. and R. Sitia (1990). "Protein degradation in the endoplasmic reticulum." Cell **62**(4): 611-4.

29.)Rogers, G. N., J. C. Paulson, et al. (1983). "Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity." Nature **304**(5921): 76-8.

30.)Connor, R. J., Y. Kawaoka, et al. (1994). "Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates." Virology **205**(1): 17-23.

31.)Matrosovich, M. N., T. Y. Matrosovich, et al. (2004). "Human and avian influenza viruses target different cell types in cultures of human airway epithelium." Proc Natl Acad Sci U S A **101**(13): 4620-4.

32.)Glaser, L., D. Zamarin, et al. (2006). "Sequence analysis and receptor specificity of the hemagglutinin of a recent influenza H2N2 virus isolated from chicken in North America." Glycoconj J **23**(1-2): 93-9.

33.)Russell, R. J., D. J. Stevens, et al. (2006). "Avian and human receptor binding by hemagglutinins of influenza A viruses." Glycoconj J **23**(1-2): 85-92.

34.)Kumari, K., S. Gulati, et al. (2007). "Receptor binding specificity of recent human H3N2 influenza viruses." Virol J **4**: 42.

35.)Matrosovich, M., T. Matrosovich, et al. (2007). "Avian-virus-like receptor specificity of the hemagglutinin impedes influenza virus replication in cultures of human airway epithelium." Virology **361**(2): 384-90.

36.)Tumpey, T. M., T. R. Maines, et al. (2007). "A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission." Science **315**(5812): 655-9.

37.)Cross, K. J., L. M. Burleigh, et al. (2001). "Mechanisms of cell entry by influenza virus." Expert Rev Mol Med **3**(21): 1-18.

38.)Basler, C. and P. Palese (2002). Influenza Viruses. Encyclopedia of Molecular Medicine. T. Creighton. New York, John Wiley and Sons**:** 1741-1747.

39.)Steinhauer, D. A. and J. J. Skehel (2002). "Genetics of influenza viruses." Annu Rev Genet **36**: 305-32.

40.) Spackman, Erica Avian influenza virus (2008) Springer.

41.) "Amplex Red Reagent"2009. www.invitrogen.com Invitrogen. 9 May 2009. <http://probes.invitrogen.com/media/pis/mp33851.pdf>

42.) "Human Trachea Epithelia Image". 2009. University of Washington. 9 May 2009. <http://www.lab.anhb.uwa.edu.au/mb140/CorePages/Epithelia/Images/trachea041he.jpg>

43.)Shinya, K., M. Ebina, et al. (2006). "Avian flu: influenza virus receptors in the human airway." Nature **440**(7083): 435-6.

Chapter 2

2 O-Glycan Analytical Method Development

O-glycan analysis can be divided into three distinct processes, all of which can be optimized independently and can have a tremendous impact on the final information obtained about the O-glycan spectra of a protein. The first process typically involves the chemical or enzymatic treatment of the protein, with the end goal being either the liberation of the O-glycan or fractionation of the glycoprotein into peptides and glycopeptides. These processes reduce the molecular weight of the analyzed species, such that they can be more accurately identified by mass spectrometry instruments, which, in general, tend to yield lower resolutions for species with high molecular weight (> 20,000 Da) [1]. This process is then followed by glycan purification and MS analysis, from which the full O-glycan spectra of a glycoprotein can be obtained.

However, just as car parts have to be correctly assembled to create a fully functional car, the successful analysis of O-glycans requires that the optimization of each process be designed with the other processes in mind. For example, if the chemical or enzymatic treatment of the glycoprotein requires large amounts of an agent that is difficult to remove in the purification process, then the final analysis of the glycan or glycopeptide can be significantly impaired. It is essential to understand each of the three processes individually in order to both grasp how the success of one process depends on another, and to understand how to effectively address any complications that arise during the optimization process.

Optimization of O-glycan retrieval, purification, and analysis was approached in a way that is the reverse of the actual experimental procedures. The rationale behind this methodology was that O-glycan

release and purification procedures are best optimized after the development of a sound mass spectrometric method. The analysis of O-glycans on the Bruker and Voyager MALDI-TOF machines were carried out first with a set of three commercial O-glycan standards in order to obtain a clear understanding of the sensitivity of the machine to both sample amount and quality. Once these methods were standardized, two purification procedures were applied to the O-glycan standards in order to determine the procedure that resulted in the highest O-glycan intensities, minimal sample losses, and the fewest adduct peaks. Lastly, one method of O-glycan release was tested under various experimental conditions to determine the conditions that resulted in the best O-glycan signal for a given amount of starting material. The optimized analytical method was then used to perform a sialic acid linkage analysis on bovine fetuin and mucin glycans to obtain additional structural information about these glycans.

2.1 O-GLYCAN MS ANALYSIS

2.1.1 Background

2.1.1.1 MALDI-TOF MS

One of the most popular methods of analyzing glycans is Matrix Assisted Laser Desorption/Ionization (MALDI) mass spectrometry [2]. MALDI is versatile because glycans can be detected in their native states (no derivatization) without destruction, and can provide information about their structure and branching [2]. Furthermore, many samples can be analyzed in one use of the machine since the average run time for each sample is less than a minute. In MALDI analysis, the sample is embedded in a chemical matrix whose role is thought to be to absorb the UV laser light and transfer the energy to the analyte. This absorption causes the substrate under analysis to vaporize and ultimately "fly" towards a detection plate that calculates its mass. From these masses, we are able to determine a mass and hence elicit the possible monosaccharide components of each glycan. The ions generated in this machine are then transported to the mass analyzer, which separates the ions according to their mass to charge ratio, via a magnetic (sector analyzer) or electric field (time-of-flight (TOF) analyzer). More information about the glycan (linkage,

etc) can be determined if the sample is fragmented by collision induced dissociation (CID) or laser induced dissociation (LID).

2.1.1.2 Important MALDI-TOF MS Parameters

Important parameters to consider when optimizing MALDI-MS analysis of ionized species are the mass resolution, mass accuracy, and signal to noise ratio. Mass resolution describes the degree to which a mass spectrometer can differentiate two masses. It is typically defined according to the type of analyzer used by the mass spectrometer; those that use a magnetic sector to separate ions typically define the mass resolution as the quotient of the mass number of the observed mass and the peak width ($R = m/\Delta m$) [3]. For these machines, the peak width is taken to be that at 10% of the peak intensity; in other words, two peaks are considered to be separated if there is a valley at an intensity less than 10% of the intensity of the higher of the two peaks [3]. Usually, this specification is acceptable for the differentiation of isotopes of the same species [3]. Mass spectrometers that use Time-of-Flight analyzers, however, use the full width at half maximum intensity to calculate the resolution (fwhm) because of peak broadening near the baseline that is typical for this type of analyzer [3]. This means that for a MALDI-TOF instrument, a mass resolution of 2000 will allow one to differentiate between species with m/z 750.0 and 750.375 or 2500.00 and 2501.25. Furthermore, a resolution of 3,000 will differentiate 750.00 from 750.25 and 2500.00 from 2500.83.

In order to fully understand the importance of mass resolution, it is necessary to understand mass accuracy, or the degree to which the mass spectrometer can correctly identify the mass of a particular ion [4]. Mass accuracy depends on the distribution of the ion species. The presence of isotopes at their natural abundances can confound MS spectra because of additional peaks that are close in mass to the peak of interest. Depending on the resolution of the instrument, species with different isotopic distributions can either appear as a cluster of monoisotopic peaks (high resolution), or a single broad peak (low resolution). For a single broad peak, the centroid mass is usually estimated and deemed the "average" mass. This calculation can be fairly tricky for isotopic distributions that are not symmetrical, since the centroid mass

will not necessarily arise from the most abundant peak (Figure 2.1). For these reasons, higher resolutions suggest higher mass accuracies since there is less of an error associated with estimating monoisotopic masses than average masses. Furthermore, if there is no trade-off between mass resolution and mass sensitivity (the sensitivity of the spectrometer to species of different mass ranges), then the mass accuracy of large molecular weight species should be the same as those of low molecular weight species [4]. However, for most spectrometers there is a trade-off between sensitivity and resolution, such that species with larger molecular weights and isotopic distributions are only observed in an MS spectrum at reduced resolutions, and thus, reduced mass accuracies [4]. On the other hand, smaller ions (carbohydrates, peptides, etc), have smaller isotopic distributions, such that the first monoisotopic mass can be taken as the predominant mass of the species, and the average mass (i.e. at low mass resolution) is within 1 amu of the predominant monoisotopic mass of the species.

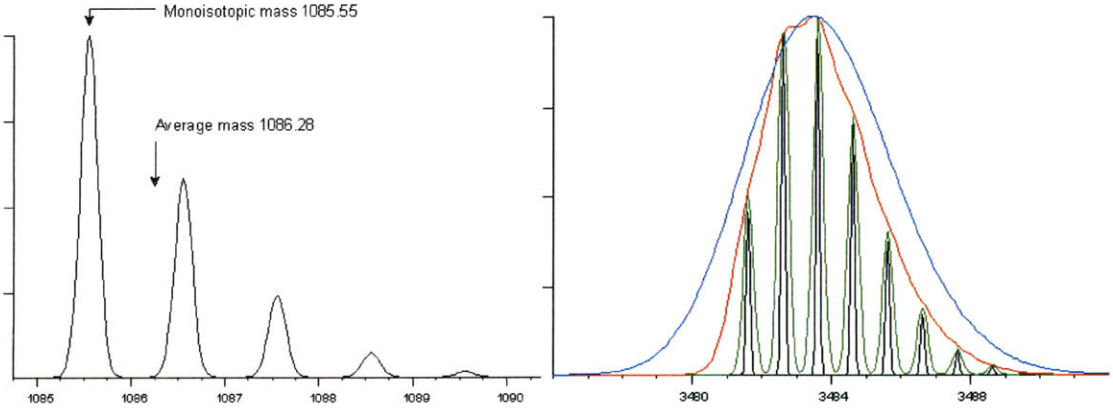


Figure 2.1. MALDI-TOF Mass Spectra for Resolved and Unresolved Protein Species. The left image shows the difference between average and monoisotopic mass calculations. For these well-resolved peaks, the monoisotopic mass is properly identified, but the average mass is skewed because of the asymmetrical shape of the isotopic distribution. This is true for smaller species (< m/z 3000). For larger molecules like glucagon, the average mass is closer to the most prominent monoisotopic peak (right panel), and so average mass calculations tend to be more acceptable. The different colors indicate resolutions of 1000 (blue), 3000 (red), 10,000 (green) and 30,000 (black) [4].

In addition to the mass resolution and mass accuracy, the signal to noise ratio, is also important in assessing the quality of an MALDI-MS spectra. This ratio is simply the ratio of the signal intensity to the background noise, and is primarily used to differentiate between potential ion peaks and peaks generated from background noise. More specifically, one of the key parameters in the identification of an ion peak with respect to baseline intensity is a signal to noise ratio of at least 3:1 [5].

For O-glycan analysis, the mass resolution, mass accuracy, and signal to noise ratios greatly influence the decision to analyze the sample in the linear or reflective modes. On TOF instruments in the linear-mode, ions with identical masses can fly with slightly different kinetic energies so that they arrive at the detector at slightly different times [6]. As a result, the peak width broadens, and mass resolution (and accuracy) decrease. In the reflective mode, however, an ion mirror called a "reflectron" focuses the energy of ions with the same molecular weight by causing the ion beam to reflect back down the flight tube toward the detector, thereby improving mass resolution and accuracy [6].

In this study, both linear and reflective modes are used, and mass resolutions of at least 500 were considered to be excellent, such that species with masses of at least 500.00 Da could be resolved down to a difference of at least 1.5 Da. This difference allows for the identification of potential glycan structures (via databases such as GlycoMod, the Consortium for Functional Glycomics), all of which must be further corroborated by additional MS/MS studies. Signal to noise ratios above 20 were also considered to be acceptable for the identification of a glycan peak. Furthermore, if minimal resolutions were not obtained, most likely due to high laser intensity, lack of species ionization in the specific matrix, small species concentrations in the sample, or impurities that suppress ionization, peaks were marked as "possible peaks" if their signal to noise ratios were above 20.

2.1.1.3 MALDI-TOF MS Matrices

The most popular MALDI matrix for carbohydrate analysis is 2,5-dihydroxybenzoic acid (DHB) [2]. The matrix is usually dissolved in some combination of ethanol, acetonitrile, or water, and will crystallize to form spiky matrix "needles" that radiate into the center of the spot [2]. Sugars and other contaminants (i.e. reducing agents, etc) are found as a heterogeneous mixture in the center of the spot, while peptides and glycoproteins are usually found on the periphery of the spot [2]. Furthermore, the high detection limits offered by this matrix (500 fmol for glycans) make it a very popular choice for carbohydrate analysis [2].

Even though DHB is most often used in the positive mode and typically produces $M + Na^+$ species as the major molecular ion, it can also produce $(M-H)^-$ ions if analyzed in the negative mode. However, analysis of acidic carbohydrates using DHB tends to be more difficult because the negatively charged residues form salts with alkali metals, resulting in multiple adduct peaks. Furthermore, fragmentation of sialylated, acetylated, or sulfated glycans, as a result of loss of a carboxylic acid, acetyl, or sulfate group, respectively, can further complicate the analysis of these sugars. The addition of co-matrices such as spermine can improve spectra by reducing salt formation and improving sample crystallization [7].

Because of the aforementioned difficulties with acidic glycan analysis, matrices like 6-Aza-2-thiothymine (ATT) and 2,4,6-trihydroxyacetophenone (THAP) are preferred over DHB [2]. These matrices provide significant increases in the sensitivity of acidic glycans. ATT provides the same detection limits for acidic glycans as other commonly used matrices do for neutral glycans (~50fmol), and also provides higher signal to noise ratios [2]. THAP, on the other hand, prevents salt formation and yields a single peak for sialylated N-glycans at ~ 10fmol with no evident fragmentation [8]. Lastly, acidic glycan analysis is even further complicated by the fact that it is harder to form a negatively analyte because of the presence of sodium and hydrogen, both of which are naturally present in common matrices [9]. However, with the appropriate matrix formulation and rigorous MALDI parameter optimization, excellent spectra can be obtained for acidic glycans.

2.1.2 Materials and Methods

Glycans were analyzed either on Bruker's Omniflex or on Applied Biosystem's Voyager MALDI-TOF machines with Delayed Extraction. Both machines use pulsed nitrogen lasers (337 nm). O-glycan analysis was optimized using a combination of previously published methods that were successful in obtaining good MS spectra for O-linked glycans. Four matrices were chosen as preliminary matrices. These matrices, which include DHB in Ammonium Citrate, DHB and Spermine in Acetonitrile, 10mg/mL ATT in ETOH (spotted on a perfluorinated Nafion spot), and 30mg/ml DHB in DIW (recrystallized in deionized water (DIW)) were freshly prepared and combined with O-glycan standards in ratios of 1:1, 1:2, 1:5, and 1:10. Furthermore, a number of drying procedures (vacuum, etc) were tested to determine the spotting method that resulted in the best matrix crystallization.

Since a method for O-glycan analysis in the negative and reflective modes had not yet been developed on the Bruker MALDI-TOF, all of these matrices were initially analyzed using previously established parameters for the analysis of negative ions with an $m/z > 5100$. This analytical method was then further optimized by varying the method parameters during data acquisition. The parameter optimization aimed to achieve the following goals: minimal peak splitting, maximal peak intensities, high mass resolution, and acceptable signal to noise ratios. The goal of obtaining a high mass resolution was expected mostly in the reflective mode because isotopes for species with molecular weights smaller than 3,000Da can be resolved. In terms of the order of parameter optimization, the second ion source voltage was optimized first, followed by the lens voltage, delay, electron gain and laser intensity. Laser intensity was usually adjusted as the data was acquired so as to obtain maximal glycan intensities without disrupting the spectra baseline.

Unlike that of the Bruker MALDI-TOF, O-glycan analysis on the Voyager MALDI-TOF was conducted by using a previously established matrix formulation and method for acidic glycans. Usually, 1µl of sample was combined with 9µl of 10mg/ml ATT in ethanol and spotted on a previously applied perfluorinated Nafion spot. The spot was allowed to dry in a humidity controlled chamber (23% H_2O

content) prior to analysis. The laser intensity was usually optimized on a sample to sample basis. Parameters for this method are listed in Table 2.1.

Table 2.1 Voyager MALDI-TOF MS Parameters Used in the Analysis of Acidic Glycans

Parameter	Acidic Glycans
Mode (+ or -)	Negative
Mode (Ref or Lin)	Linear
Accelerating Voltage	22,000 V
Grid Voltage	93 %
Guide Wire	0.3 %
Delay	150 nsec

2.1.3 Results and Conclusions

2.1.3.1 Bruker Optimization

Of the preliminary matrices selected for Bruker-MS O-glycan analysis, 30mg/mL DHB in DIW (re-crystallized in ETOH) showed the most promise. All three O-glycan standards were present in this spectrum (Figure 2.2A). This preliminary spectra was then improved by parameter optimization to yield the glycan spectra in Figure 2.2 C and the optimized method parameters (for glycans within the mass range of 600-1400) listed in Table 2.2.

Table 2.2 Bruker MALDI-TOF MS Parameters Used in the Analysis of Acidic Glycans

Ion Source 1	Ion Source 2	Lens	Reflector	Delay
19.00 kV	13.55 kV	10.00 kV	20.00 kV	200 ns

It is important to note that for this MALDI-TOF machine and method, it was difficult to obtain any peaks for acidic glycans above a mass of 1400 Da, most likely because of the method parameters, which were optimized for m/z 632 – m/z 997. Further attempts at parameter optimization for acidic glycans above an m/z of 1400 using acidic N-glycan standards were not successful using the optimized DHB matrix formulation in the negative mode. It is possible that this particular matrix has a lower ionizing

capacity for acidic ions above this mass. Therefore, the matrix method and optimized parameters discussed for the Bruker MALDI-TOF MS will be used for glycans within the mass range of 500 and 1400 Da.

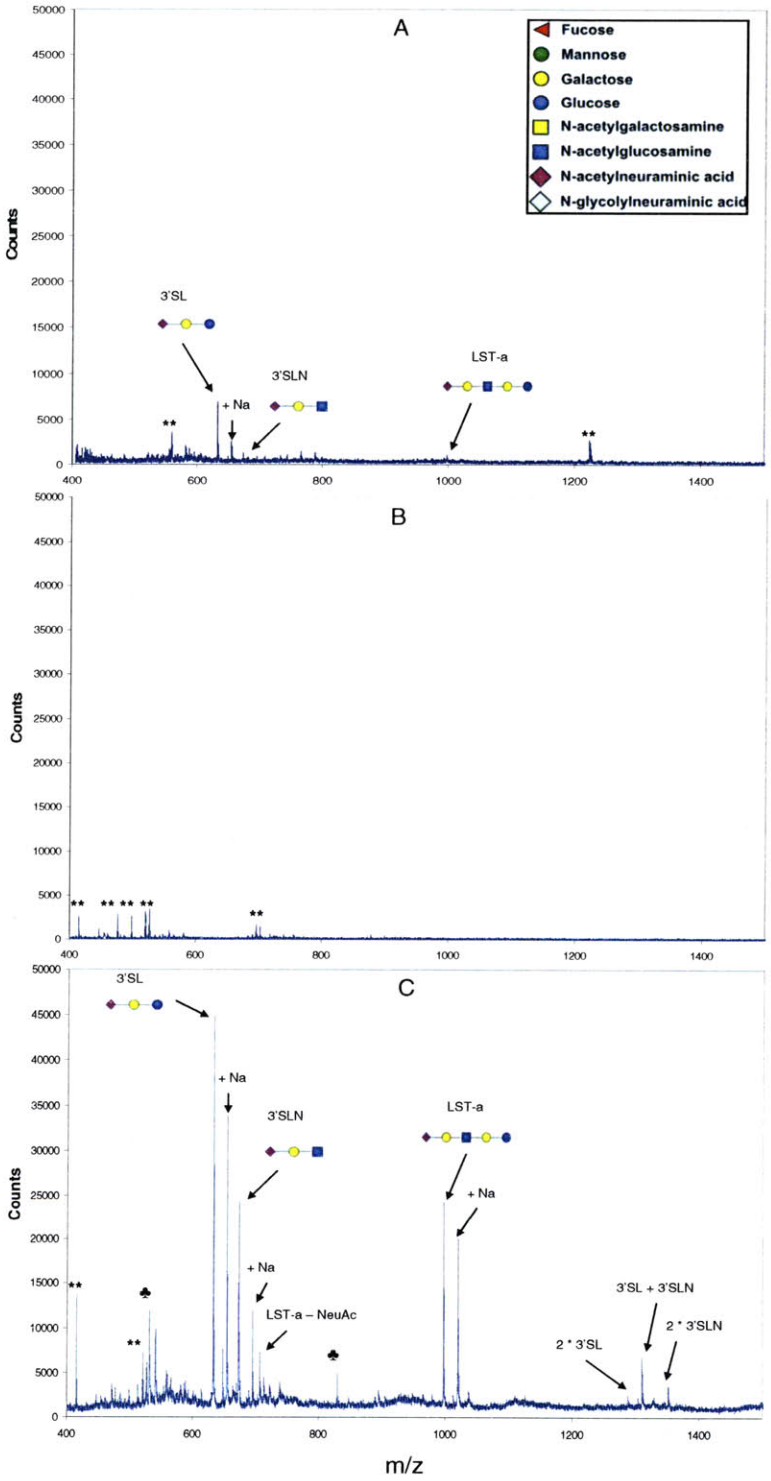
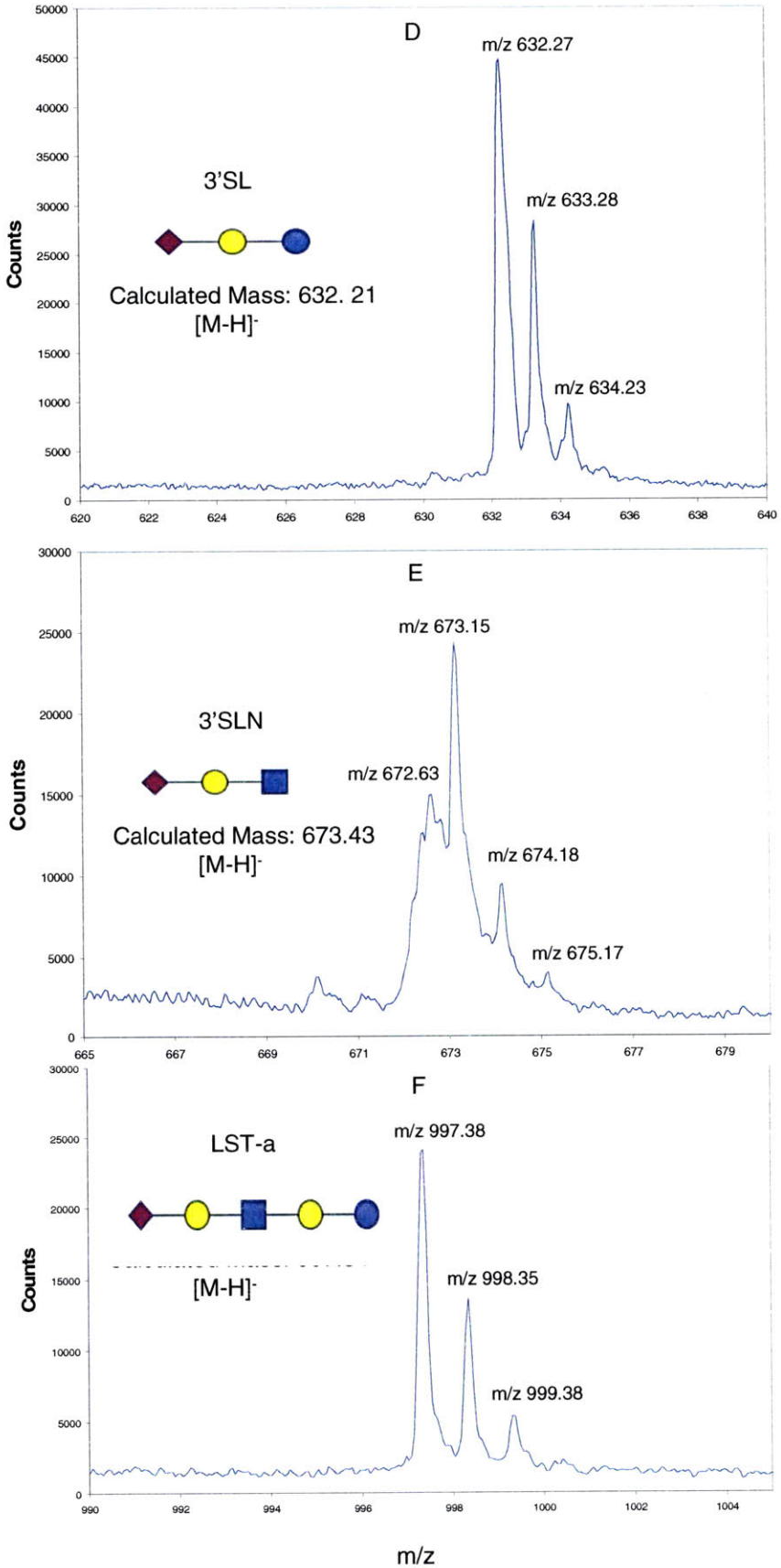


Figure 2.2. O-glycan Standard MS Method Optimization on the Bruker MALDI-TOF. Spectra were acquired in the negative and reflective modes. A. The preliminary O-glycan spectra acquired using saturated DHB in DIW, re-crystallized in ETOH. B. Matrix background for the optimized method and matrix. C. Optimized O-glycan spectra utilizing 30mg/mL DHB in 50% ACN, recrystallized in ETOH. D. Close-up 800pmol 3'SL. E. Close-up 200pmol 3'SLN. F. Close-up 200pmol LST-a. ♣ = peaks arising from the 3'SL sample. ** = matrix peaks.


D
m/z 632.27
m/z 633.28
m/z 634.23
3'SL
Calculated Mass: 632. 21
[M-H]-
Counts
E
m/z 673.15
m/z 672.63
m/z 674.18
m/z 675.17
3'SLN
Calculated Mass: 673.43
[M-H]-
Counts
F
m/z 997.38
m/z 998.35
m/z 999.38
LST-a
[M-H]-
Counts
m/z

The optimized Bruker method yielded high O-glycan standard intensities while also providing high peak resolutions and signal to noise ratios (>1000 and >500, respectively) (Table 2.3). Furthermore, glycans were best viewed in the reflective mode because their isotopes were clearly resolved, allowing for the differentiation between unique glycan peaks and isotopes of one particular glycan. Peaks resulting from a loss of sialic acid were not dominant in the spectra. The matrix formulation and sample to matrix ratios were finalized as 30mg/mL DHB in ACN/DIW (pH 10) (v:v, 1:1).

Table 2.3. Mass Resolution and Signal to Noise Ratio for O-glycan Standards Analyzed on the Bruker MALDI-TOF

O-Glycan Standard	Resolution†	Signal/Noise†
3'SL	1504.7	9476.6
3'SLN	1749.0	513.3
LST-a	3925.6	509.2

†Value given is that of the highest monoisotopic peak for each glycan. The signal to noise ratio is calculated as the ratio between the peak intensity and the background noise, as calculated by the machine.

The resolutions of each O-glycan standard ranged from 1500 to 3925, and at least three monoisotopic peaks for each glycan, representing the naturally occurring isotopic distribution of carbon, are present (Figure 2.2D-F). While the monoisotopic peaks of 3'SL and LST-a are clearly distinguishable and highly resolved, that of 3'SLN is not. This glycan seems to have 4 partially resolved monoisotopic peaks, all of which are contained within a single broader peak. It is possible that the peak shape and characteristic of 3'SLN arise from the quality of the particular commercial O-glycan standard.

Minimal background noise and excellent signal to noise ratios were also observed in samples analyzed with the optimized Bruker MS method (Figure 2.2B). Since this project aims to analyze all glycans above an m/z of 500, it was necessary to ensure that matrix noise would not suppress glycans in the lower limit of this mass range. The only matrix-associated peaks above an m/z of 500 are those at m/z 527 and 697, and only the former is apparent in spectra in which O-glycans are present. Thus, it should be

fairly easy to separate matrix and glycan-derived peaks above an m/z of 500 using this method. Furthermore, signal to noise ratios above 500 were obtained for each O-glycan standard species.

In terms of glycan intensities, all acidic standards were clearly observed at intensities above 25,000. More importantly, the peak intensities of 3'SLN and LST-a are equal and smaller than that of 3'SL, which is present at 4 times the concentration, showing that accurate relative quantification of glycans can be achieved using this methods for glycans within the range of m/z 500 to 1400. It is important to note that all peaks above an m/z of 1500 result from glycans flying as dimers and can be minimized by decreasing the laser intensity. Furthermore, all peaks clustered around m/z 531 and the single peak at m/z 830 come from species in the 3'SL sample. These peaks are therefore associated with the quality of the sample and not the MS method.

In terms of mass accuracy, the most prominent monoisotopic peaks for each glycan are within .05% of the theoretical mass after calibration (calculated with average values of each atomic species) (Figure 2.2 D-F). This degree of mass accuracy is sufficient for the purposes of this project because MALDI-TOF analysis will be used to propose potential glycan structures, which will need to be further validated with additional MS/MS and exoglycosidase studies.

2.1.3.2 Voyager Optimization

Optimization of O-glycan standards on the Voyager MALDI-TOF MS was investigated to a lesser extent, since a matrix and method formulation had already been well established in the lab for acidic glycan standards. Figure 2.3 shows the optimized spectra for the three O-glycan standards, as well as close-ups of their individual peaks. This method had been well standardized for acidic glycans above m/z 1000 (data not shown), and this optimized spectra shows that it can also effectively detect glycan species within the m/z range of 500-1000. For this range, there is only one prominent matrix associated peak (at m/z 568), such that all other peaks can be safely deemed glycan-associated peaks. The parameters utilized for this method are shown in Table 2.1.

In contrast to the spectra shown in Figure 2.2, Figure 2.3 shows O-glycan mass assignments that have been averaged because of the inability to effectively resolve monoisotopic peaks for each glycan in the linear mode. In this case, the linear mode was chosen because of the excessive loss of sialic acid seen in the reflective mode when the same matrix was used. Therefore, each glycan is represented by a single broad peak, and the mass assignment is made based on the centroid mass for that peak. As expected, the resolutions for each individual peak, shown in Table 2.4, are low (less than 500) because of the resolving capacity of the linear mode. This means that it may be difficult to distinguish between two species that have masses separated by 3 Da. However, the signal to noise ratios for each peak are well above 500, indicating a definite presence of an ionic species. Therefore, peaks can be justifiably identified, and scrutinized just on the basis of signal to noise ratio. Further structural elucidation of the proposed glycan structures for each peak may then be determined with additional MS/MS experiments.

Table 2.4 Mass Resolution and Signal to Noise Ratio for O-glycan Standards Analyzed on the Voyager MALDI-TOF

O-Glycan Standard	Resolution†	Signal/Noise†
3'SL	167	1813.6
3'SLN	266	748.8
LST-a	190	1837.9

†Value given is based on the average mass for each glycan. The signal to noise ratio is calculated as the ratio between the peak intensity and the background noise, as calculated by the machine.

The intensity of 3'SL is on the order of that of LST-a, which is present at 4 times lower of a concentration. These inconsistencies result from an optimization that prioritized higher molecular weight species, as recommended by the Voyager-DE STR user manual. In the linear mode especially, it was suggested that for a particular m/z range, the higher molecular weight species takes "optimization priority" over the other species (with regard to resolution) because smaller species are naturally easier for

the machine to resolve. This means that, for the m/z range of 500 to 1000, accurate relative quantitative analysis of glycans may not be possible. However, if relative quantification is necessary, analysis of this range on the Bruker MALDI-TOF machine can be performed because that method works very well for glycans in the m/z range 500 to 1400. For acidic glycans above an m/z of 1000, however, accurate relative quantitative analysis can be achieved because accurate relative ratios of the larger acidic N-glycan standards were achieved (data not shown). In terms of mass accuracy, the average mass calculated for each glycan is within 0.05% of the theoretical mass after calibration (calculated with average values of each atomic species, Figures 2.3 B-D). This degree of mass accuracy is sufficient for the purposes of this project.

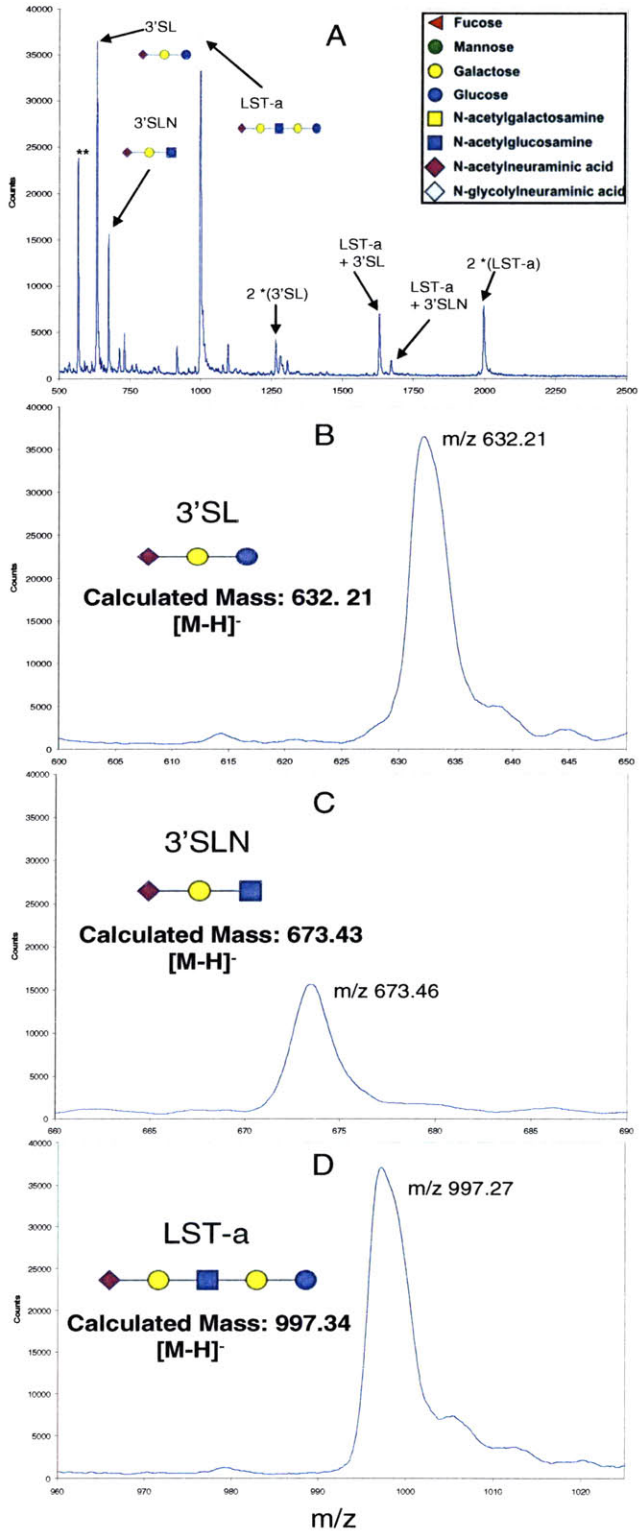


Figure 2.3 O-glycan Standard MS Spectra on the Voyager MALDI-TOF. A. Full O-glycan standard spectra. B. Close-up of 800 pmol 3'SL C. Close-up of 200 pmol 3'SLN. D. Close-up of 200 pmol LST-a. Spectra were obtained in the negative and linear modes using 10mg/mL ATT in ETOH and the parameters in Table 2.1. ** = matrix peaks.

2.1.3.3 Detection Limits for Voyager and Bruker Machines

Serial dilutions of the O-glycan standards were analyzed on both MALDI-TOF instruments in order to determine the detection limit of each machine (Figure 2.4). Glycans were analyzed in ratios ranging from nanomolar to femptomolar amounts (Table 2.5).

Table 2.5 Molar Amounts of O-glycan Standards per Spot for the Detection Limit Assay

Dilution #	**Bruker Ratios (pmoles 3'SLN: pmoles LST-a: pmoles 3'SL)**[†]	**Voyager Ratios (pmoles 3'SLN: pmoles LST-a: pmoles 3'SL)** [†]
1	100:100:400	20: 20: 80
2	20:20:80	4: 4: 16
3	4:4:16	0.8: 0.8: 3.2
4	n/a	0.162: 0.162: 0.648

[†]Molar amounts listed reflect the actual amount of glycan on the spot

Four glycan mixtures were prepared using 1:5 serial dilutions. The stock solution contained a ratio of 200pmol 3'SLN: 200 pmol LST-a: 800pmol 3'SL, while the other three mixtures contained 1:5 serial dilutions of this stock solution. However, because the sample to matrix ratios were different for Voyager and Bruker analysis (9:1 for ATT matrix, but 1:1 for DHB matrix, respectively), the amount of glycan that actually ended up on each spot varied for the same stock solution. Regardless, Figure 2.4 shows that the Bruker method has a much higher glycan detection limit; the detection limit for the Bruker machine was only about 20 pmoles, while that for the Voyager machine was ~ 0.8 pmoles, although two of the three glycans were still detectable at a ratio of 0.162 pmoles 3'SLN: 0.162 pmoles LST-a 0.648 pmoles 3'SL (Figure 2.4 G).

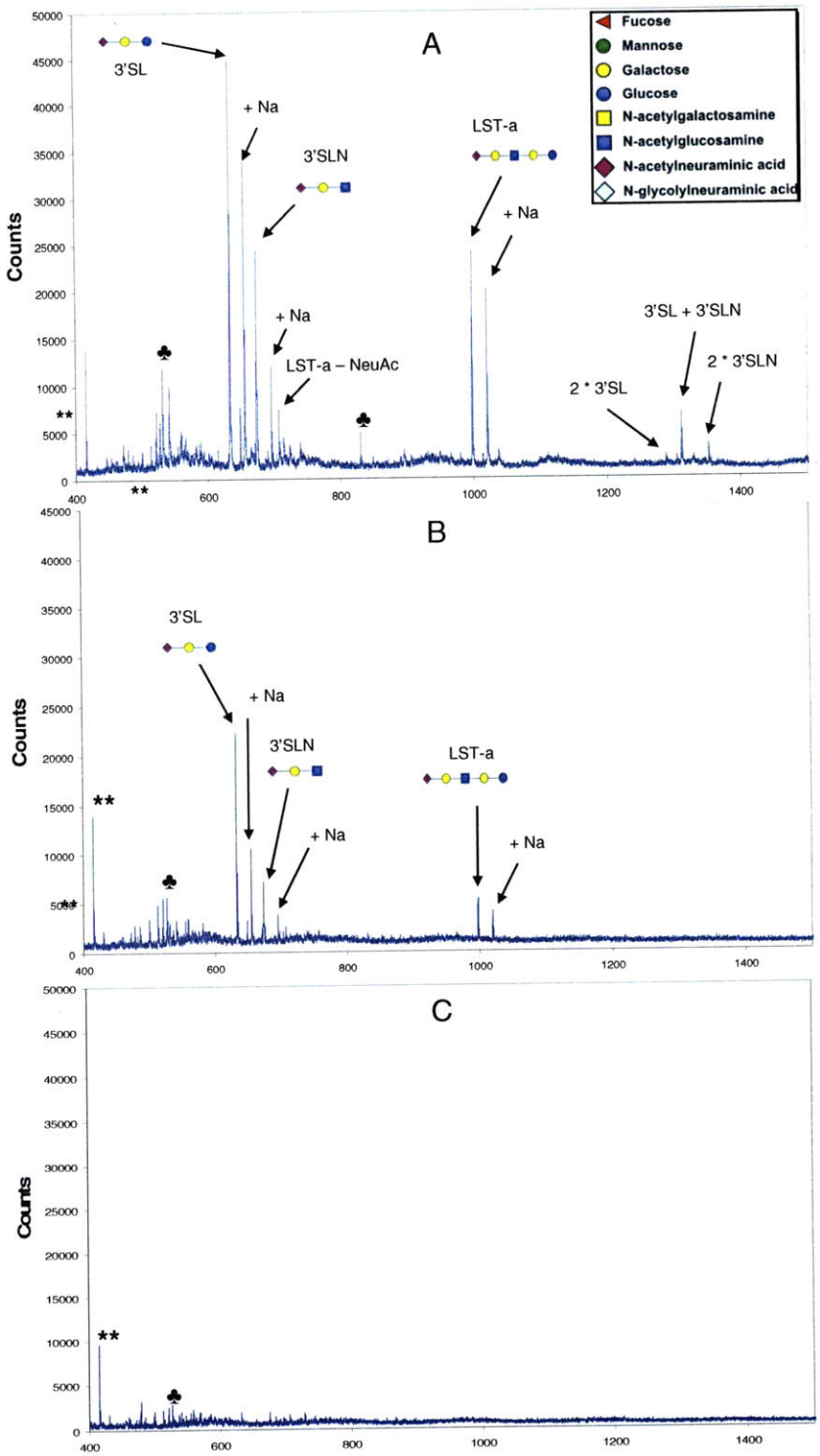
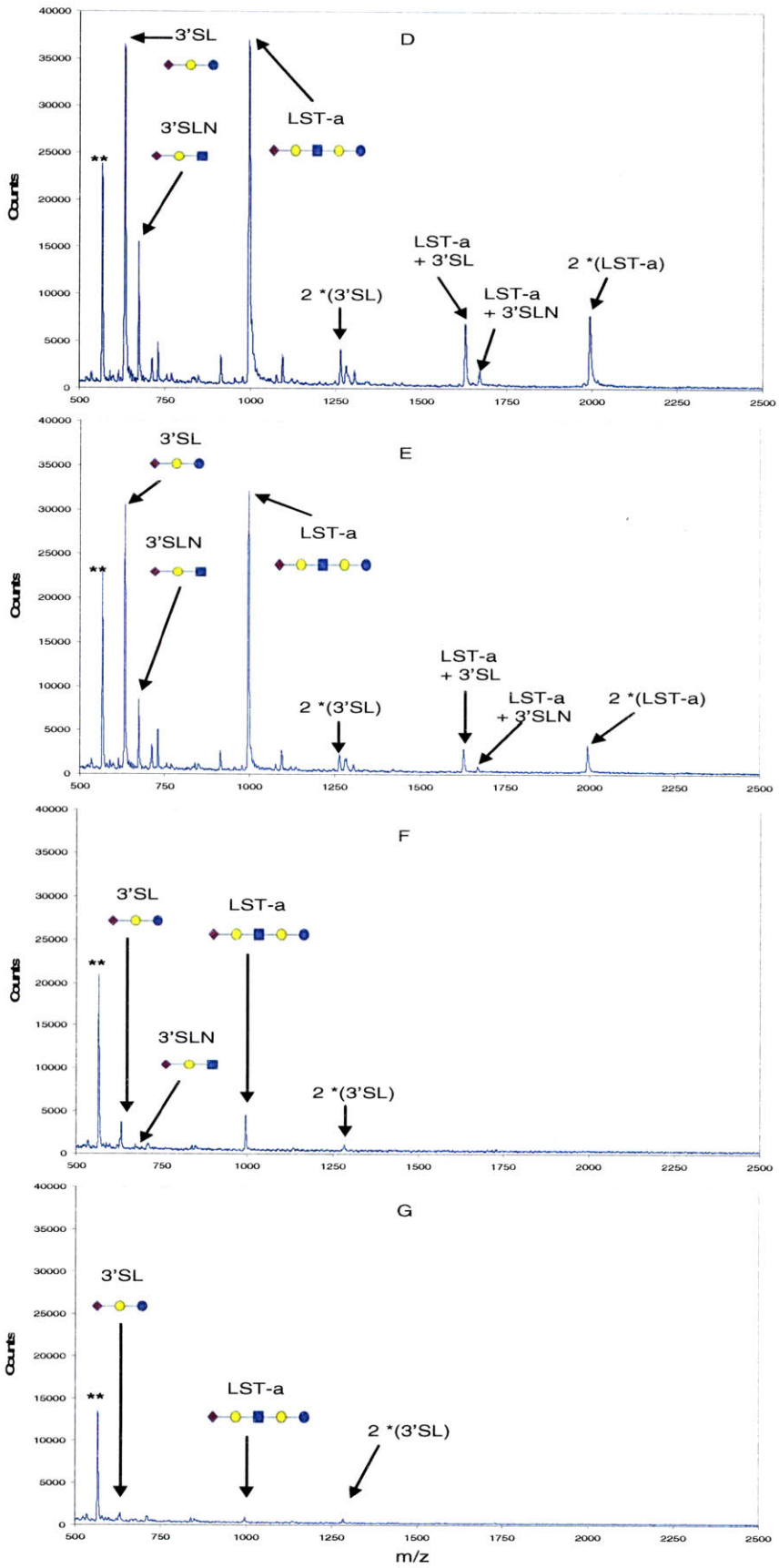


Figure 2.4. O-glycan Standard Molar Detection Limit on the Bruker and Voyager MALDI-TOFs. Spectra were taken with the following pmole ratios of 3'SLN:LST-a:3'SL on spot, for Bruker A. 100:100:400 B.20:20:80 C. 4:4:16, and for Voyager D. 20:20:80 E. 4:4:16 F. 0.8: 0.8: 3.2 G. 0.162: 0.162: 0.648 Spectra in A-C were acquired using the optimized matrix and method parameters listed in Figure 2.2, while D-G were acquired that of Figure 2.3. ** = matrix peaks, ♣= 3'SL associated peaks.


D
3'SL
3'SLN
LST-a
**
LST-a
+ 3'SL
LST-a
+ 3'SLN
2 *(3'SL)
2 *(LST-a)
E
3'SL
3'SLN
LST-a
**
LST-a
+ 3'SL
LST-a
+ 3'SLN
2 *(3'SL)
2 *(LST-a)
F
3'SL
LST-a
**
3'SLN
2 *(3'SL)
G
3'SL
LST-a
**
2 *(3'SL)
Counts
m/z

2.1.3.4 Conclusion

Two analytical methods were optimized for the analysis of acidic O-glycan standards. The method optimized on the Bruker MALDI-TOF utilizes the reflective mode, and is effective at resolving monoisotopic peaks for glycans in the range of m/z 500 -1400, while that optimized on the Voyager MALDI-TOF utilizes the linear mode, and is effective at detecting all glycans above an m/z of 500.

Each method, however, also has its advantages and disadvantages. The Bruker method has superior resolving power, but can only be used for the analysis of glycans below an m/z of 1400. This range covers some of the smaller O-glycans, but inevitably misses the longer O-glycan species which may play biologically significant roles in cell surface binding. For example, O-glycans on mucins secreted in breast cancer tumors are reported to contain many lactosamine units (GalNAc-Gal units); thus, an O-glycan with 3 lactosamine units might not be detected using this method. Furthermore, because DHB can create both positive and negative ions, neutral glycans are just as susceptible to ionization as acidic ions in the negative mode. This could further complicate analysis of biological samples, which tend to be "messy" and contain both neutral and negatively charged species. Lastly, the detection limit for the Bruker machine is on the order of 20 pmoles, which is relatively high considering that some glycans are present in samples in femptomolar amounts. Still, if the goal of the analysis is to identify the most predominant glycans in a sample, this method will effectively work for glycans within the m/z range of 500 – 1400.

The Voyager method, however, even with its lower resolving power, can effectively detect all acidic glycans above an m/z of 500 with minimal matrix peaks and background noise. The advantage of this method is that mainly acidic glycans are detected (femptomolar amounts). The major disadvantage to this method is that the lower resolving power can make it difficult to resolve unique glycan species with masses differing by less than 4 Da. However, depending on the goal of the MS analysis, these shortcomings may not matter much, especially if further MS studies are done on the species. For most of this project, the Voyager method is used because of its ability to detect all glycans above an m/z of 500 and its selectivity to acidic glycans. The Bruker method was mainly used to confirm or elaborate on

glycan species within the mass range of 500 – 1400 Da, or to assess the efficiency of salt removal in the purification process (since the Voyager method doesn't allow for the formation of sodium adduct peaks).

2.2 O-GLYCAN PURIFICATION

2.2.1 Background

Glycan purification is a crucial step in the glycan analytical process because of the ability of sample contaminants to hinder glycan ionization on MS machines. The N-glycan purification procedure previously established in the lab uses a normal phase bench-top column (C_{18} in nature) to separate glycans from peptides. This process is then followed by a second purification that is reversed phase in nature and separates the glycans from salts. These two fairly standardized procedures can be done at bench-top or HPLC levels, and are crucial to the optimum detection of N-glycans. O-glycan purification procedures, on the other hand, are far less standardized, particularly because of the properties of the oligosaccharides themselves. O-glycans tend to be smaller and less hydrophobic in character than N-glycans, and are thus not reliably captured on bench- top reversed phase columns. Furthermore, sample losses can be a problem in bench-top purification procedures, which typically require the elution of glycans in volumes larger than 2mL. For all of these reasons, O-glycan purification strategies have remained somewhat un-standardized and diverse. Currently, fast and high performance liquid chromatography (FPLC and HPLC, respectively), lectin affinity chromatography, resin, and size-exclusion chromatography based strategies are employed.

O-glycan purification strategies either rely on the differences in hydrophobicity between the protein and the glycans, or the ability of the glycan to bind glycan-binding proteins called lectins. For example, Sephadex G50 columns, which contain cross-linked dextran resins commonly used in gel-filtration columns, purify O-glycans via size-exclusion chromatography. Typically, the sample is applied to a pre-equilibrated column, flows down through the column, and is washed with buffer. O-glycans can be retained in the column according to their degree of hydrophobicity, but some flow through and exit the

column without being retained. Captured glycans are typically eluted using a more hydrophobic solvent. Fractions of the eluent are collected, and the glycan-containing fraction can be determined using UV absorbance at 206nm if the glycans are present at a high enough concentration [10]. Even though this process efficiently separates glycans from proteins and can potentially fractionate glycans into neutral and acidic fractions, large column volumes of resin (5mL) are typically employed, and sample losses can be a problem.

Additionally, Qproteome sells an O-glycoprotein purification kit containing amylase inhibitor-like lectin (AIL) and PNA lectin, which bind to NeuAcα2-6Galβ1-3GalNAc and Galβ1-3GalNAc residues, respectively [11]. Glycans are purified chromatographically, with specific elution buffers competing with the glycans for lectin binding.

Lastly, a number of resins or solid fixed adsorbents exploit the hydrophobic properties of proteins and glycans to separate them. These chromatographic materials can be used in the bench-top or FPLC columns. Sepabeads SP20SS resin, for example, is an aromatic synthetic adsorbent based on a crosslinked polystyrenic matrix with a large surface area and activated carbon [12]. This resin can effectively adsorb organic compounds like proteins from aqueous solutions because of the compatibility between the protein and resin hydrophobicities [12]. O-glycan desalting is typically done using a resin like Dowex's Sulfonated Copolymer of Styrene and Divinylbenzene ion exchanger resin [13]. As always, the main issue with purifying O-glycans with resins or fixed adsorbents is the possible sample loss due to the addition of at least one purification step and unforeseen glycan-resin interactions.

The goal of the O-glycan purification optimization was to identify the purification process that most effectively removed chemical contaminants from samples (i.e. salts) while minimizing sample loss and the number of purification steps. Two methods of purification, each involving only one major step, were initially scrutinized. Supelco column purification could remove salts and contaminants while also separating neutral and acidic glycans by using increasing concentrations of acetonitrile to elute them.

However, this method has several drawbacks. Since glycans and proteins are retained in the column, and since it is still unclear how Supelco columns separate glycans based on hydrophobicity and charge, it is possible that some proteins can elute with glycans and further complicate MS analysis. Lastly, since the supplier recommends that elution take place with 3ml of elution solvent, sample losses can be significant since several concentration steps need to be employed to get the sample down to a working volume of ~ 5ul.

The second purification procedure only involves one step, and has been shown to effectively remove salts and other contaminants from biological samples using SP2055 hydrophobic beads and Dowex Cation Exchange Beads packed into a GELoader tip [7]. The only disadvantage to this method is that acidic and neutral glycans cannot be separated because there is no elution step. There is also a minor chance that other contaminants such as reducing agents, could be retained in the sample because of their polar or charged nature. However, care was taken to minimize the presence of these species during the glycan extraction step.

2.2.2 Materials and Methods

O-glycan Supelco cleanup was performed with Supelco Envi-Carb 3mL columns. Columns were equilibrated with 3mL of ACN followed by 3mL DIW. The sample was added in 50ul, and then 3mL of DIW was added to wash the column. Elutions of 25%, 50%, and 100% ACN (3mL) were carried out, and the flow-through was collected in each case. Samples were lyophilized, re-diluted in 400ul DIW and transferred to smaller Eppendorfs. Three successive dilution/lyophilization processes were preformed to concentrate the glycans (in 80ul and 20ul). Glycans were then stored at -20C or diluted in 3ul of DIW prior to MS analysis.

O-glycan GELoader Sample Cleanup was carried out using SP20SS and Dowex H^+ Cation Exchange resin, as previously described [7]. Briefly, GELoader tips were pinched at the bottom prior to being filled with the following ratios of beads: 20µl SP2055: 40µl of Cation Exchange, 20µl SP2055: 60µl Cation

Exchange, 30µl SP2055: 60µl Cation Exchange, 30µl SP2055: 80µl Cation Exchange. After being washed with approximately 200µl of DIW, the sample was added in 20µl, and the glycans were caught in the flow-through. The beads were washed with an additional 200µl to remove any loosely bound glycans such that a total fluid volume of 220µl was collected. When several samples needed to be purified, four samples could be simultaneously purified using a multi-channel pipetteman. Care was taken to avoid drawing air into the micropipette column so that the resin bed was not disturbed. Post-purification, samples were lyophilized and subjected to two additional dilution/lyophilization cycles (in 40µl of DIW followed by 10µl of DIW) in order to concentration the glycans. The final samples were either stored at -20° C or diluted in 3µl of DIW directly prior to MS analysis.

2.2.3 Results and Conclusions

In order to determine the optimum O-glycan purification procedure, purified commercial O-glycan standards were utilized to first assure that extreme sample losses would not affect the quality of the O-glycan spectra. Figure 2.5 shows O-glycan standards purified using Supelco columns for both MS methods. It is important to note the trifluoroacetic acid (TFA) was not used to enhance elution of the glycans (as it is for the N-glycan purification process) because the TFA produces multiple adduct peaks for each O-glycan standard (data not shown).

The Bruker data (Figures 2.5 A-D) shows that regardless of the concentration of acetonitrile used to elute the O-glycans from the Supelco, the original spectra (Figure 2.5 A) cannot be reclaimed. The Voyager data in Figure 2.5E-H yields similar conclusions about the efficacy of Supelco purification for O-glycans. Because of the perfluorinated Nafion spot, all adduct related peaks are not present in any spectra. However, the intensities of 3'SL, 3'SLN and LST-a are extremely reduced compared to the original unpurified spectra (Figure 2.5 B-D). Therefore, the main conclusion from the data in these spectra is that the O-glycan standards, which are purified and contain no protein or chemical contaminants, were not detectable at acceptable levels after purification by Supelco columns. Both Bruker

and Voyager spectra support this conclusion, and thus, this purification method was abandoned as a potential method for O-glycan purification.

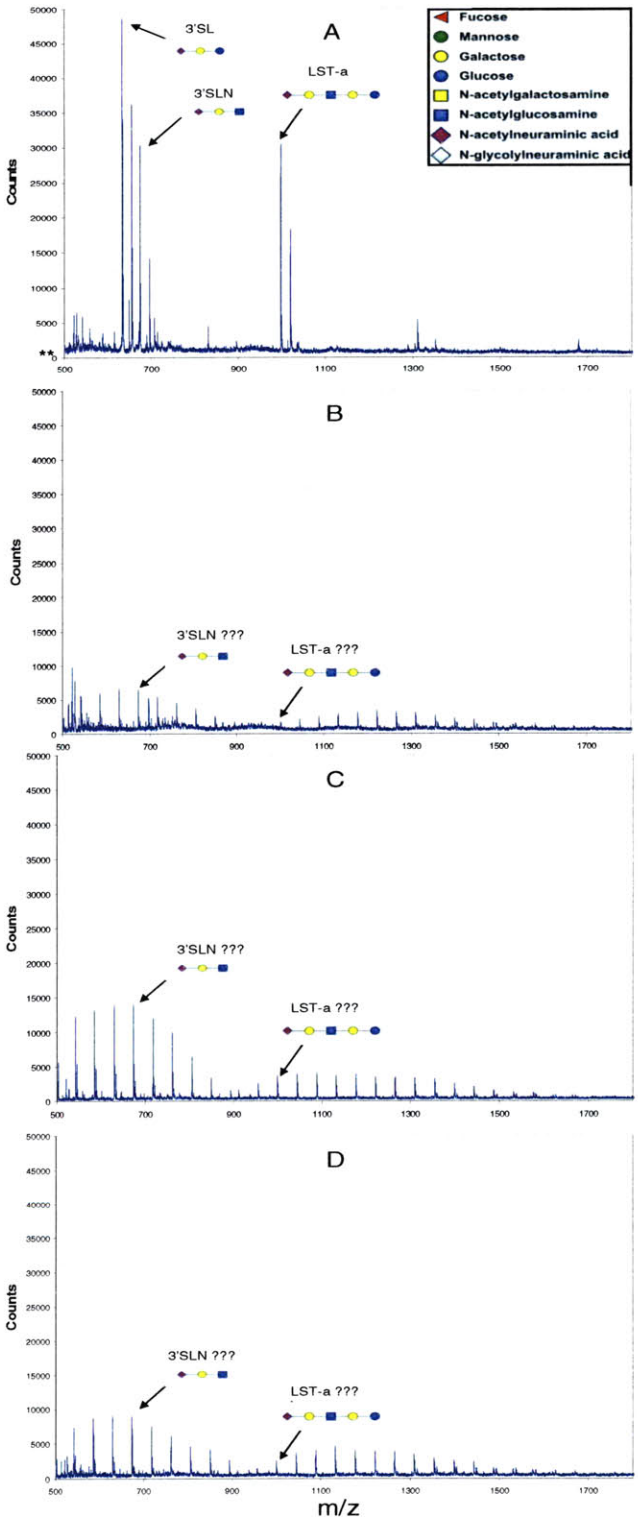


Figure 2.5. Supelco Purification of Acidic O-Glycan Standards. Spectra in A-D (in the ratio 33pmol LST-a: 33pmol 3'SLN: 132pmol 3'SL) is for the Bruker MALDI-TOF, while E-H is for the Voyager MALDI-TOF (in the ratio 6pmol LST-a: 6pmol 3'SLN: 24pmol 3'SL). A,E. Unpurified commercial O-glycan standards. Standards were eluted with B,F. 15% ACN in DIW, C,G. 50% ACN in DIW D,H. 100% ACN in DIW.

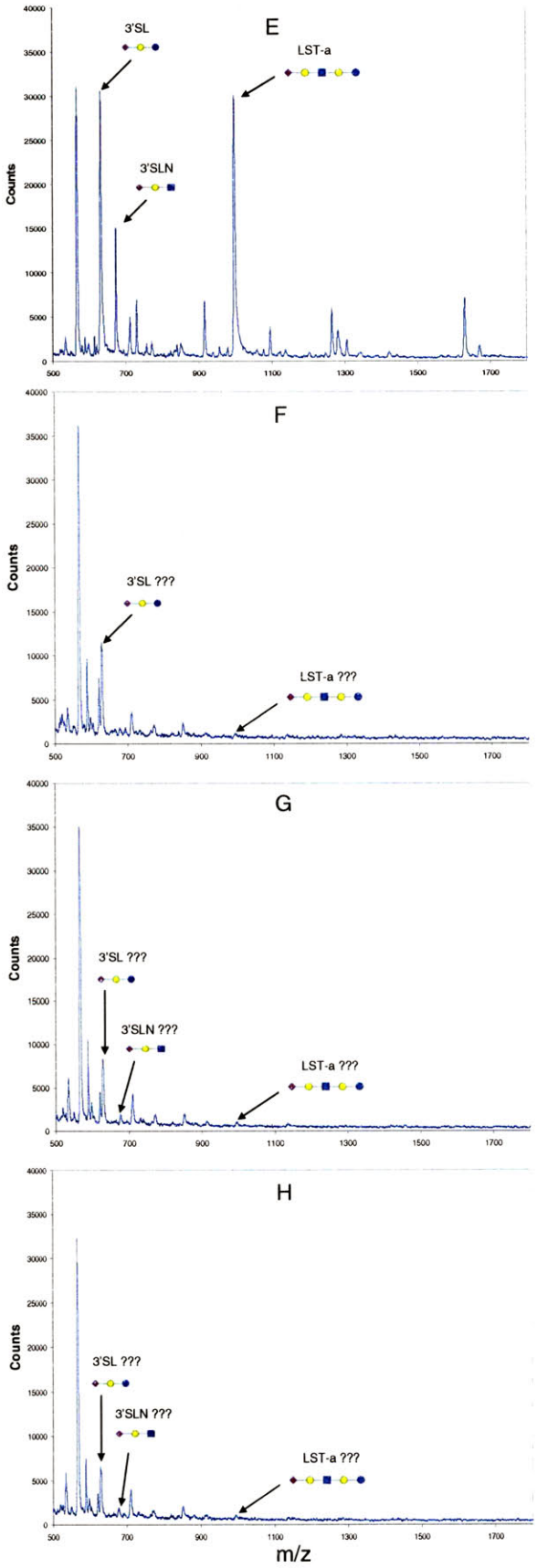

E
3'SL
LST-a
3'SLN
Counts
F
3'SL ???
LST-a ???
G
3'SL ???
3'SLN ???
LST-a ???
H
3'SL ???
3'SLN ???
LST-a ???
m/z

Figure 2.6 shows the purification results for the GELoader tip filled with 20ul of SP20SS resin and 40ul of Dowex Cation Exchanger Resin on both the Voyager and Bruker MALDI-TOF machines. Figure 2.6B shows the efficacy of this strategy because all salt-produced peaks are removed from the O-glycan standard spectra. However, the standard peak intensities are much lower than the original, indicating that sample losses are present and can be significant. It was expected that this purification would have a larger impact on the Bruker analysis than that of Voyager because of the lower tolerance of the former to glycan amount. Figure 2.6C and 2.6D, which both come from the Voyager instrument, fulfill this expectation by showing that the purified and un-purified samples both show strong O-glycan standard peaks. The purified peaks show a loss of half of the intensity of the unpurified standards, but even this signal is excellent. Overall, because Figure 2.6 shows that this method can efficiently purify commercial O-glycan standards with minimal signal loss, this method was further scrutinized for its effectiveness in purifying β-eliminated O-glycans. Furthermore, since the Voyager method has a lower detection limit for both unpurified and purified glycans, this method was the primary method used in the analysis of all β-eliminated samples.

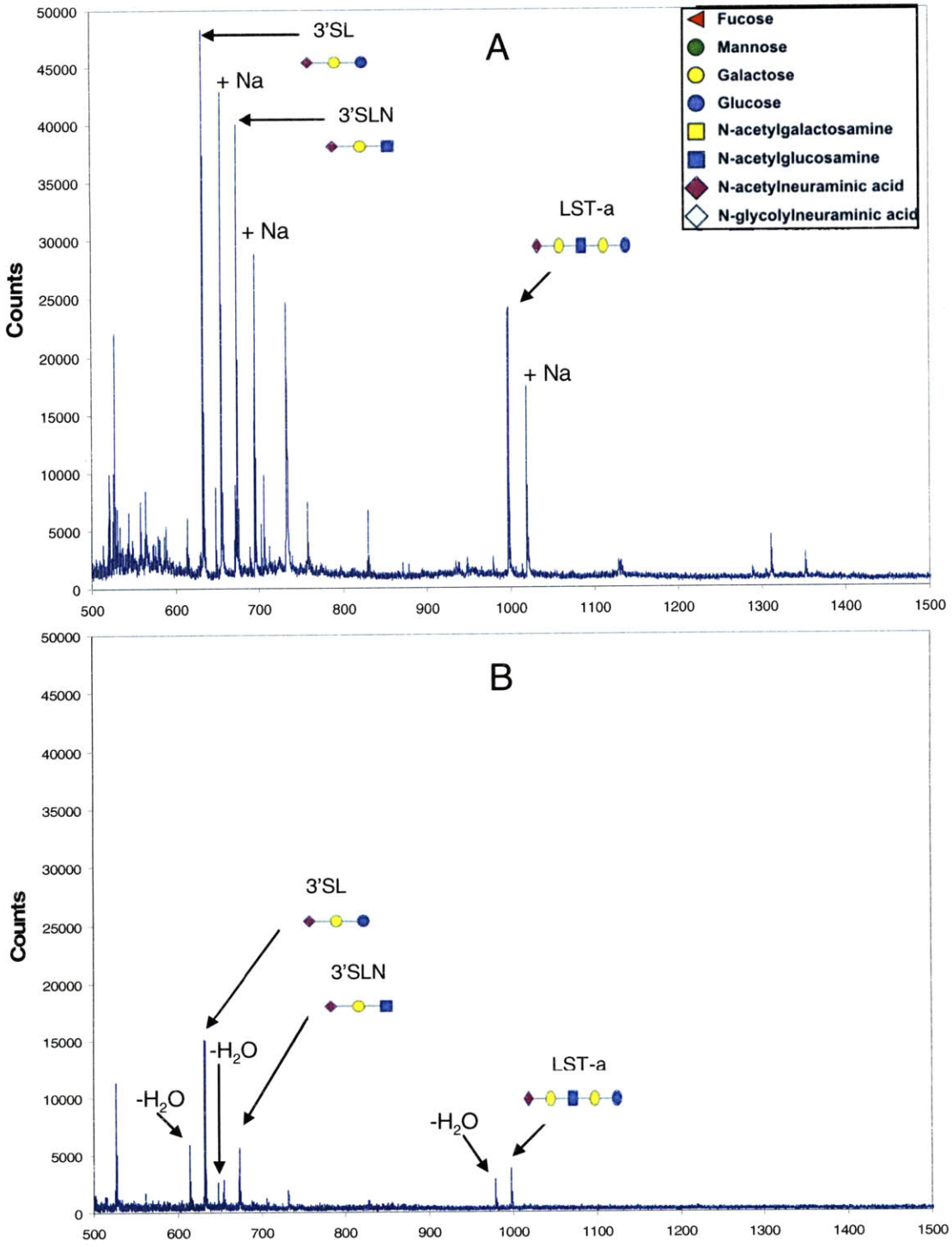
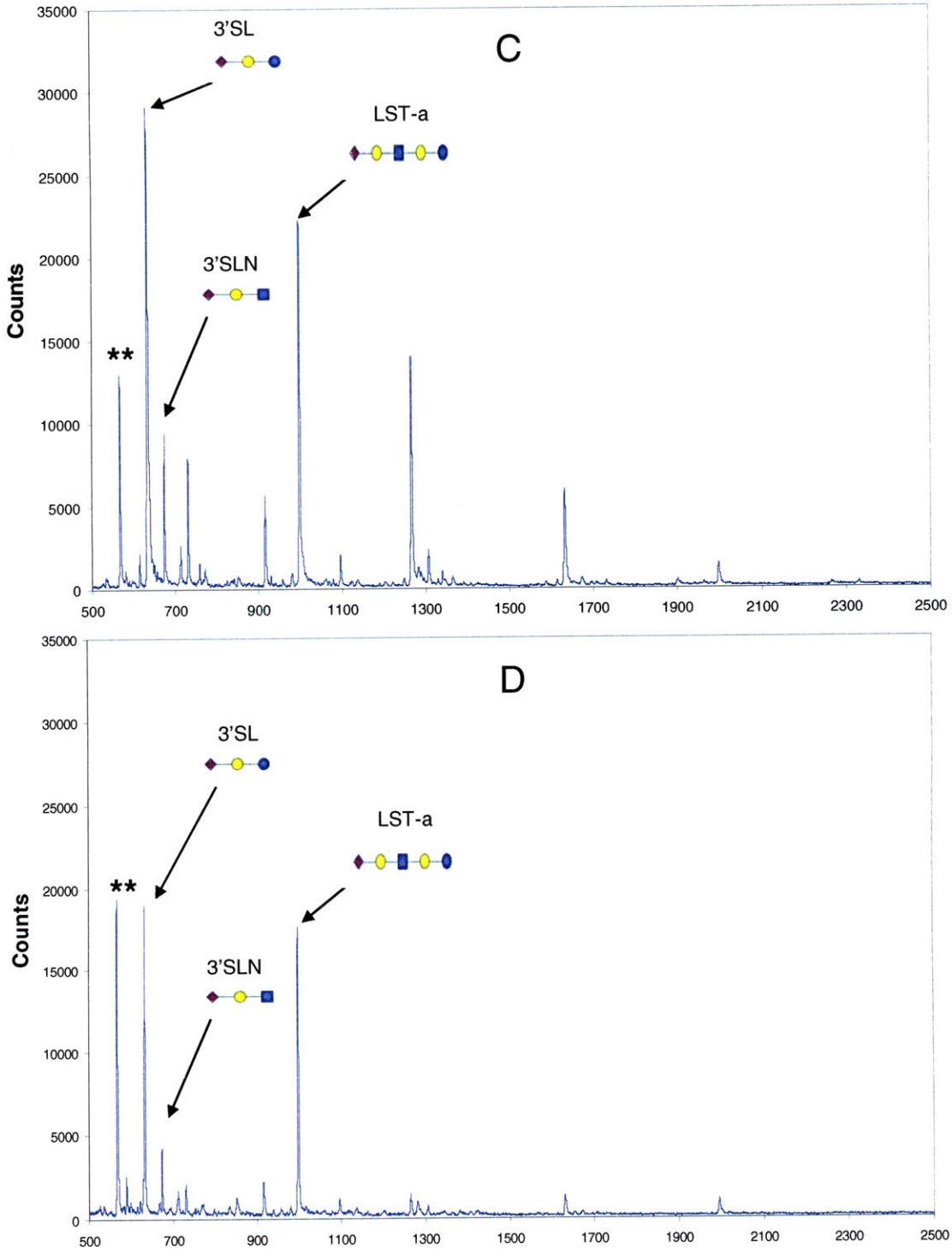


Figure 2.6. GELoader Purification of O-glycan Standards. O-glycan standards were purified using a GELoader microcolumn filled with 20ul SP2OSS resin and 40ul of H^+ Cation Exchanger. Samples were analyzed on the Bruker (A,B) and Voyager (C,D) MALDI-TOF machines. A,B. Unpurified O-glycan standards C,D GELoader purified standards. Standards are present in the same amounts on the spot for each MALDI-TOF machine. ** = matrix peaks.


C
3'SL
LST-a
3'SLN
**
Counts
35000
30000
25000
20000
15000
10000
5000
0
500
700
900
1100
1300
1500
1700
1900
2100
2300
2500
D
3'SL
LST-a
3'SLN
**
Counts

Since the SP20SS/Dowex Cation Exchanger resin purification was superior to the Supelco method, this method was further scrutinized to determine the maximum amount of material that could be passed through the GELoader tip before the purification efficiency was compromised. In order to address this issue, different amounts of bovine fetuin (1ug, 5ug, 25ug, and 125ug), beta-eliminated with the optimized procedure described in the "Section 2.3 O-glycan Release", were added to a GELoader tip filled with 20ul of SP20SS and 40ul of Dowex cation exchanger resin. Bovine fetuin was chosen instead of commercial O-glycan standards because it's a "messier" sample; processed bovine fetuin contains residual reducing agent, salts, glycans and protein while commercial glycan standards only contain glycans and salt. This "messier" sample can give a more realistic idea of the purifying capacity of the GELoader tips, which will become particularly important when biologically significant samples are analyzed.

Figure 2.7 shows the purifying capacity of a 20:40 mixture of SP20SS resin and H^+ Dowex resin. Only Voyager MS analysis was performed for this experiment because it represents the most sensitive method for glycan detection. At minimum, the glycans must be detected on this machine in order for the purification procedure to be acceptable. Even though the three fetuin O-glycans are detectable at 1ug molar amounts (Figure 2.7A), the signal to noise ratios of the respective glycans quite low, and it is unlikely that these peaks would be scrutinized as potential glycan peaks in any sample. Furthermore, this amount of fetuin roughly correlated to approximately .6pmol on a MALDI plate, which is around the 3.75 times the detection limit on the Voyager machine. Thus, this data is in agreement with the detection limit data in Figure 2.4. Overall, Figures 2.7B-E show that the GELoader purification scheme is robust, and can efficiently purify samples with protein amounts up to 125ug, an amount which is probably over the realistic amount of protein acquired from a biological sample.

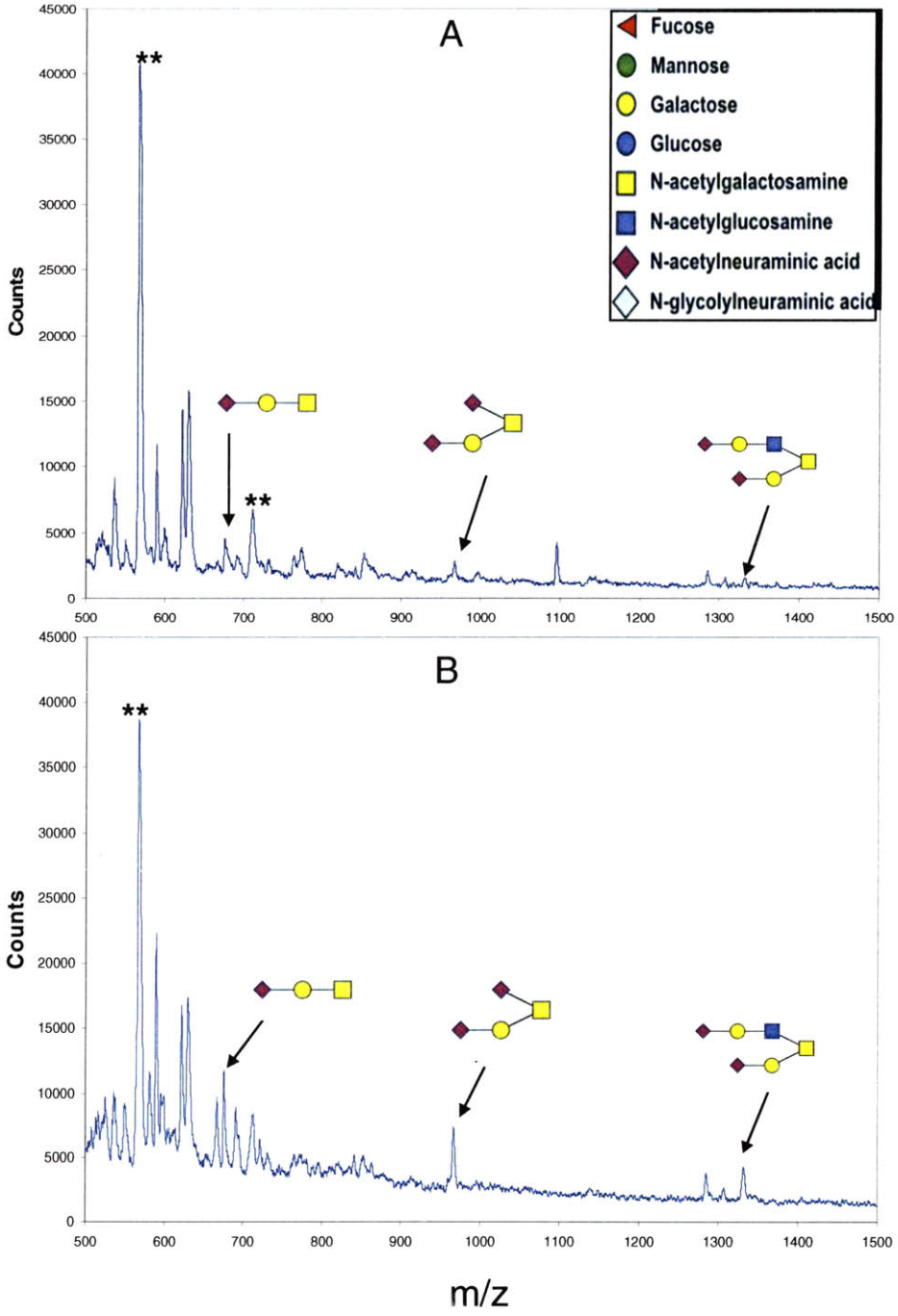
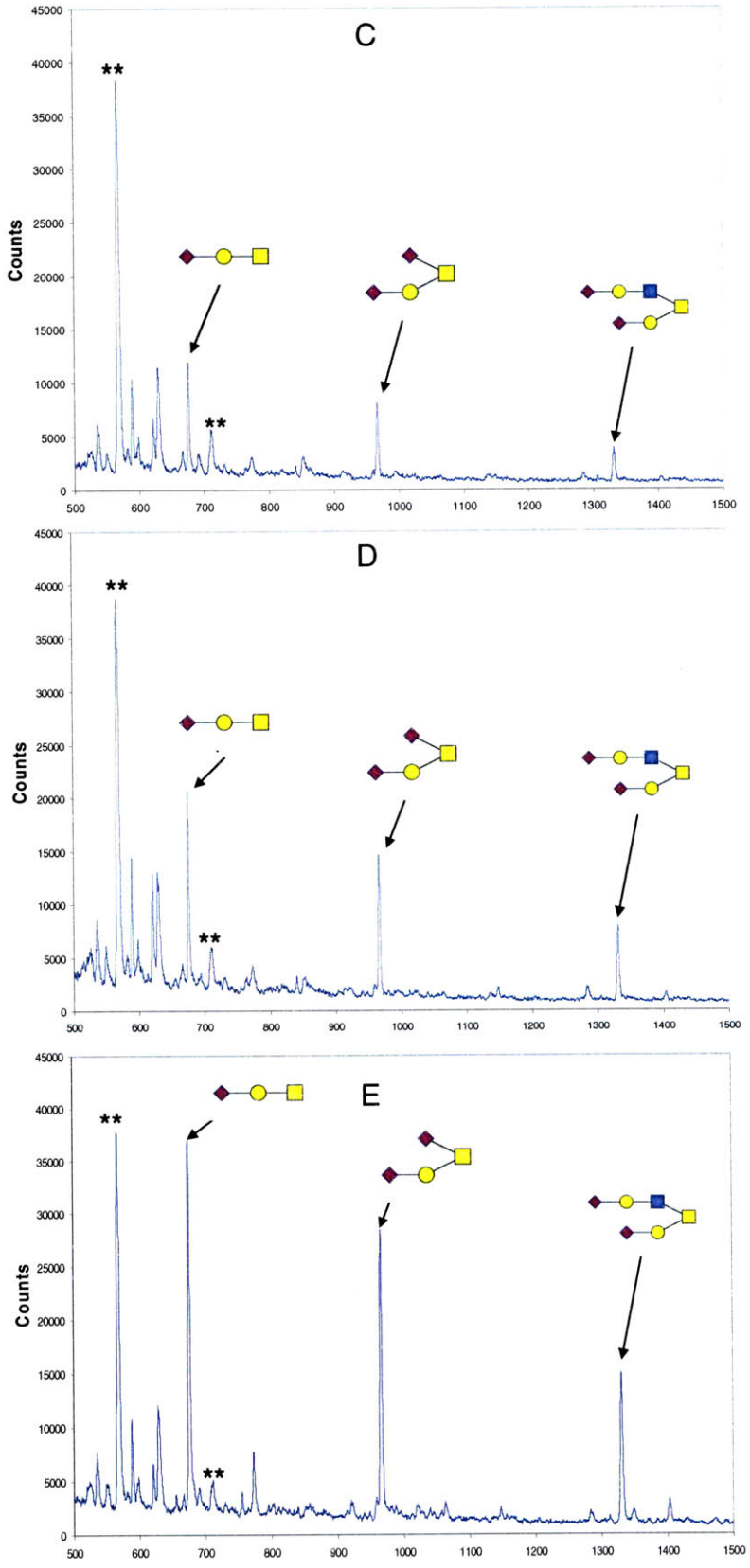


Figure 2.7. The Purification Limit of GELoader Micro-purification Columns. All columns were filled with 20ul SP20SS resin followed by 40ul of H^+ Dowex Cation Exchanger, as described in the Materials Section. Fetuin was prepared as a single batch in 1mg aliquots. The appropriate volume was then purified and collected. An additional 200ul (to wash the column) was collected for each case. The following amounts of bovine fetuin (and their released glycans) were purified: A. 1ug, B. 5ug C. 10ug D. 25ug E. 125ug. ** = matrix peaks.


C
D
E
Counts
m/z
**

In conclusion, the purification optimization strategy succeeded in identifying a method that effectively purified beta-eliminated sample from salts and protein, while minimizing sample loss. This method involves using two types of resins; SP20SS resin serves as a reverse phase resin that captures the hydrophobic protein while allowing the polar glycans to pass, while the H^+ Dowex Cation exchanger resin effectively removes alkali salt, which can suppress glycan ionization from the sample. It is possible that there is a more optimum ratio of resins for a specific amount of material, but since the amount of material in a biological sample is often unknown, it was best to test the feasibility of one purification process on a range of sample amounts. It is also important to note that, while this project focuses on the cleanup related to beta-elimination of O-glycans, it might be possible to use this purification process for other glycan-related purposes. For example, one could perform exoglycosidase studies on a small fraction of released (but unpurified) O-glycans, and then use this method to purify the sample of the exo-glycosidase and other salts. However, this method might not be appropriate for the purification of samples processed with contaminants like reducing or alkylating agents, since these species will not be removed by the resin, and might consequently obstruct glycan ionization. In these cases, an additional purification step would be required, or an additional resin would need to be added to the purification scheme.

The Supelco method of purification, although effective for N-glycans, results in unacceptable O-glycan sample loss, presumably because O-glycans less hydrophobic, tend to elute sooner, and must be concentrated down to a 3ul working volume from a 3mL elution volume. Therefore, this purification method was abandoned.

2.3 O-GLYCAN RELEASE

2.3.1 Background

Even though glycan structural information can be obtained from both O-glycans and O-glycopeptides, this thesis focuses on the development of a method that generates free O-glycans. O-glycans can be released both chemically and enzymatically, although the former is widely preferred. While N-glycans can be enzymatically released from their protein backbones by PNGase F, O-glycan enzymatic release is hindered by the limited availability of O-glycan specific enzymes. The most readily available commercial enzyme, O-glycanase, has a specificity restricted to the Gal-GalNAc disaccharide sequence, making the release of complex oligosaccharide chains from proteins with significance to human health and disease difficult [14]. Furthermore, resistance to de-glycosylation by O-glycanase does not necessarily confirm the absence of O-glycans. Due to these confounding issues, most researchers rely on chemical methods to ensure a more extensive de-glycosylation.

O-glycans are most commonly chemically released by either hydrazine treatment or alkaline beta-elimination, although other chemical methods have been established [7, 16, 17]. Hydrazine removes both N- and O- glycans by cleaving the glycan-peptide bond, but has been reported to selectively remove O-glycans at a temperature of 60° C [15]. A major advantage of hydrazine treatment is that the released sugars can still be labeled with chromophores or other tags since the reducing terminus of the glycan is intact. However, hydrazine treatment has several major disadvantages. Firstly, no information on the protein glycosylation site can be obtained because all peptide bonds are destroyed. Secondly, since acyl groups are cleaved from N-acetylamino sugars and sialic acids, a re-acetylation step, which can increase sample losses, must be added to the release procedure [15]. This step can also cause a small amount of acetyl substitution on hydroxyl groups as well as acetylation of sialic acids that did not previously contain acyl groups. Analysis of hydrazine treated glycans is further complicated by the fact that the re-acetylation procedure increases the glycan mass by an amount equal to that between a hexose and an aminohexose group [15].

For the aforementioned reasons, alkaline beta elimination is the preferred method of O-glycan release (Figure 2.8). Just like hydrazine, beta elimination releases both N- and O- linked glycans, but is reported to selectively release O-glycans as oligosaccharide-alditols when a dilute alkali (50-500mM) is used [17]. Because certain glycans are more resistant to beta-elimination than others, the elimination reaction is typically carried out for one to forty-eight hours, and is often optimized with respect to the protein and glycans being analyzed. The elimination reaction converts the glycan-bound amino acids to unsaturated hydroxyamino acids, such that an increase in absorbance can be detected at 240nm, and the glycosylation site can be determined [17].

In typical beta-elimination procedures, a reducing agent such as sodium borohydride is included in the alkaline solvent because even dilute alkali solvents can cause destructive glycan peeling, which ruins the glycan and further complicates MS analysis [17]. However, the reducing agent has been shown to cause protein degradation, and it is therefore omitted when the experimental goals include obtaining glycosylation site information [17, 18, 19].

The reducing agent concentration can also affect the efficiency of glycan purification and glycan stability. High concentrations can lead to downstream purification problems, since the reducing agent is packaged with an alkali metal, which can ultimately shield the negative charges that would normally allow the glycan passage through the hydrophobic purification media. Furthermore, even though cation exchanger is supposed to solve this problem by removing all alkali metals prior to the reversed phase separation, excessive amounts of salts as a result of high reducing agent concentrations can even make this step inefficient. On the other hand, lower reducing agent concentrations can be less efficient at preventing glycan degradation in alkaline conditions. Therefore, the optimal reducing agent concentration must ultimately maximize the ease of glycan purification while minimizing the likelihood of glycan degradation. In this study, both 2M and 0.8M reducing agent concentrations were used, with the former and latter representing the upper and lower concentration bounds typically used in beta-elimination procedures [17]. Lastly, the choice of alkaline solvent can also influence purification results because of

the different binding strengths of the alkali metals to the charged species; therefore, the optimum solvent will ultimately provide cleaner MS spectra because the associated alkali metal will have been removed the most efficiently.

However, it is important to note that the protective capabilities of the reducing agent also lead to the removal of the reducing terminus, which prevents labeling of the glycan with chromophores and other tags [17]. The reducing terminus of the oligosaccharide can be labeled in this procedure only if the reducing agent is removed and the labeling agent is added after the elimination (Figure 2.8). By doing this, the stability of the glycan can be compromised because of the time lag in between glycan elimination and labeling. Glycan stability in this situation can be controlled by adjusting the temperature of the reaction and shortening the experimentally determined incubation time.

In summary, the release of O-glycans is mainly accomplished by chemical treatment via hydrazine or alkaline beta elimination, since the specificity of O-glycanases severely restricts O-glycan research efforts. Of these two chemical methods, alkaline beta elimination is by far the most utilized because of the additional steps and analytical ambiguities conferred by hydrazine treatment. Beta-elimination can provide a quick and reliable source of O-glycans, and has very little glycan specificity. Depending on the experimental goals, reaction temperature, incubation time, and reducing agent concentration can be easily adjusted. Therefore, the experimental flexibility of the beta-elimination procedure provides a great framework for different experimental goals if the conditions are chosen wisely.

2.3.2 Materials and Methods

2.3.2.1 Optimization of Reagent Concentrations

In order to simultaneously test reducing agent concentration and choice of alkaline solvent, fetuin (100ug at 10mg/ml concentration) was incubated in either of the following reagent combinations: 2M KBH_4 in 0.1M KOH, 2M $NaBH_4$ in 0.1M NaOH, 0.8M KBH_4 in 0.1M KOH, 0.8M $NaBH_4$ in 0.1M NaOH. These particular reducing agent concentrations (2M and 0.8M) represent the upper and lower

limits of reducing agent concentrations typically used in beta elimination procedures [4]. Beta elimination of fetuin was carried out according to a slightly modified method that was previously published [4]. Briefly, an equal volume (as that of 10mg/mL protein) of the reducing agent/alkaline solution was added to the aqueous protein solution. Samples were incubated at 45°C for 16 hours. Post-incubation, the sample was cooled to room temperature, and the excess $NaBH_4$ was quenched by adding 5 sample volumes of 0.25M acetic acid in methanol. The acetic acid and methanol was then evaporated using a SpeedVac (Savant). To ensure that all reducing agent was destroyed, this step was repeated twice more. To remove the boric acid, this step was repeated twice more with only methanol. Samples were then stored at -20C until further use, at which point they were diluted in water at a ratio of 20ul to 10ug of protein. Twenty microliters was then taken for each purification procedure unless otherwise stated.

2.3.2.2 Length of Incubation Studies

The optimal beta elimination incubation time was determined by incubating six 50ug samples of fetuin and bovine mucin (in duplicate) for 1hr, 6 hrs, 12hrs, 18hrs, 24hrs, or 48 hrs. The reaction was carried out with 1M KBH_4 in 0.1M KOH, and was completed as previously described in Section 2.3.2.1.

2.3.2.3 Sialic Acid Linkage Analysis

Sialic acid linkage analysis was performed by extracting glycans with the optimized beta elimination extraction protocol, and then treating 20ul of this mixture (diluted in water at a ratio of 10ug protein: 20ul DIW) with 3ul of Sialidase A or S. These mixtures, along with an untreated sample, were incubated at 37C for 16 hours on a rocker, and then purified and analyzed according to the optimized protocols described in Sections 2.1 and 2.2. For this experiment, only acidic glycans were analyzed on the MALDI-TOF instrument.

2.3.3 Results and Conclusions

2.3.3.1 Reagent Choice & Concentration Results

Since alkaline beta elimination was the primary method of glycan release used in this project, the reducing agent concentration and choice of alkaline solvent was scrutinized. Figure 2.8 illustrates the effect of reducing agent concentration and choice of alkaline reagent on the Voyager MS spectra of fetuin O-glycans. Since the matrix formulation used in this figure only allows the ionization of acidic glycans, salt-derived peaks were not considered in the analysis of this figure. This figure was primarily used to establish a preliminary beta elimination reagent "recipe" that would be used to further explore O-glycan analysis.

The results of this assay show that samples incubated with 2M $NaBH_4$ or 2M KBH_4 (Figures 2.8A and 2.8C) yield a cleaner spectra and fewer peaks associated with the loss of a carboxylic acid. It is possible that such a high reducing agent concentration most efficiently protects the glycan from degradation, such that higher glycan intensities and fewer degradative peaks are observed. These two panels also suggest that samples incubated with KBH_4 yield a cleaner MS spectrum that those incubated with $NaBH_4$. These results could arise from the fact that the potassium ion is easier to remove during the purification process because it has a weaker energy of interaction with negatively charged molecules when compared to sodium, which is much smaller and contains fewer atomic orbitals. Overall, the results of this experiment suggest that KBH_4 and KOH are the best reagent choices for the beta-elimination procedure. Furthermore, a 2M KBH_4 incubation seems to be more effective than a 0.8M KBH_4 incubation. However, because this assay was only used to establish a preliminary reagent recipe, fetuin was also incubated in other concentrations of KBH_4 in between 2M and 0.8M. Ultimately, the goal of this optimization was to find the lowest reagent concentration that yielded the most optimum fetuin spectra with the smallest amount of alkali metal to remove in the purification process. The results of this additional experiment showed that 1M KBH_4 in 0.1 KOH worked most efficiently (data not shown), and so this reagent combination is used for all future beta elimination experiments.

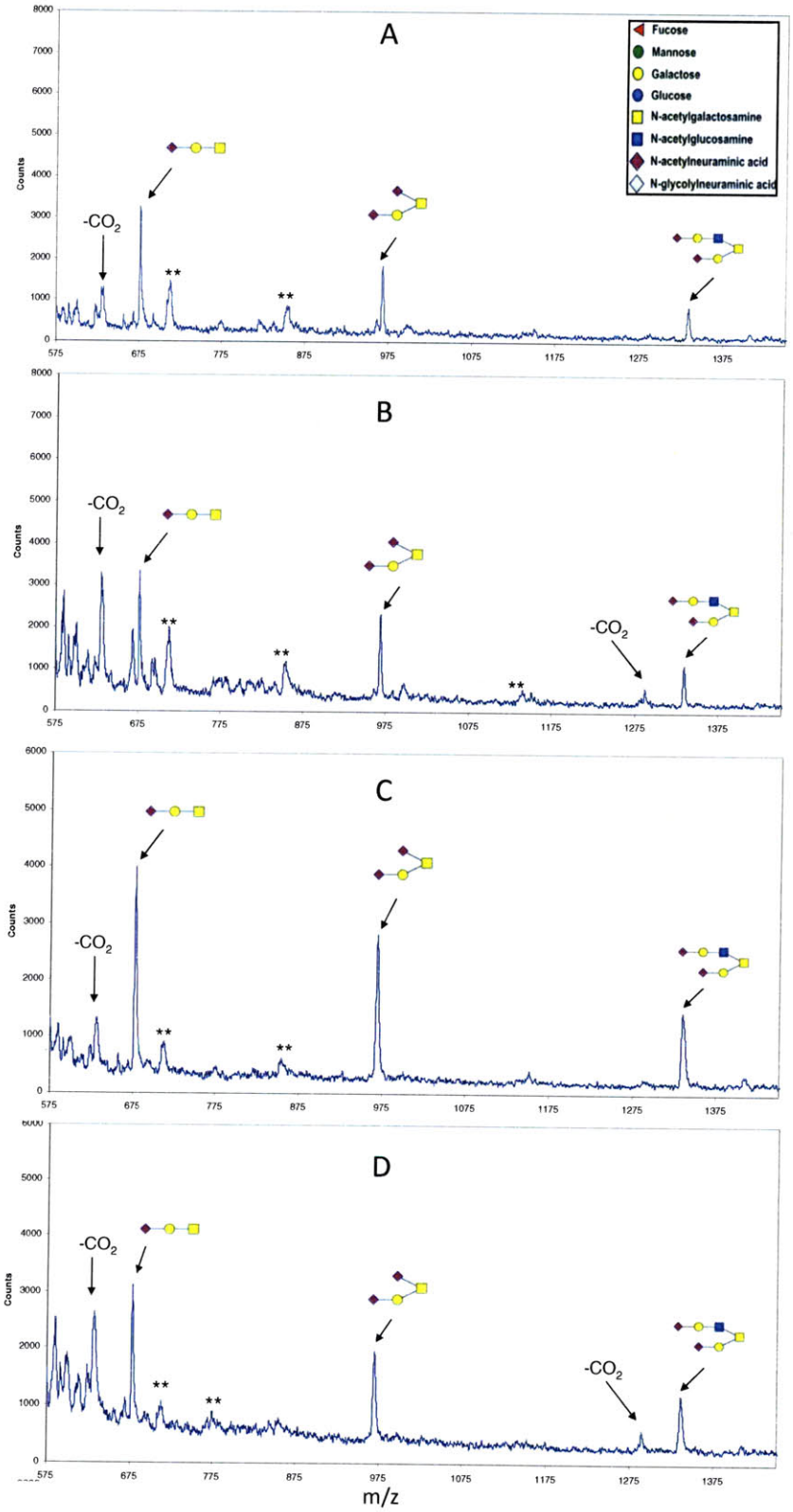


Figure 2.8. MALDI MS Spectra for Fetuin O-glycans Treated with Different β- Elimination Reagent Combinations. A. 2M $NaBH_4$ in 0.1M NaOH B. 0.8M $NaBH_4$ in 0.1M NaOH C. 2M KBH_4 in .1M KOH B. 0.8M KBH_4 in 0.1M KOH. ** = matrix peaks.

2.3.3.2 Length of Incubation Study Results

In order to investigate the effect of the length of β-elimination incubation on the quality of O-glycan mass spectra, fetuin and bovine mucin samples were incubated for 1, 6.5, 12.5, 18.5, 25, and 48 hours with 1M KBH_4 in 0.1M KOH. Bovine submaxillary mucin was used in order to better understand the relationship between the number and type of O-glycans on a glycoprotein and the incubation length, while bovine fetuin was used to determine the susceptibility of N-glycans to cleavage in the O-glycan elimination process. Since glycans from both of these proteins have been extensively profiled, the MS spectra obtained from these experiments could easily be compared to previously published MS spectra.

Figures 2.9 shows the mass spectra for bovine fetuin at every incubation time point on the Voyager MALDI-TOF machine. In analyzing the spectra of fetuin from Figure 2.9, it is immediately evident that N-glycans are just as susceptible to cleavage by beta elimination in dilute alkaline conditions as O-glycans. Figure 2.9A shows that at 1hr, all three O-glycans are fully distinguishable and have signal to noise ratios that are well above the accepted ratio of 3:1, while three of the five prominent N-glycans are only partially distinguishable. However, all of fetuin's N- and O-glycans are distinguishable by 6hrs (Figure 2.9B), and continue to be apparent, although to a lesser extent, at the 48hr mark (Figure 2.9F). Furthermore, Figure 2.9 suggests that O-glycans are present at maximal intensities after a 25 hr incubation, after which the intensities slightly drop. Interestingly, even though N-glycan release was not an intended result of this experiment, this data suggests that it is inevitable for all incubation times.

It is important to note that since this data was acquired in the linear mode, the resolution of all glycans shown in Figure 2.9 ranges from 133 to 280, which indicates that isotopes of each glycan are not fully resolvable. However, at this stage in glycan analysis, an average molecular weight, as opposed to a monoisotopic weight yielded at higher resolutions, is sufficient to propose glycan structures for a particular m/z, especially since the mass accuracy of the Voyager MALDI-TOF is less than 1 Da for the mass range of 300 to 5000.

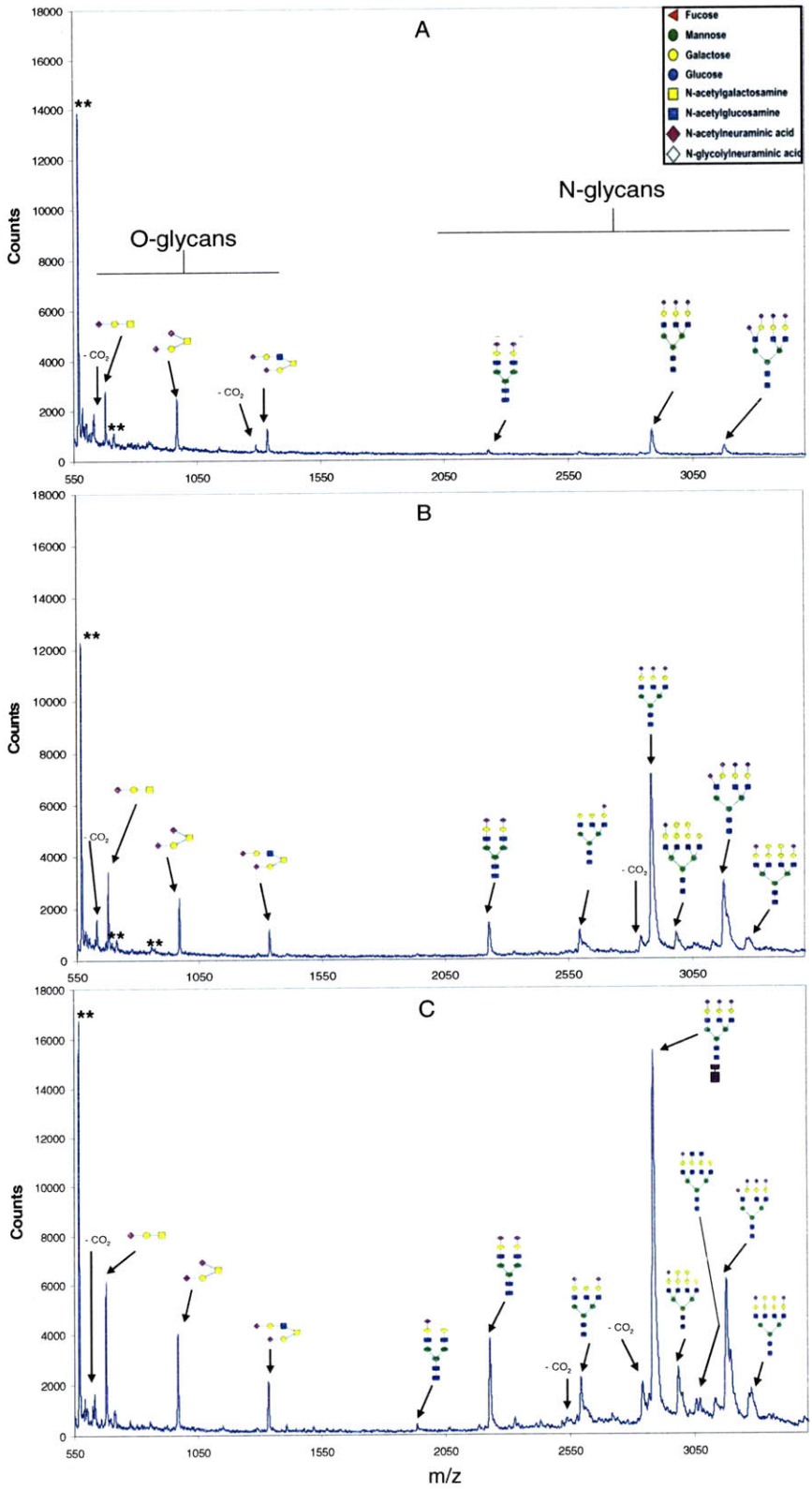
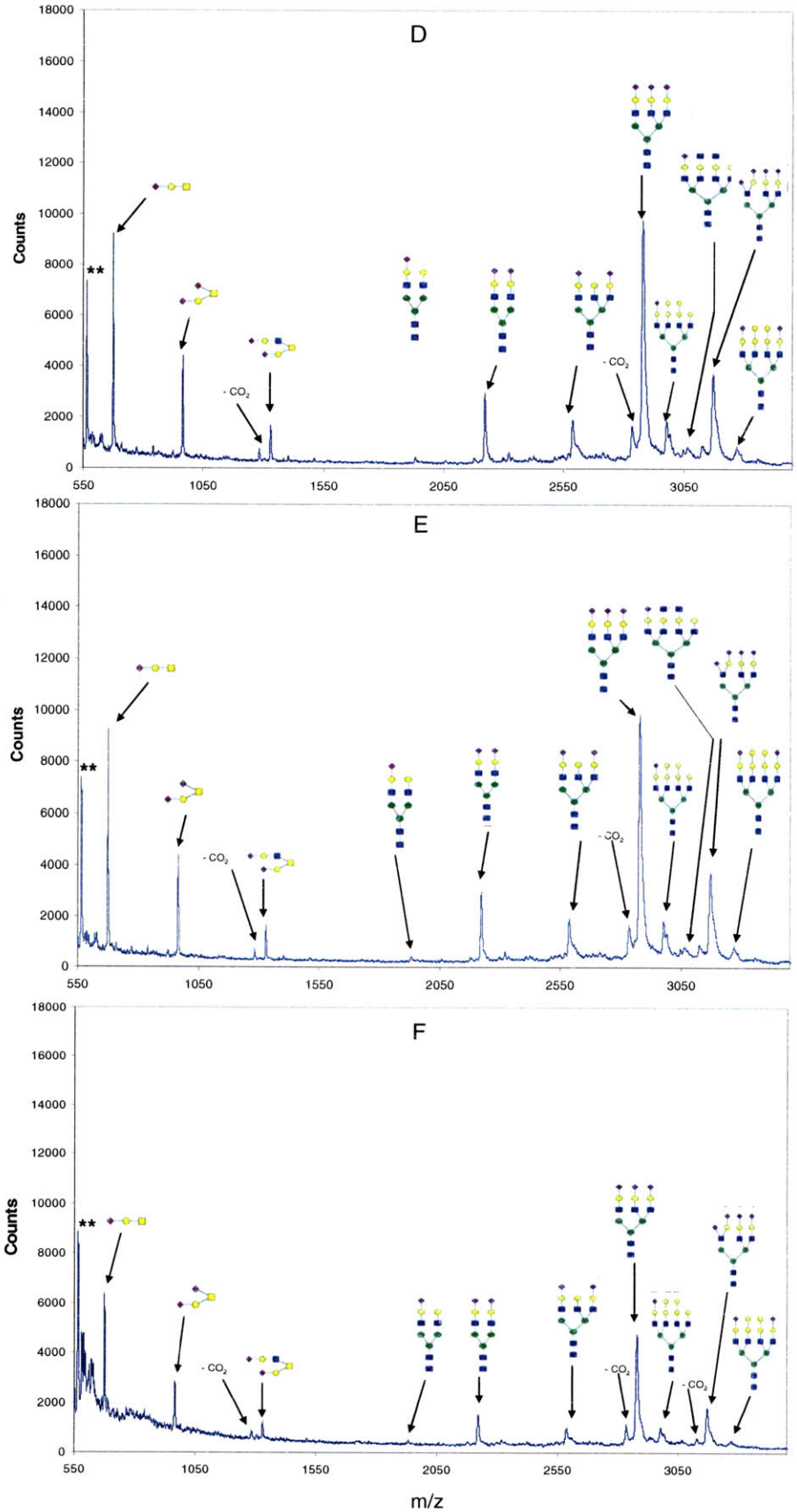


Figure 2.9. The Effect of β-elimination Incubation Length on Fetuin O-glycan Spectrum. A. 1hr B. 6hrs C. 12hrs D. 18hrs E. 24hrs F. 48hrs. ** = matrix peaks.


D
E
F
Counts
m/z
**
- CO2
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
550
1050
1550
2050
2550
3050

Because of the large diversity of glycans on bovine submaxillary mucin, only four previously reported glycans (Table 2.6) were profiled over the course of 48 hours. Tracking these glycans provides an understanding of how incubation length corresponds to release of O-glycans with a range of sizes and compositional complexities. Figure 2.10 shows that the intensities for each glycan generally increase with time, and don't seem to diminish much by 48 hours, even though there is a slight dip at 18hrs, presumably because of the variable nature of sample purification and preparation. Generally, the data in Figure 2.10 suggest that heavily O-glycosylated proteins, such as mucins, will yield strong MS spectra after 12-48 hour incubation periods.

Table 2.6 Profiled O-glycans of Bovine Submaxillary Mucin

Glycan m/z (calculated)	**Proposed Composition**
514.45	NeuAc GalNAc-ol
717.65	GlcNAc (NeuAc) GalNAc-ol
1042.06	NeuAc Gal_2 GlcNAc GalNAc-ol
1246.00	GalNAc (Fuc) Gal GlcNAc (NeuGc) GalNAc

As with the results for bovine fetuin, the resolution of bovine mucin O-glycans was on the low end because the analysis was carried out in the linear mode. The MS resolution of O-glycans from heavily glycosylated proteins can be particularly important because of the presence of glycans that are very close in molecular weight. In order to detect these glycans, and even suggest the presence of new structures, the O-glycan peaks must be distinguishable from the baseline and must stand alone as a single peak. Furthermore, since some glycans on heavily glycosylated proteins are present in minute amounts, resolutions above 2000 are sometimes unattainable, and thus the signal to noise ratio can be used to propose potential glycan peaks. In this study, bovine mucin peaks with signal to noise ratios above 50 are considered peaks of interest, and were eligible for structural scrutiny.

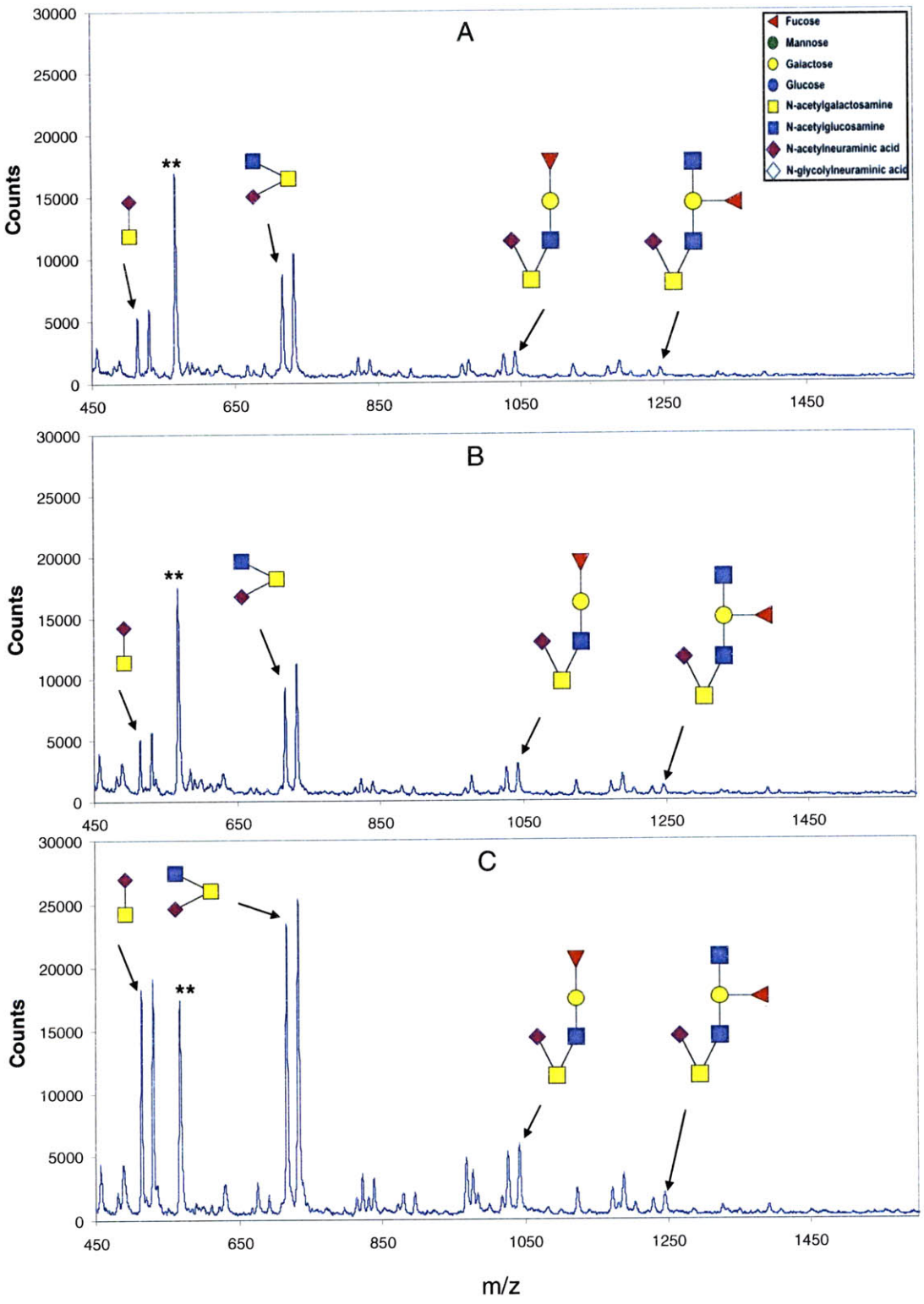
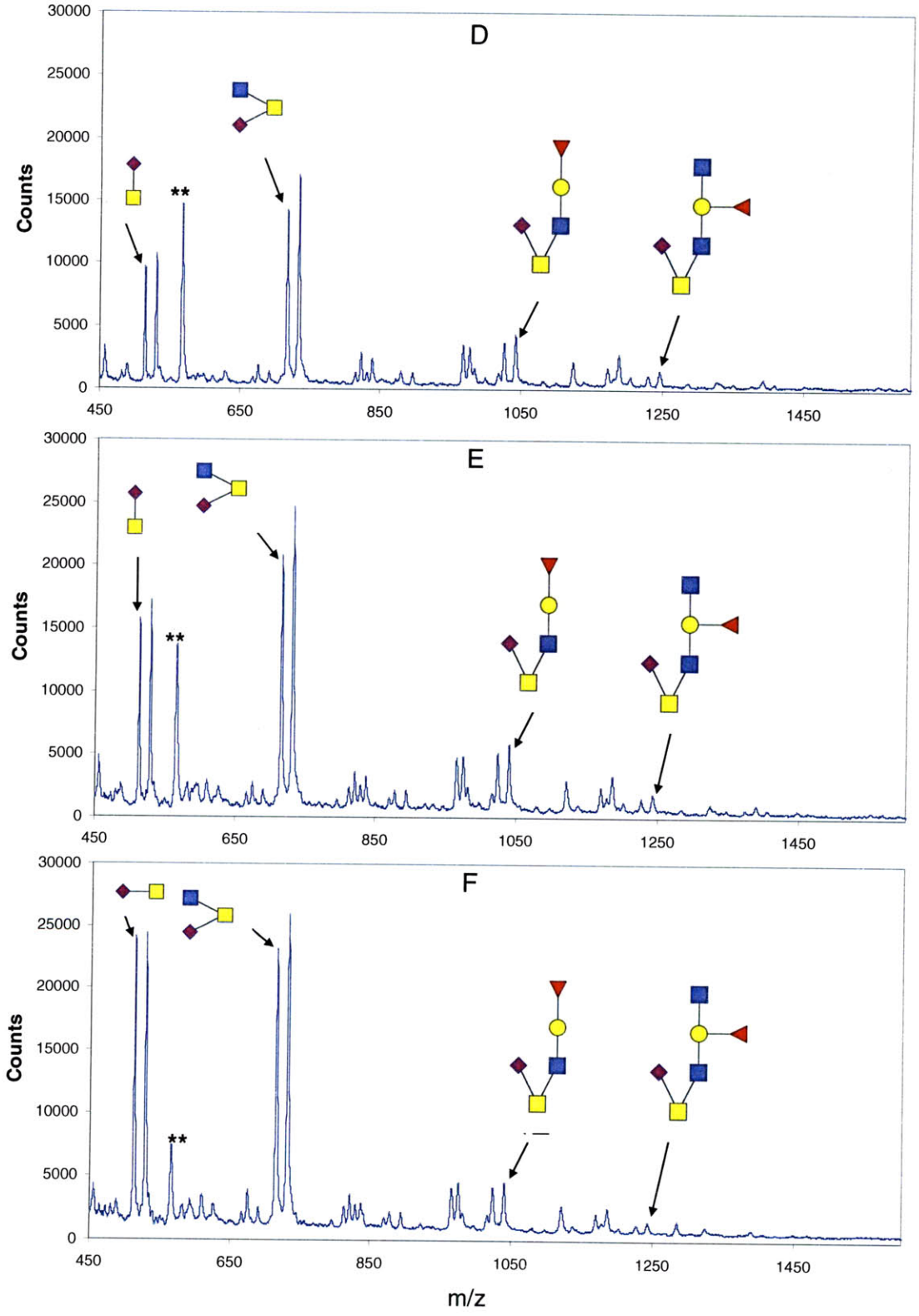


Figure 2.10. The Effect of β-elimination Incubation Length on the Bovine Mucin O-glycan Spectrum . A. 1hr B. 6hrs C. 12hrs D. 18hrs E. 24hrs F. 48hrs. ** = matrix peaks.


D
E
F
Counts
m/z
**
30000
25000
20000
15000
10000
5000
0
450
650
850
1050
1250
1450

For the Voyager analysis of bovine mucin, all previously published acidic O-glycans were evident as early as 1hr (Figure 2.10A, Table 2.7). More importantly, while other investigators fail to obtain every reported acidic glycan in each study, this study captures all of them in one spectrum. The glycan structures in this table were taken from four of the most cited papers involving bovine submaxillary mucin glycan analysis [7, 20-24]. This unprecedented display of all acidic glycans further validates the efficacy of the entire optimized method.

Figure 2.10 also reveals several additional unreported acidic glycan structures that have signal to noise ratios larger than 50. A list of the masses and their proposed structures are reported in Table 2.7. Most of these structures have masses larger than m/z 1243, and are consistently present in the spectra for every time point, even though the majority of these peaks have less than optimal resolutions. The presence of so many unreported acidic glycans provides further evidence that the optimized O-glycan analytical process is sound for acidic glycans, and can be used to determine new glycan species that are of interest to researchers.

Table 2.7 Reported Glycan Compositions for Bovine Mucin from Five Separate Studies. All reported glycan species are clearly resolved in the optimized glycan extraction method in this project. Furthermore, besides the study articulated in this project, there isn't any one study that has identified all reported bovine mucin glycan structures.

	Proposed Glycan Composition	Robbe et al, 2003	Savage et al, 2004	Chai et al, 1992, 1997	Novotny et al, 2002	Karlsson et al, 2001	Optimized Method
		MALDI-TOF	NMR	LSI-MS/ NMR	MALDI-TOF	MALDI-TOF	MALDI-TOF
513.45	NeuAcGalNAc-ol	Y		Y	Y	Y	Y
529.45	NeuGcGalNAc-ol	Y		Y	Y	Y	Y
674.00	$NeuAc_1$ Gal_1 GalNAc-ol		Y				Y
690.00	$NeuGc_1$ Gal_1 GalNAc-ol						Y
716.65	GlcNAc(NeuAc)GalNAc-ol	Y	Y	Y	Y	Y	Y
732.75	GlcNAc(NeuGc)GalNAc-ol	Y	Y	Y	Y	Y	Y
820.00	NeuAc-Gal-(Fuc)GalNAc-ol						Y
836.00	NeuGc-Gal-(Fuc)GalNAc-ol						Y
877.00	$NeuAc_1$ Gal_1 $GlcNAc_1$ GalNAc-ol	Y		Y			Y
894.00	$NeuGc_1$ Gal_1 $GlcNAc_1$ GalNAc-ol				Y		Y
965.00	$NeuAc_1$ Fuc_2 Gal_1 GalNAc-ol $NeuAc_2$ Gal_1 GalNAc-ol						Y
981.00	$NeuAc_1$ $NeuGc_1$ Gal_1 GalNAc-ol						Y
1015.00	$NeuGc_1$ Gal_3 $GalNAc_1$						Y
1024.00	$NeuAc_1$ Fuc_1 Hex_1 $HexNAc_2$	Y		Y			Y
1040.00	$NeuGc_1$ Fuc_1 Hex_1 $HexNAc_2$ $NeuAc_1$ Gal_2 $HexNAc_2$	Y		Y	Y		Y
1121.00	$NeuAc_1$ $HexNAc_4$						Y

1137.00	$NeuGc_1\ HexNAc_4$	[redacted]	[redacted]	[redacted]	Y	[redacted]	Y
1170.00	$NeuAc_1\ Fuc_2\ Hex_1\ HexNAc_2$ $NeuAc_2\ Hex_1\ HexNAc_2$	Y	[redacted]	Y	[redacted]	[redacted]	Y
1178.00	$NeuGc_1\ Hex_4\ HexNAc_1$	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Y
1186.00	$NeuGc_1\ Fuc_2\ Hex_1\ HexNAc_2$ $NeuGc_1\ NeuAc_1\ Hex_1\ HexNAc_2$	[redacted]	[redacted]	Y	Y	[redacted]	Y
1201.00	$NeuGc_2\ Hex_1\ HexNAc_2$	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Y
1227.00	$NeuAc_1\ NeuGc_1\ HexNAc_3$ $NeuAc_1\ Fuc_1\ Hex_1\ HexNAc_3$	Y	[redacted]	Y	[redacted]	[redacted]	Y
1243.00	$NeuGc_1\ NeuGc_1\ HexNAc_3$ $NeuGc_1\ Fuc_1\ Hex_1\ HexNAc_3$	[redacted]	[redacted]	Y	[redacted]	[redacted]	Y
1283.00	$NeuAc_1\ Hex_1\ HexNAc_4$	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Y
1324.00	$NeuGc_1\ Fuc_1\ Hex_4\ HexNAc_1$ $NeuAc_1\ Hex_5\ HexNAc_1$	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Y
1372.00	$NeuAc_2\ Hex_1\ HexNAc_3$	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Y
1389.00	$NeuAc_1\ Fuc_1\ Hex_2\ HexNAc_3$ $NeuGc_1\ NeuAc_1\ Hex_1\ HexNAc_3$	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Y
1431.00	$NeuGc_1\ NeuAc_1\ HexNAc_4$ $NeuAc_1\ Fuc_1\ Hex_1\ HexNAc_4$	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Y
1447.00	$NeuAc_1\ Hex_2\ HexNAc_4$ $NeuGc_1\ Fuc_1\ Hex_1\ HexNAc_4$	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Y

2.3.3.3 Sialic Acid Linkage Analysis Results

After having optimized the O-glycan extraction, purification, and MS analysis protocols, a sialic acid linkage analysis experiment was carried out for bovine mucin and fetuin in order to see if these optimized protocols could be used to learn more information about the glycan compositions of these species. In this experiment, unpurified O-glycans extracted from bovine fetuin and mucin were treated with either

Sialidase A, Sialidase S, or neither. Sialidase S only cleaves α2-3 and α2-8 linked terminal sialic acid moieties, while Sialidase A cleaves α2-3, α2-6, and α2-8 linked sialic acids. Comparing samples treated with both enzymes provides information about the relative abundance of each linkage. After the reaction, the purified mixture was analyzed in the negative mode to view the acidic glycans. The results of these experiments are shown in Figures 2.11 and 2.12.

From Figure 2.11, it is evident that bovine fetuin contains a mixture of α2-3 and α2-6 linked sialic acids, since Sialidase S treatment results in the disappearance of all of the original acidic glycan peaks, and the appearance of new acidic glycan peaks. More specifically, for O-glycans, Sialidase S treatment resulted in the removal of all O-glycan peaks, suggesting that these moieties all contain α2-3 linked sialic acids. On the other hand, treatment of the N-linked glycans with Sialidase S resulted in the disappearance of all original acidic glycan peaks, suggesting that every N-glycan contained at least one α2-3 linked sialic acid residue. Furthermore, the fact that several new acidic peaks (with high intensities) appear in the in the Sialidase S spectrum implies that several if not all of fetuin's acidic glycans also contain α2-6 linked sialic acids. As expected, Sialidase A treatment resulted in the complete disappearance of all acidic glycan peaks. In order to absolutely confirm that all of fetuin's N-glycans contain an α2-6 linked sialic acid, it will be necessary to analyze the reaction mixture in the neutral mode, and confirm that no neutral glycans appear as a result of the Sialidase S-induced removal of sialic acids, because these neutral peaks would suggest that some of fetuin's N-glycans do not in fact have α2-6 linked sialic acids.

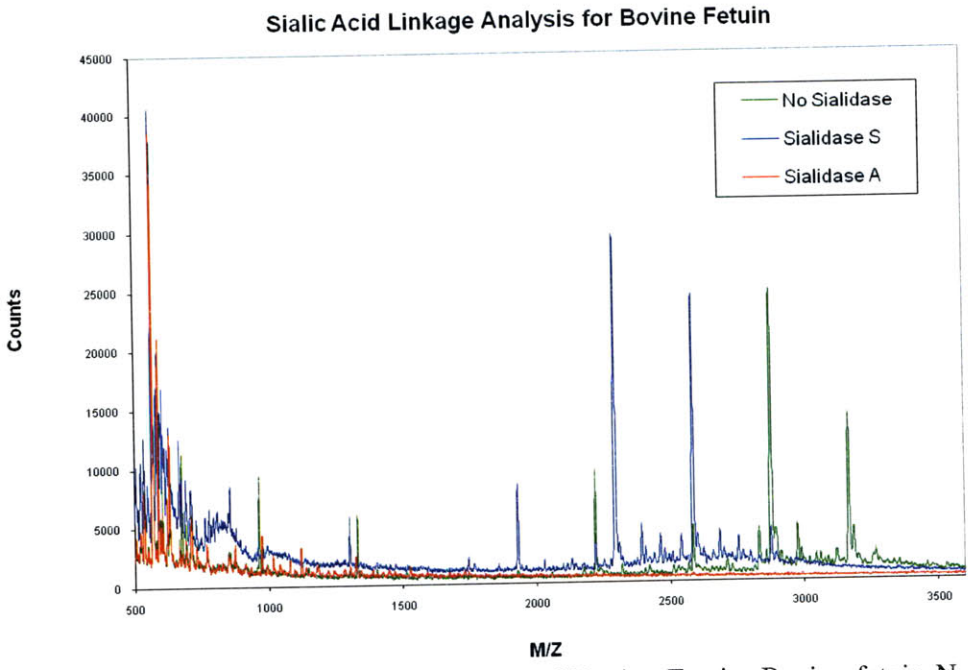


Figure 2.11. Sialic Acid Linkage Analysis of Bovine Fetuin. Bovine fetuin N- and O-glycans were released by beta elimination and subjected to Sialidase treatment. The reaction mixture was purified and the product was analyzed in the negative mode on the Voyager MALDI-TOF MS to view the acidic glycans.

Unlike that of bovine fetuin, the sialic acid linkage analysis for bovine mucin suggested that almost all of its glycans are only α2-6 sialylated, because the Sialidase S and untreated spectra almost completely overlap (Figure 2.12). If bovine mucin's O-glycans contained a mixture of α2-3 and α2-6 linked sialic acids, then most of the acidic glycan peaks would have shifted as they did in the case of bovine fetuin. A close-up of the overlap of the Sialidase A and S spectra is shown in the lower panel of Figure 2.12. As expected, the samples treated with Sialidase A showed no acidic glycans.

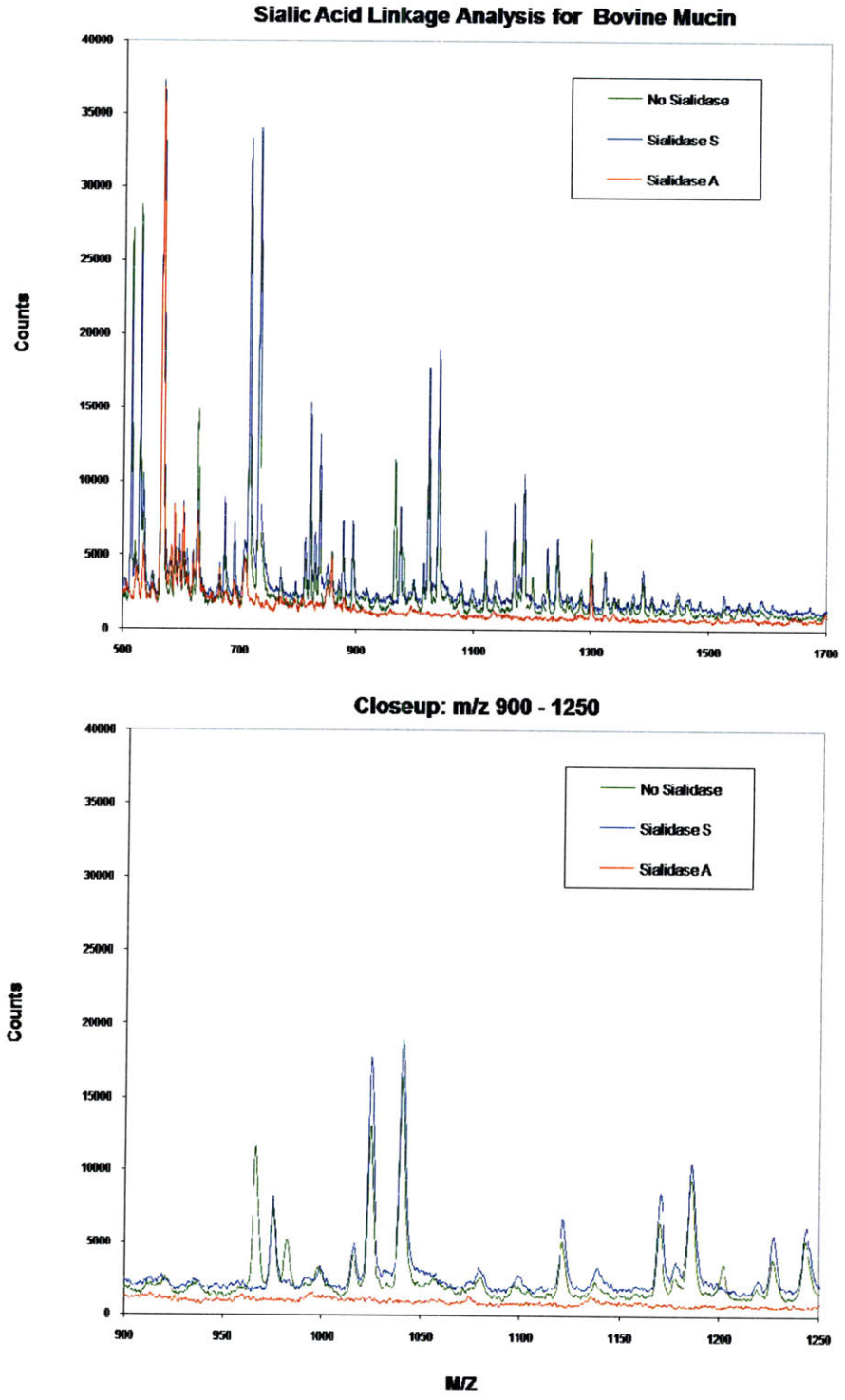


Figure 2.12. Sialic Acid Linkage Analysis of Bovine Mucin. Bovine mucin O-glycans were released by beta elimination and subjected to Sialidase treatment. The reaction mixture was purified and the product was analyzed in the negative mode on the Voyager MALDI-TOF MS to view the acidic glycans.

The success of the sialic acid linkage experiment further validates the optimized method. This set of experiment shows how modular the optimized O-glycan analysis scheme can be, since the exoglycosidase treatment was inserted in between the extraction and purification steps, and resulted in clean MALDI spectra. As expected, all Sialidase A treated samples showed no acidic glycans, and Sialidase S treatment was used to comment on the types of sialic acid linkages that were present on bovine mucin and fetuin. In the future, other types of assays (other exoglycosidase assays, derivatization reactions) may be inserted in between the modular steps of this procedure to achieve a desired goal, with the caveat that the researcher must understand all of the reaction reagents and how they might be processed at each step.

2.3.3.4 Conclusion

Overall, the length of incubation study validated the choice of reducing agent concentration and alkaline reagent by successfully revealing all major reported acidic glycans for fetuin and bovine submaxillary mucin. Furthermore, the ability of the optimized method to successfully identify several new glycans on bovine submaxillary mucin attests to its efficacy and versatility. In general, the data in this study suggested that proteins subject to beta-elimination for 12-25 hours will effectively result in the release of glycans and in spectra containing at least all prominent or reported O-glycans. Further optimization must be done if less prominent glycan residues are desired at specific intensities or resolutions. Unfortunately, this study also revealed that N-glycans are just as susceptible to cleavage as O-glycans during the beta-elimination process, even with a 1 hour incubation time. This fact will make the analysis of glycoproteins containing both N- and O- glycans more complex, but the structural composition of all peaks can eventually be scrutinized with additional MS procedures. Lastly, the sialic acid linkage analysis proved that this optimized method can be used to obtain useful information about the types of linkages and monosaccharides associated with a given glycan composition. In the future, this protocol can be used to research a number of biologically relevant O-glycosylated proteins for which the significance of the glycans is unknown or speculated to be important.

2.4 DISCUSSION

This project led to the successful optimization of an O-glycan analytical method that can be used to learn preliminary structural information about O-glycans. Furthermore, this procedure unintentionally showed how N-linked glycans could also be isolated, purified, and characterized. In this optimized method, glycosylated proteins are incubated in 1M KBH_4 in 0.1M KOH for 18 hours, after which the reaction is quenched with several acetic acid and methanol additions. Released O-glycans are then purified using a constructed microcolumn containing both SP20SS resin and Dowex Cation Exchange resin, which remove the protein and the alkali salt from the reaction mixture, respectively. This microcolumn formulation has been optimized to purify from 1ug to 125ug of proteins, and proves itself to be a robust method of purification following exoglycosidase digestions as well. Acidic glycans can then be analyzed with great efficiency using optimized Voyager and Bruker MALDI-TOF MS methods, with the former taking precedent in all experiments due to its selectivity to acidic ions and its lower O-glycan detection limit.

This project specifically showed how the optimized method could be used to obtain MALDI-MS spectra for bovine mucin and fetuin O-linked glycans, and then determine sialic acid linkage information about the glycans on these proteins. However, researchers can further exploit the benefits of this procedure by performing additional structural studies on the released glycans. If the O-glycans are present in a high enough abundance, NMR analysis can help by determining quantitative information about the types of moieties present on the batch of glycans as a whole (α2-3 vs. α2-6 linked sialic acids, lactosamine unit prevalence, etc). Permethylation of released glycans, on the other hand, can allow for the isolation and MS-MS analysis of these glycan species by HPLC and ESI-MS/MS, respectively. Lastly, further optimization of this method at the level of the glycan extraction could potentially allow researchers to label the released glycans with a chromophore or agent like biotin, so that these glycans can be separated by HPLC or adsorbed to microtiter plates in a glycan-array assay. All of these additional

steps can tremendously supplement the preliminary structural information obtained here, and will be essential if researchers hope to completely profile O-linked glycans.

Because fewer studies are carried out on O-linked glycans than to N-linked glycans, this optimized procedure has the potential to shed light on a lot of structural features of O-glycans which are presumed to be important in pathological and physiological processes. Moreover, it is the hope that this optimized procedure be eventually applied to biologically relevant human samples, such as the ascites from patients with various types of cancer, so that we can learn more information about the significance, if any, of these mucins on disease prognosis and progression. Another area of potential interest is the study of influenza binding, which is relevant to O-linked glycans because some subtypes of the influenza cell surface protein hemagglutinin preferentially bind the mucin-secreting goblet cells of the human upper respiratory tract. Thus, this analytical method could provide new and interesting information about the role of glycans in disease progression and pathogenesis.

2.5 REFERENCES

1.) Chang, W. C., L. C. Huang, et al. (2007). "Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited." Anal Chim Acta 582(1): 1-9.

2.) Harvey, D. J. (1999). "Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates." Mass Spectrom Rev 18(6): 349-450.

3.) "Mass Spetrometry". 20009. Friedrich-Schiller University. 9 May 2009 < www.med.uni-jena.de/vbmf/SELDI-MALDI/MStheory.doc>.

4.) "Accuracy & Resolution" www.matrixscience.com. 2007. Matrix Science. 9 May 2009. <http://www.matrixscience.com/help/mass_accuracy_help.html>.

5.) Busse, K., M. Averbeck, et al. (2006). "The signal-to-noise ratio as a measure of HA oligomer concentration: a MALDI-TOF MS study." Carbohydr Res 341(8): 1065-70.

6.) "MALDI-TOF Instruments" www.biocompare.com. 2009. Biocompare. 9 May 2009. <http://www.biocompare.com/Articles/BuyingTip/38/MALDI-TOF-Instruments.html>

7.) Huang, Y., T. Konse, et al. (2002). "Matrix-assisted laser desorption/ionization mass spectrometry compatible beta-elimination of O-linked oligosaccharides." Rapid Commun Mass Spectrom 16(12): 1199-204.

8.) Zhu, Y. F., C. N. Chung, et al. (1996). "The study of 2,3,4-trihydroxyacetophenone and 2,4,6-trihydroxyacetophenone as matrices for DNA detection in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry." Rapid Commun Mass Spectrom 10(3): 383-8.

9.) "Voyager Biospectrometry Workstation with Delayed Extraction Technology". Framingham: Perseptive Biosystems and PE Corporation. 1999.

10.) Robbe, C., C. Capon, et al. (2004). "Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract." Biochem J 384(Pt 2): 307-16.

11.)"Glycoprotein Fractionation Handbook" www. qiagen.com. 20005. Qproteome. 9 May 2009. <wolfson.huji.ac.il/purification/PDF/Lectins/QIAGEN_GlycoproteinFractionHandbook.pdf>

12.) "DIAION(R) & SEPABEADS(R) Synthetic Adsorbents"2000. www.diaion.com. Mitsubishi Chemical. 9 May 2009. <http://www.diaion.com/Index_E.htm>

13.) "Cation Exchange Resin, Dowex C-211, H+ Form" 2007 www.jtbaker.com. JT Baker Corporation. 9 May 2009. <http://www.jtbaker.com/msds/englishhtml/i6840.htm>

14.) Dwek, R. A., C. J. Edge, et al. (1993). "Analysis of glycoprotein-associated oligosaccharides." Annu Rev Biochem 62: 65-100.

15.) Hames, B.D. et.al. Gel electrophoresis of proteins: a practical approach Oxford: Oxford University Press. 1998.

16.) Edge, A. S. (2003). "Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: elucidation of molecular structure and function." Biochem J 376(Pt 2): 339-50.

17.) Beta-Elimination To Release O-Glycans Santosh Patnaik, Jμly 2005 http://stanxterm.aecom.yu.edu/wiki/index.php?page=Beta_-_elimination

18.) Shimamura, M., Y. Inoue, et al. (1984). "Reductive cleavage of Xaa-proline peptide bonds by mild alkaline borohydride treatment employed to release O-glycosidically linked carbohydrate units of glycoproteins." Arch Biochem Biophys 232(2): 699-706.

19.) Rademaker, G. J. et al (1996) "Mass Spectrometry: A Modern Approach to Solving Biological Structural Problems". Ph.D. Thesis, Utrecht University. 9 May 2009.

20.) Robbe, C., C. Capon, et al. (2003). "Microscale analysis of mucin-type O-glycans by a coordinated fluorophore-assisted carbohydrate electrophoresis and mass spectrometry approach." Electrophoresis 24(4): 611-21.

21.) Savage, A. V., C. M. Donoghue, et al. (1990). "Structure determination of five sialylated trisaccharides with core types 1, 3 or 5 isolated from bovine submaxillary mucin." Eur J Biochem 192(2): 427-32.

22.) Chai, W. G., E. F. Hounsell, et al. (1992). "Neutral oligosaccharides of bovine submaxillary mucin. A combined mass spectrometry and 1H-NMR study." Eur J Biochem 203(1-2): 257-68.

23.) Chai, W., T. Feizi, et al. (1997). "Nonreductive release of O-linked oligosaccharides from mucin glycoproteins for structure/function assignments as neoglycolipids: application in the detection of novel ligands for E-selectin." Glycobiology 7(6): 861-72.

24.) Karlsson, N. G. and N. H. Packer (2002). "Analysis of O-linked reducing oligosaccharides released by an in-line flow system." Anal Biochem 305(2): 173-85.

3 Mucin Direct Binding Assay Exploration

3.1 PROJECT MOTIVATION

While the O-glycan analytical method focused on the structural characterization O-linked glycans mucins, this project explores the potential benefits of a direct binding assay that would provide quantitative information on HA-mucin binding specificities. In the same way that synthetic glycan motifs were used to study N-glycan-HA binding, a direct binding assay that utilizes mucin as a "template" for large scale O-glycan presentation could be used to study mucin-HA binding. Furthermore, this "platform" would be more biologically relevant, since O-glycans that play a key role in binding are rarely presented as single, isolated species. This assay is more specific than the tissue binding assay because of the presence of characterized glycans, and also more relevant than the glycan array, since O-glycans are rarely presented in a biological setting as single isolated species. Previous research has suggested that certain HA isotypes preferentially bind to the mucin-rich goblet cells in the trachea and bronchia instead of the N-glycan rich ciliated epithelial cells. The potential significance of this preference is completely unknown, but it is hypothesized that this binding preference is due to a high affinity or multivalent interaction between the O-glycan mucin coatings on the goblet cells and the HA isotype.

Typically, binding assays using unmodified mucins have been used to study bacteria-mucin binding, since bacterial infection of host cells is thought to be prevented by binding to the mucosal layer [5]. Because the bacteria bind in a multivalent fashion to this mucin-filled protective layer, they are prevented from reaching the epithelial surface, where they can infect these cells and form colonies [5]. Furthermore, mucin-bacterial binding assays are used to study potential inhibitors of bacterial binding. Typically, microtiter wells are covered with mucin and left to incubate overnight either at 4C or 37C. Bacteria are subsequently added, and the wells are visualized for bacterial binding [5]. The assay developed in this protocol relies on the principles used in the bacteria-mucin binding assays, but is optimized to allow for the acquisition of quantitative binding information describing the interactions

between glycan binding proteins (lectins or HA) and glycoproteins like mucin. The experiments in this project are only meant to give an initial glimpse of the applicability of this tool. In the future, more studies would have to be performed to optimize the entire system to ensure accurate and reliable results.

In this project, two commonly used lectins were utilized. The elderberry lectin from Sambucus Nigra (SNA) binds the glycan motif Neu5Acα2-6Gal(NAc). This means that SNA commonly binds to α2-6 sialylated N-linked glycans because of the presence of the Neu5Acα2-6Gal motif at the end of glycan arms [9]. SNA can also bind to α2-6 sialylated GalNAc residues that are prominent on O-glycans, but only if these glycans are physically accessible to SNA [9]. Because of the dense population of O-glycans on mucins, SNA binding to these moieties may prove ineffective, especially if the sialylated species in question is the core GalNAc, but binding is still very much possible if the glycans contain other GalNAc residues that are α2-6 sialylated. Amylase inhibitor lectin (Jacalin) from Artocarpus integrifolia, on the other hand, is known to strongly bind the T-antigen (Galβ1-3GalNAcα1-Ser/Thr), which can be abundant on mucin O-glycans [9]. Jacalin only binds the sialylated form of the T-antigen (sialyl-T antigen) if the glycan is α2-3 sialylated. Jacalin also binds a variety of other glycan motifs containing Gal and GalNAc, but to a lesser extent. Therefore, according to these binding specificities, it is expected that SNA will strongly bind bovine fetuin, since it contains α2-6 sialylated Gal residues, and might bind to bovine mucin O-glycans, which may contain the motif Neu5Acα2-6Gal(NAc). However, Jacalin is expected to heavily bind to bovine mucin because of its abundance of GalNAc residues, but not as strongly to fetuin, which contains fewer glycans that are partially α2-3 sialylated.

3.2 MATERIALS AND METHODS

3.2.1 Optimization of Mucin Binding Conditions

Initially, a number of unmodified polystyrene plates were utilized, but were soon abandoned because of the difficulty of adsorbing the highly charged mucin to the plate. Maxisorp polystyrene immuno-plates were finally chosen because the surface of the plate is modified to have a high affinity for polar groups

[47]. Typically, each well was incubated with 100ul of a specific concentration of mucin (dissolved in 0.05M sodium carbonate buffer, pH 9.0), and incubated at 4C on a rocker overnight. Bovine fetuin was also used in the optimization process as an additional candidate glycoprotein because it has well-characterized N- and O-linked glycans. Wells were then washed four times with 200ul of TBS containing 0.05% Tween, after which 100ul of the appropriate concentration of FITC- conjugated lectin (diluted in TBS) was incubated in the wells at room temperature on a rocker for 3 hours. Wells were once again washed four times with 200ul of TBS containing 0.05% Tween, and then four times with just 200ul of TBS. The Anti-FITC antibody (in 100ul) was then added in a 1:200 dilution from the stock, and incubated in the wells at room temperature for 1 hour.

The quantitative aspect of the mucin direct binding assays involves the use of Fluorescein isothiocyanate (FITC)-conjugated lectins, which can be fluorometrically detected after the addition of a horse radish peroxidase (HRP) conjugated Anti-FITC antibody. HRP reacts with hydrogen peroxide (H_2O_2) and Amplex-Red Reagent in a 1:1 ratio to produce a brightly fluorescent product that has an excitation and emission wavelength of ~568 and 581 nm, respectively [46]. Typically, 8mL of DIW was combined with 2mL of the supplied 5X buffer, 8ul of H_2O_2, and 10ul of Amplex Red Reagent. Immediately prior to reading the plate, the wells were washed six times with 200ul of TBS, and then 100ul of the Amplex mixture was added. The fluorescent signal was detected on a Molecular Devices Spectra-Max 190 spectrophotometer instrument. The signal was recorded in intervals of 25 seconds for a total time of 20 minutes.

A number of controls were performed throughout this project to further validate all results. Wells with only glycoprotein, only lectin, only anti-FITC antibody, or only the Amplex mixture were treated according to the aforementioned protocol, and used to rule out the contribution of non-specific binding to all assay results.

3.2.2 Optimization of Mucin Concentration for Lectin Assays

The quantitative information obtained from these assays could ultimately be used to determine binding constants of proteins of interest such as lectins or HA. In order to obtain accurate constants, the mucin wells must be completely covered with mucin, such that an increase in the concentration of mucin results in no change in the fluorescent signal for the same concentration of FITC-conjugated protein. In order to determine the optimal mucin concentrations for this assay, wells were incubated with either 200ug/mL, 1000ug/mL, 2000ug/mL, 5,000ug/mL, or 10,000ug/mL bovine mucin, and treated according to the aforementioned protocol. Quantitative binding measurements were accomplished by using FITC-conjugated Sambucus nigra lectin (SNA) at a concentration of 1ug/mL. This concentration was initially chosen because it represents the concentration of SNA at which signal saturation occurs on the biotinylated-glycan array using long and short synthetic glycan arms that are α2-6 sialylated (6'SLN and 6'SLN-LN).

3.2.3 Lectin Dose Response

Because quantitative binding differences are calculated by comparing binding constants (K_d), it was essential to standardize the protocol for obtaining the binding constant on the mucin array. For this experiment, 1mg/mL bovine mucin and 200ug/mL bovine fetuin were adsorbed to the wells. The aforementioned washing and antibody/lectin binding protocols were used, except wells were incubated with lectins at concentrations of 0.005ug/mL, 0.01ug/mL, 0.05ug/mL, 0.1ug/mL, 0.5ug/mL, 1ug/mL, 2ug/mL, 5ug/mL, 10ug/mL, 20ug/mL, and 50ug/mL. These concentrations were chosen because they represent a 3 fold difference in the lectin concentrations, and would therefore provide reliable information about the binding constants. The lectins that were used for this assay were SNA and Jacalin, which binds strongly to a variety of glycan motifs containing GalNAc residues.

3.2.4 Lectin Binding Specificity Assays

Since binding specificity assays are done using exoglycosidases on the bound mucin, the whole plate has to be incubated at 37C for at least 16 hours. Therefore, it is necessary to determine if such a long

incubation time at 37C affects the amount of mucin bound to the plate, and ultimately the fluorescent signal obtained. For this assay, 2mg/mL mucin was adsorbed to the plate, and following washing, the wells were subsequently incubated for 18 hours at 37C to simulate the incubation phase with exoglycosidases. The plate was then treated using the aforementioned general protocols with either 1ug/mL SNA, 1ug/mL Jacalin, or 10ug/mL Jacalin.

In order to optimize an experimental protocol for determining the sialic acid lectin binding specificities, mucin that was adsorbed to the wells at a concentration of 2mg/mL, and treated either without Sialidase, with Sialidase S, or with Sialidase A. The plate was incubated for 18-24 hours at either room temperature or 37C. Both temperatures were chosen to observe the effects, if any, of temperature on the ability of the mucin to stay adsorbed to the plate, since the mucins were initially adsorbed at 4C. The rest of the procedure was carried out according to the aforementioned general protocol, with the wells being incubated in 1ug/mL SNA for 3 hours are room temperature.

3.3 RESULTS

3.3.1 Optimization of Mucin Concentration for Lectin Assays

In order to determine the proper concentration of mucin with which to coat the microplate wells, a mucin dose response assay was performed. In this assay, different concentrations of mucin were incubated in the wells and exposed to 1ug/mL SNA in order to determine the minimum concentration at which the fluorescent signal was most intense. SNA was the preferred lectin for this assay because it binds specifically to Neu5Acα2-6Gal(NAc) residues on either N- or O-linked glycans, and it was suspected that bovine mucin has only α-2-6 linked sialic acid residues on its glycans (from the research detailed in Chapter 2). Figure 3.1 shows the initial slopes for each binding curve (associated with a specific mucin concentration). From the data in Figure 3.1, signal saturation (with an intensity of ~ 1.5 x 10^6) begins at a bovine mucin concentration of 2mg/mL. This concentration suggests that the majority of the well is covered with mucin. Therefore, this mucin concentration was determined to be suitable for all

future mucin binding assays. It is important to note while a concentration of 2mg/mL seems quite high for any biological protein, 70-80% of that weight is attributed to the carbohydrates on the mucin. Furthermore, mucins are thought to have a molecular weight above 1 million Daltons, such that the actually number of moles of mucin being added to the wells is much lower than it appears. Also important to note in the results is the fact that control wells containing only mucin, only SNA, or only anti-FITC showed insignificant background signal (data not shown).

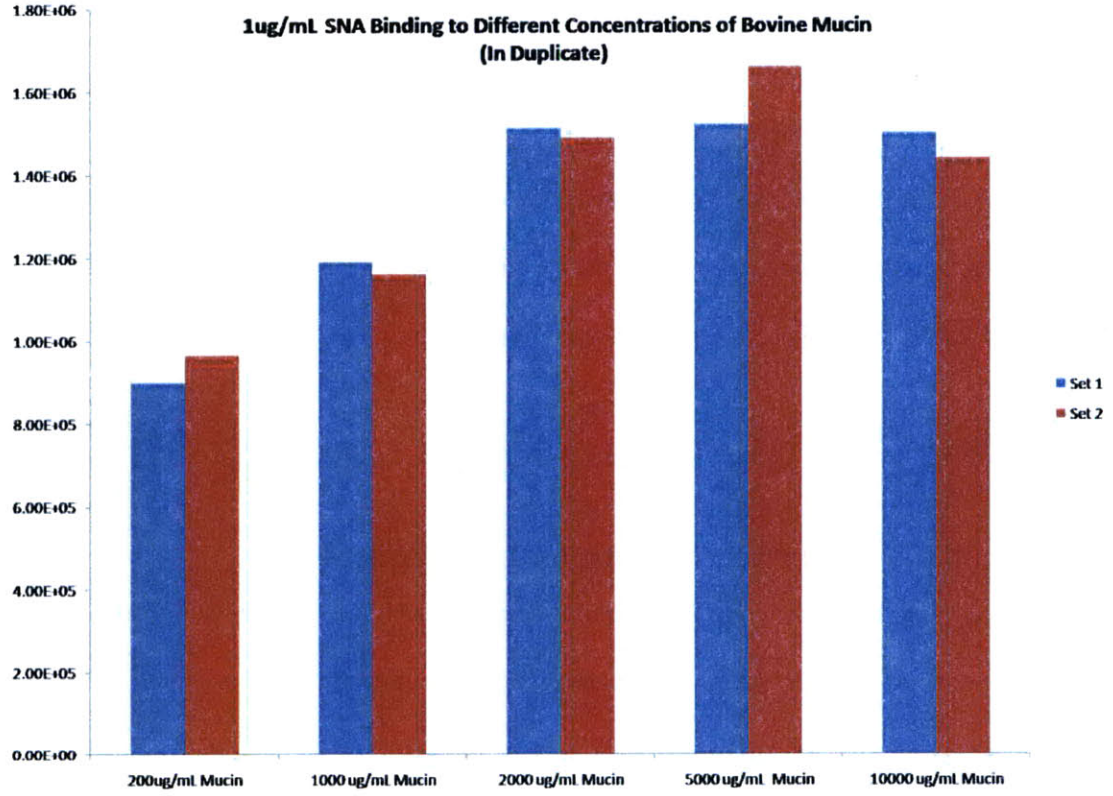


Figure 3.1. Optimization of Mucin Concentration for Lectin Assays. Different concentrations of mucin were incubated in the wells, which were then subject to 1ug/mL SNA incubation for 3 hours at room temperature. The experiments were done in duplicate (Set 1 and Set 2).

3.3.2 Lectin Dose Response Assay Results

In order to explore the potential quantitative results offered by the mucin direct binding assay, a lectin dose-response assay was performed. Ideally, the trends seen in this assay should be comparable to those

seen on the glycan array. The initial flourescent signal per time should be highest for the highest lectin concentration, and lowest for the lowest lectin concentration. This trend ultimately proves that binding is specific and due to the presence of the adsorbed mucin. In the first dose response experiment, 8 concentrations of SNA and Jacalin were applied to wells containing 200 ug/mL bovine fetuin. This fetuin concentration was chosen because of previously optimized experiments in the lab that yeilded successful SNA binding results with 200ug/mL fetuin (data not shown). Jacalin was chosen because it shows a high degree of specificity towards any motifs containing GalNAc residues, which includes the O-linked glycans of bovine fetuin. From the sialic acid linkage experiments detailed in Chapter 2, it is evident that bovine fetuin contains a mixture of α2-3 and α2-6 linked sialic acids on its N-glycans. Furthermore, the N-glycans of fetuin have been extensively characterized, and these glycans appear to be short and branched, such that the results of fetuin binding experiments can be compared to the glycan array results for short glycan arms containing α2-3 or α2-6 linked sialic acids. These assumptions are also valid for fetuin's O-linked glycans, which also appear to have short arms. Therefore, the lectin dose response data for bovine fetuin will provide some insight on the ability of lectins like SNA and Jacalin to bind to N- and O-linked glycans with short α2-6 sialylated arms.

Figure 3.2 shows the initial slopes for each flourescence curve associated with a specific concentration of SNA and Jacalin. SNA appeaars to saturate bovine fetuin at a concentration of ~ 2ug/mL, while Jacalin binding saturates fetuin at a concentration of ~ 0.5ug/mL – 1ug/mL. The saturation concetration data for SNA is not too far off from the 1ug/mL saturation concentration obtained from previous glycan array experiments using short glycan chains with using α2-6 linked sialic acids (6'SLN). Jacalin's saturating concentration, on the other hand, is less than half of that of SNA, and is probably due to the presence of fetuin's O-linked glycans. Furthermore, previously performed tissue binding data (data not shown) showed that Jacalin didn't bind to the ciliated bronchial and tracheal epithelial cells (which contain only N-linked glycans), but did strongly bind to the mucin secreting goblet cells, which are covered in motfis containing GalNAc residues.

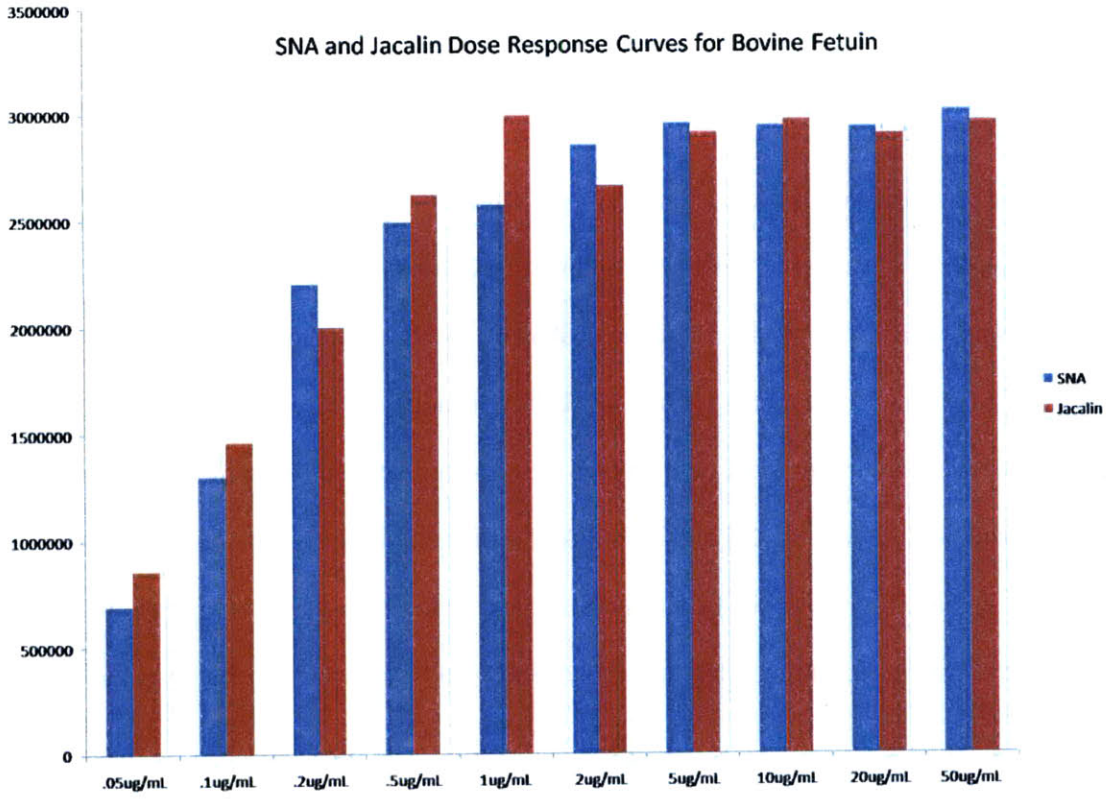


Figure 3.2. SNA and Jacalin Dose Response Binding Assay on Bovine Fetuin. Ten concentrations of SNA or Jacalin were incubated in the wells containing 200ug/mL adsorbed fetuin.

Figure 3.3 shows the dose response data plotted in the form of the linearized Hill equation (Equation 2, Section 1.1.4.3). The value of "y", the fractional saturation, was taken to be the quotient of the initial slope at a specific lectin concentration and the maximum initial slope recorded for that lectin. These "maximum" values are 2.95×10^6 and 2.99×10^6 for SNA and Jacalin, respectively. A linear trendline was created using the data points up until the maximum slope value, and the equations and R^2 values for the linear trendlines are shown. The calculated K_d values for SNA and Jacalin (for bovine fetuin) are shown in Table 3.1. The K_d value calculated for SNA ($\sim 10^{-11}$) was actually in the same order of magnitude as that calculated for 6'SLN on the glycan array. The K_d for Jacalin was also on the order of 10^{-11}, suggesting that Jacalin binds strongly to the three O-glycan residues present on bovine fetuin. One could verify that the K_d was solely due to the presence of O-glycans by exposing the adsorbed fetuin to

PNGase F, followed by an SNA incubation. In this case, the K_d value should not change, since PNGase F only removes the N-linked glycans, which Jacalin does not bind. In this table no data is available on the K_d for Jacalin bound to 6'SLN because Jacalin isn't expected to bind to any glycans that contain GlcNAc residues. Therefore, in order to obtain an analgous K_d value for Jacalin on the synthetic glycan array, biotinylated synthetic glycans analogous to those present on bovine mucin should be used.

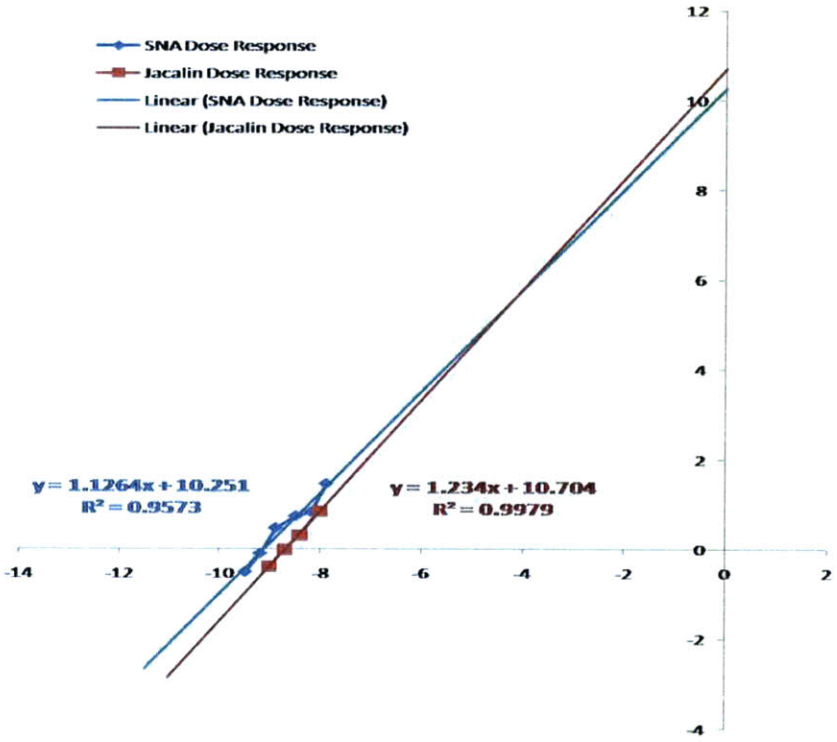


Figure 3.3. Linearized Hill Equations using Fractional Saturation Values for SNA and Jacalin on Bovine Fetuin. Typically, only the data points before that of the maximal fractional saturation were used. These values were linearized and plotted. The concentrations of SNA and Jacalin were converted to the proper units using molecular weights of 150,000 Da and 50,000 Da, respectively. A linear trendline was created using these data points, and the y-intercept was used to calculate the K_d.

Table 3.1. Dissociation Constant Calculated for SNA and Jacalin on both Fetuin and Synthetic Glycan Arrays.

Lectin on 200ug/mL Bovine Fetuin	**Saturating Concentration**	**K_d on fetuin array**	**K_d on glycan array using 6'SL (for SNA)**
SNA	2ug/mL	5.61×10^{-11}	$\sim 10^{-11}$
Jacalin	0.5ug/mL – 1ug/mL	1.98×10^{-11}	No data available

In the second does response experiment, the same conenctrations of Jacalin and SNA were applied to wells containing 2mg/mL mucin. Figure 3.4 shows the initial slopes for each binding curve associated with a specific concentration of SNA and Jacalin. SNA saturates bovine mucin at a concentration of ~ 20ug/mL, while Jacalin saturates at a concentration of ~ 0.5ug/mL. Unlike the saturation concentration for bovine fetuin, that for SNA on bovine mucin is much higher than the 1ug/mL saturation concentration obtained from the glycan array. This high saturation concentration suggests a much weaker interaction between SNA and bovine mucin, possibly because of the presence of densly packed short glycan arms with α2-6 sialic acids, to which SNA may not be able to bind because of steric reasons. Furthermore, a large number of bovine mucin glycans could be α2-6 sialylated on GlcNAc residues instead of Gal or GalNAc residues. If this is the case, SNA will not bind to these glycans. On the other hand, the saturating concentration for Jacalin was much lower, presumably because Jacalin binds to a wide vareity of glycan moeities containing GalNac residues, which are an abundant on bovine mucin. This data affirms the results of previous tissue binding experiments, which showed that Jacalin heavily binds to the surface of mucin secreting goblet cells, but minimally binds to the N-glycan containing epithelial cell surfaces.

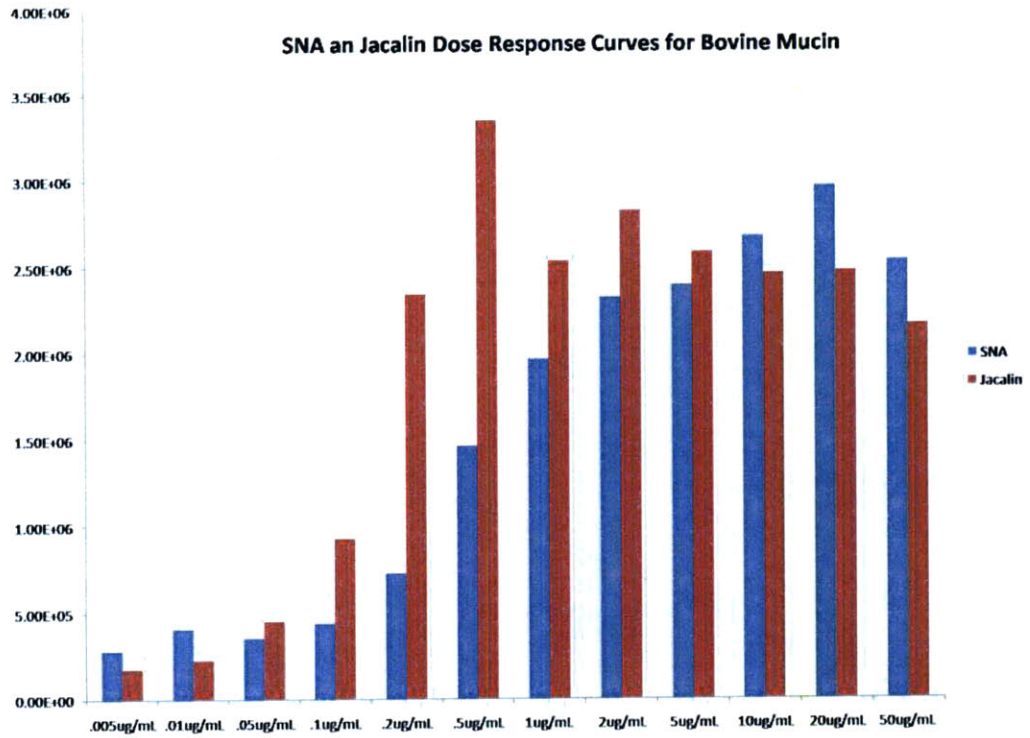


Figure 3.4. SNA and Jacalin Dose Response Binding Assay on Bovine Mucin. Ten concentrations of SNA or Jacalin were incubated in the wells containing 200ug/mL adsorbed fetuin.

Figure 3.5 shows the dose response data from bovine mucin plotted in the form of the linearized Hill equation. The value of "y" was taken to be 2.96×10^6 and 3.34×10^6 for SNA and Jacalin, respectively. The calculated K_d values for SNA and Jacalin are shown in Table 3.2. The K_d value for SNA was much lower than the same value for bovine fetuin, and this result supports the theory that either SNA cannot bind to a densly packed short O-glycan network or doesn't bind to α2-6 sialylated GlcNAc residues that could be abundant on bovine mucin O-glycans. As expected, the K_d for Jacalin was much lower (on the order of 10^{-17}), confirming the theory that Jacalin binds strongly to mucin. Also interesting is the fact that this K_d was several order of magnitude lower than the K_d for Jacalin on bovine fetuin. This disparity can be explained by the fact that fetuin has significantly fewer O-glycans, and subsequently, fewer GalNAc residues.

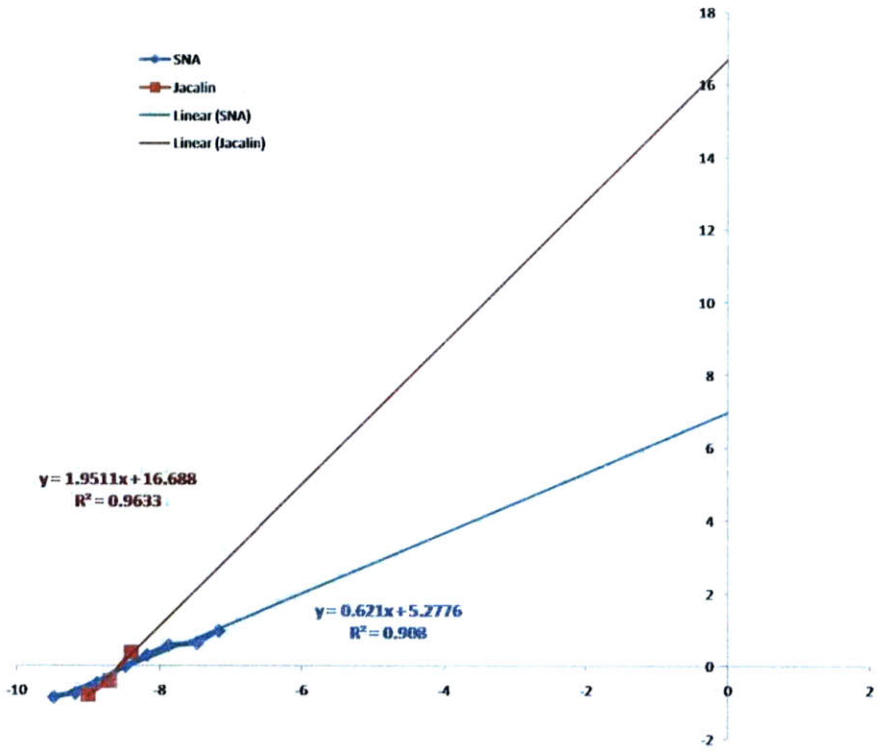


Figure 3.5. Linearized Hill Equations for Fractional Saturation Values of SNA and Jacalin Bound to Bovine Mucin. Typically, only the data points before that of the maximal fractional saturation were used. These values were linearized and plotted. The concentrations of SNA and Jacalin were converted to the proper units using molecular weights of 150,000 Da and 50,000 Da, respectively. A linear trendline was created using these data points, and the y-intercept was used to calculate the K_d.

Table 3.2. Dissociation Constant Calculated for SNA and Jacalin on Bovine Mucin and Synthetic Glycan Arrays.

Lectin on 200ug/mL Bovine Fetuin	Saturating Concentration	K_d on fetuin array
SNA	10 ug/mL	55.28×10^{-6}
Jacalin	0.5ug/mL	2.05×10^{-17}

The preliminary results for SNA and Jacalin on bovine mucin and fetuin therefore suggest that the glycoprotein direct binding assay can potentially be used to calculate binding constants for glycan-binding proteins, and that the binding constants will at least make sense in the context of the specific

glycoporotein target. However, additional optimization must be performed with these assays to ensure repeatable and reliable results.

3.3.3 Sialic Acid Specificity Assay Results

The glycoprotein direct binding tool can potentially be used to learn information about the critical monosaccharide components responsible for lectin or HA specificity. In order to explore this possibility, it was first necessary to determine the effects, if any, of incubation at 37C on the ability of the glycoprotein to stay adsorbed to the plate. Therefore, a temperature-specific assay was performed in which plates adsorbed with mucin were incubated at either room temperature or 37C for 18 hours (to mimic exoglycosidase experiments) prior to SNA incubation. From the data in Figure 3.6, it is apparent that an incubation at 37C has little to no effect on the signal obtained from a well incubated with SNA, but it has a slightly larger effect on the signal from a well incubated with 1ug/mL Jacalin because the signal decreases by 33%. It is important to note that the saturating concentration of SNA was not used in this experiment, and therefore, a dependence on temperature might be observed at this concentration. Regardless, useful and informative information can still be obtained from experiments with these conditions if the proper untreated control is performed alongside treated wells. Overall, this data is very useful because it illustrates the importance of doing a control at every temperature at which the wells are exposed.

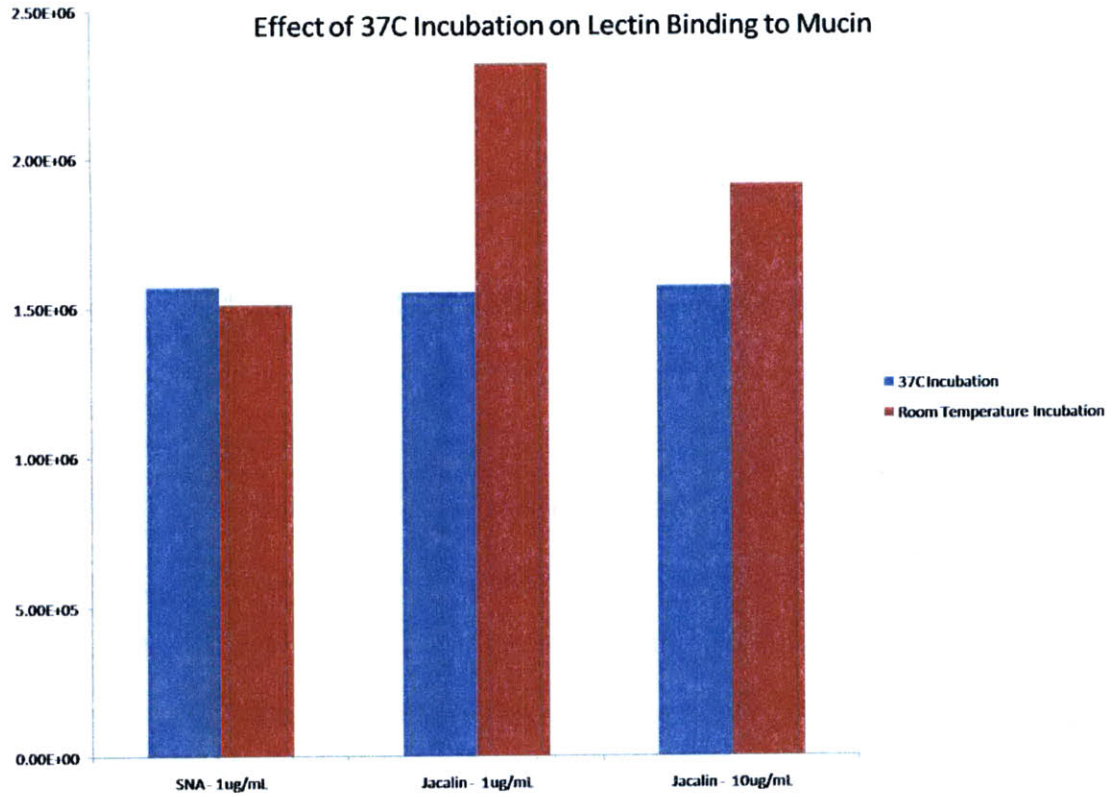


Figure 3.6. Effect of 37C Incubation on Lectin Binding to Mucin. In this assay, wells plated with either 2mg/mL mucin and exposed either to an 18 hour incubation at room temperature or 37C (with no lectin). After this period, the relevant concentrations of lectins were added to the wells for a 3 hour incubation period at room temperature. Concentrations of 1 ug/mL and 10 ug/mL were chosen for Jacalin to make sure that any temperature specific trends occurred regardless of lectin concentration.

Finally, binding experiments using SNA were performed on Sialidase A and S treated mucin previously adsorbed to the wells in order to both confirm reported sialic acid binding specificities, and further validate the assay. Sialidase treatment was applied to previously adsorbed mucin or fetuin to ensure that the same amount of material was bound to the wells, and that the binding properties of the mucin were not altered for each specific Sialidase treatment. Furthermore, Sialidase treatment on bound glycoprotein suggests the removal of only those residues that are exposed and available for binding. The results of this assay are shown in Figure 3.7. As expected, Sialidase A treated bovine fetuin showed little SNA binding because SNA binding requires the presence of α2-6 linked sialic acid. Conversely, Sialidase S treated bovine fetuin showed binding signals similar to that of untreated samples, confirming the fact

that fetuin has at least some α2-6 linked sialic acid residues. Furthermore, the fact that the signals are close in intensity for these two samples suggests that α2-6 sialylation is prominent in bovine fetuin glycans, since evidence for the contrary would show reduced SNA binding signals for Sialidase S treated wells.

Similarly, the results for bovine mucin show that Sialidase A treatment reduces SNA binding, while Sialidase S has no effect. Once again, these results further validate the results of the glycan linkage analysis detailed in Chapter 2, which suggest that bovine mucin contains glycans with only α2-6 linked sialic acids, since Sialidase S treatment (which removes only α2-3 linked sialic acids), does not affect SNA binding. The one major difference between the results seen here and those for bovine fetuin is that the overall affect of Sialidase A treatment on SNA binding is less dramatic; the SNA binding signal only decreased by ~ 45% for Sialidase A treated bovine mucin, while that for Sialidase A treated fetuin decreased by more than 95%. These differences could arise from the fact that there is significantly more (sialylated) glycan in bovine mucin plated wells, such that an 18 hour incubation with 10ul of Sialidase A may have been insufficient for the removal of most of the sialic acid. In the future, this theory can be tested either by adding more Sialidase A and keeping the same incubation time, or by using the same amount of Sialidase A and increasing the incubation time. The former option seems more reasonable, since the latter requires that the incubation times be optimized, and that the appropriate set of temperature and incubation time specific controls are performed. Lastly, it is important to note that this experiment was performed using 1ug/mL SNA instead of the saturating concentration of 20ug/mL. In the future, this assay should be repeated with the saturating concentration to see the maximum effect of the Sialidase treatment.

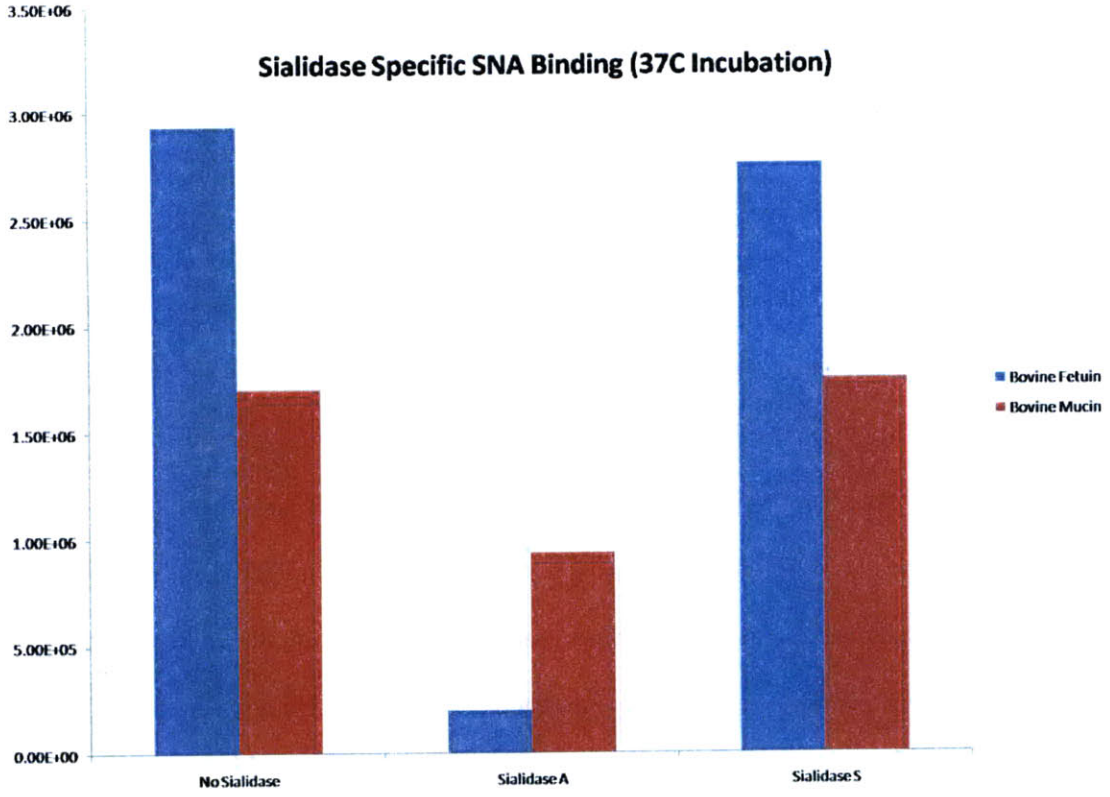


Figure 3.7. Sialic Acid Binding Specificity for SNA Bound to Fetuin and Mucin. In this assay, wells plated with either 2mg/mL mucin or 200ug/m fetuin were incubated with either no Sialidase, Sialidase A, or Sialidase S and left to incubate for 18 hours at 37C. The wells were subsequently incubated with 1ug/mL SNA and left to incubate for 3 hours at room temperature.

3.4 DISCUSSION

In this Project, a quantitative binding assay to study lectin binding to mucin or glycoproteins was explored. Bovine fetuin and mucin were effectively bound to microtiter plates, and lectins like SNA and Jacalin successfully bound to these glycans in a dose dependent manner. These results suggest that other glycan-binding proteins of interest may be studied in a quantitative manner with the actual glycoproteins that they are suspected to bind. These adsorbed glycoproteins represent a more natural substrate for the glycan-binding protein, especially in the case of mucins, since synthetic O-glycan arrays would fail to recapitulate the densely packed O-glycan networks that mucins present. Another benefit of this array is

that cell surface binding can be simulated with these bound glycoproteins and studied through the application of the relevant glycan-binding protein. This is particularly relevant in the case of HA binding to cell surface glycoproteins, since up until now, studies have only focused on HA binding to individual glycan motifs and not individual glycoprotein species. More specifically, this assay could provide insight on the significance, if any, of HA-mucin binding, since it is unclear why certain subtypes of HA preferentially bind mucin rich goblet cells instead of the ciliated epithelial cells of the human upper respiratory tract.

Furthermore, the sialic acid specificity experiments associated with this direct binding assay potentially allow researchers to uncover critical binding specificity information. In the experiments in this project, only Sialidase treatment was used as a potential mechanism for uncovering binding specificities of glycan-binding proteins. This information is especially critical in the efforts to understand HA binding and transmission efficiency, since the sialic acid residues are needed for the initial viral binding event. However, this direct binding assay can also be used to look at other kinds of binding specificities. Different exoglycosidases and endoglycosidases could potentially be used to discover information about binding specificities relating to linkage, monosaccharide, and even conformation types.

Overall, this direct binding assay could provide some insight into the binding specificities and quantitative binding information associated with glycan-binding proteins and their natural glycoprotein substrates. This assay is probably not well suited for studies which seek to directly compare the binding of one protein to two or more substrates because of the multitude of other factors that will contribute to the quantitative binding results, such as the number and spacing of the glycans. It can, however, be used to study the binding of two or more proteins to one specific substrate, and is therefore well suited for influenza studies which seek to determine quantitative binding differences between two or more HA subtypes. When used in conjunction with glycan arrays and glycan characterization experiments, a more complete picture of the binding dynamics can be obtained, and researchers can specifically focus on

therapeutic and vaccine targets with particular characteristics, such as glycan topologies and linkage types.

3.5 REFERENCES

1.) Gendler, S. J. and A. P. Spicer (1995). "Epithelial mucin genes." Annu Rev Physiol 57: 607-34.

2.) Bowen, R. "Goblet Cells" 1998. Biomedical Hypertexts. 9 May 2009. http://www.vivo.colostate.edu/hbooks/pathphys/misc_topics/goblets.html

3.) "Human Trachea Epithelia Image". 2009. University of Washington. 9 May 2009. <http://www.lab.anhb.uwa.edu.au/mb140/CorePages/Epithelia/Images/trachea041he.jpg>

4.) Srinivasan, A., K. Viswanathan, et al. (2008). "Quantitative biochemical rationale for differences in transmissibility of 1918 pandemic influenza A viruses." Proc Natl Acad Sci U S A 105(8): 2800-5.

5.) Mucin-Bacterial Binding Assays Nancy A. McNamar[a2], Robert A. Sac[k3] and Suzanne M. J. Fleiszi[g2]

6.) McNamara, N. A., R. A. Sack, et al. (2000). "Mucin-bacterial binding assays." Methods Mol Biol 125: 429-37

7.) "Specialised Surfaces" 2009 www.nuncbrand.com. Nunc. 9 May 2009. <http://www.nuncbrand.com/us/page.aspx?id=1275#Subsection2>

8.) "Amplex Red Reagent"2009. www.invitrogen.com Invitrogen. 9 May 2009. <http://probes.invitrogen.com/media/pis/mp33851.pdf>

9.) "Glycan Array: Primary Screen Data" www.functionalglycomics.org 2009 The Consortium for Functional Glycomics. 9 May 2009. <http://www.functionalglycomics.org/glycomics/publicdata/selectedScreens.jsp>.

Chapter 4

4 Cell Surface Glycan Characterization

4.1 PROJECT MOTIVATION

The purpose of the cell surface glycan characterization project is to compare and contrast the glycan species present on the cell surfaces of chicken/turkey erythrocytes and human tracheal epithelial cells in order to draw conclusions about the ability of the agglutination assay to capture the HA-glycan interactions that are key determinants of virus adaptation. As stated in the Introduction, erythrocytes represent the only model system for influenza binding and inhibition studies, and even though this system is widely used, it has not been investigated for its relevance to the HA-glycan binding interactions that are key in human influenza infection. This means that vaccines or inhibitors that are identified using these assays may show decreased binding specificities in humans because the chicken, turkey, horse, and guinea pig agglutination assays fail to fully recapitulate the key glycan moieties uniquely implicated in human influenza infection. Cell surface glycan characterization was previously performed only on human bronchial epithelial cells, which are present in the lower part of the upper respiratory tract. Therefore, characterization of the glycans present on the surface of human tracheal epithelial cells, which are more likely to be exposed to the influenza virus due to their location in the upper airway, will supplement the information obtained for human bronchial epithelial cells. Only chicken and turkey erythrocytes were characterized in this project because of the wide-spread availability of both, but in the future, it will be necessary to characterize all of the currently utilized erythrocytes to provide a more complete understanding of the relevance of these cell types. In this project, glycan profiling entails the isolation, purification, and MALDI-MS analysis of cell surface glycans, followed by sialic acid linkage analysis of the acidic glycans for each cell type in order to determine the abundance of α2-3 to α2-6 linked sialic

acids for each cell type. All of these experiments represent the first steps in profiling the glycans present on the surfaces of these cells and on chicken and turkey erythrocytes. These characterization methods were used to provide a preliminary profile of the relevant cell surface N-glycans, and the results of this project will provide the platform on which additional studies can be designed to specifically study the glycan topologies (glycan chain length, extent of fucosylation/branching, etc) present on the cell surfaces of each of these cell types.

4.2 MATERIALS AND METHODS

4.2.1 Red Blood Cell Surface Glycan Extraction

Glycans were extracted from the surface of chicken and turkey red blood cells according to a modified version of a previously published protocol [5]. RBC stocks were diluted in PBS to obtain an approximate concentration of 400 million cells per mL. Thirteen aliquots of 1mL were utilized for the each cell type. The following steps were then repeated twice; cells were spun down at 2,000g for 10 minutes at 4C, the supernatant was aspirated, and the pellet was re-suspended in 0.5mL of PBS + 1% protease inhibitor. Cells were then lysed for 15 minutes under gentle agitation at room temperature in 0.5mL of deionized water (DIW) containing 1% protease inhibitor. The suspension was then spun down at 2000g for 10 minutes at 4C, and then resuspended in 0.5mL of PBS + 1% protease inhibitor after the supernatant was removed. An additional spin down cycle was performed using the same buffer volume at 4,000g for 10 minutes at 4C. The supernatant was removed, and the pellet was resuspended in 20ul of DIW and 230ul of an aqueous solution of 1% SDS + 20mM 2-Mercaptoethanol. The suspended pellets were boiled in a hot water bath for 10 minutes, after which 30ul of 10% NP40 (PNGase F Kit), 40ul of G7 Buffer (PNGase F Kit), and 3ul of PNGase F (glycerol free) [Prozyme] were added to the mixture. The pellets were incubated for 24 hours at 37C under gentle agitation. After incubation, 60ul of Calbiosorb beads were added to the mixture to remove SDS, and this mixture was incubated for 15 minutes under gentle agitation at room temperature. At this point, the sample was ready for purification.

4.2.2 Epithelial Cell Surface Glycan Extraction

Glycans were extracted from the surface of human tracheal epithelial cells according to a modified version of a previously published protocol [5]. HTE cells were grown to confluency in a 1:9 mixture of fetal bovine serum and MEM media (Dibco), and were not harvested until at least the 3rd passage. Typically, four large culture plates (500cm^2 surface area) were used for each assay. The cells were removed from the surface of the plate using 50mL of sterile 100mM sodium citrate buffer. Cells were incubated for 30 minutes at 37C, at which point a thin cell layer was readily peeling off of the culture plate surface. The plate surface was washed multiple times in the same buffer to ensure that all cells were removed. The suspension was then gently mixed several times to de-clump the cells in preparation for cell counting. The suspension was transferred to a 50mL Falcon tube, sampled to check the cell viability, and then spun down at 2000g for 10 minutes at 4C. The following steps were then repeated twice; cells were spun down at 2,000g for 10 minutes at 4C, the supernatant was aspirated, and the pellet was re-suspended in 20mL of PBS + 1% protease inhibitor. For each step the cell viability was maintained over 80%. Cells were then lysed with ~ 15 strokes of a tissue homogenizer. Cell viability was no higher than 35%. The suspension was then transferred to a new Falcon tube and spun down at 4000g for 10 minutes at 4C, and then resuspended in 1mL of PBS + 1% protease inhibitor. This suspension was transferred to an Eppendorf tube, and spun down at 4000g for 10 minutes at 4C. The supernatant was removed, and the pellet was resuspended in 20ul of DIW and 230ul of an aqueous solution of 1% SDS + 20mM 2-Mercaptoethanol. The suspended pellets were boiled in a hot water bath for 10 minutes, after which 30ul of 10% NP40 (PNGase F Kit), 40ul of G7 Buffer (PNGase F Kit), and 5ul of PNGase F (glycerol free) were added to the mixture. The pellets were incubated for 16 hours at 37C under gentle agitation. After incubation, 60ul of Calbiosorb beads were added to the mixture to remove SDS, and this mixture was incubated for 15 minutes under gentle agitation at room temperature. At this point, the sample was ready for purification.

4.2.3 Glycan Purification

Cell surface glycans were purified using two different schemes. Unlike the second scheme, the first scheme allows for fractionation of the neutral and acidic glycans. In this case, the first scheme was used for characterization of erythrocyte glycans because of their relative abundance, and the second scheme for HTE glycans to minimize sample losses, since the total yield of HTE ranges anywhere from 50 -100 million cells.

In the first purification scheme, glycans were first purified using a C18 column (Sep-Pak Vac, Waters). The columns were equilibrated with 6mL of MeOH, followed by 3mL DIW, and then 3mL of 0.05% TFA in DIW. The columns were then transferred to an empty Falcon tube , and the sample was added in at least 50ul of DIW. The column was washed with 3mL of 5%ACN in 0.05%TFA to wash loosely bound glycan from the column. The eluent was the lyophilized. Samples were then diluted in 400-500ul DIW prior to the next step. Glycans were fractionated into neutral and acidic fractions, and separated from salts using the Supelco EnviCarb column (Supelco). The columns were equilibrated with 3mL of ACN followed by 3mL DIW. The sample was added, and the column was washed with 3mL of 0.05% TFA in DIW to remove salts. Neutral glycans were eluted with 3mL of 25% ACN in DIW, and acidic glycans with 3mL of 50% ACN in 0.05% TFA . The eluent was lyophilized and the dry sample was stored at -20C prior to MS analysis.

In the second purification scheme, resins packed in a GELoader tip are used to purify glycans, according to a slightly modified method [8]. The tips of GELoader tips (rounded ends) were pinched by compressing them with a large 25ml pipette. DIW (100µl) was added to each tip. SP20SS beads (20µl) were added to the surface of the water in the GELoader tip. The beads were left to settle for ~5minutes. H^+ Dowex Cation Exchanger Beads (40µl) were added to the column in the same manner. The column was washed with 200ul DIW, and then the sample was added in ~20ul volume. The flow-through was collected, and loosely bound glycans were washed from the microcolumn with an additional 200ul DIW. The flow-through was lyophilized.

4.2.4 Glycan Sialic Acid Linkage Analysis

Sialic acid linkage analysis was performed on the cell surface glycans by adding 3ul of Sialidase A from *Athrobactur Ureafaciens* (Prozyme) or Sialidase S from *Streptococcus Pneumoniae* (Prozyme) to 10ul of the glycan solution. The samples were then incubated at 37C for 18 hours under gentle agitations. As a control, 10ul of untreated glycan was also incubated at 37C, except in the case of HTE glycans due to sample amount. Samples were either purified with the Supelco column or the GELoader tip microcolumn using the aforementioned methods.

4.2.5 Glycan MS Analysis

All glycans were analyzed using the Voyager DE-STR MALDI-TOF MS (Applied Biosystems). Acidic glycans were analyzed using 10mg/mL ATT in ethanol. The purified sample can be diluted in a number of different volumes depending on the starting amount. The sample and matrix was combined in a 1:9 ratio, respectively. Nafion (1ul) was spotted on the plate and allowed to dry for ~5minutes. The matrix-sample mixture was then spotted on top of the Nafion spot and allowed to dry in a humidity chamber (humidity 23%). The following parameters were used for acidic glycan analysis: Negative and Linear Mode, 22,000V Accelerating Voltage, 93% Grid Voltage, 0.3% Guide Wire, 150 nsec Delay.

Neutral glycans were analyzed with 20mg/mL DHB dissolved in a 50/50 mixture of ACN and 0.01% TFA in DIW. The sample and matrix were combined in a 1:1 ratio, respectively. This mixture was spotted (1ul) on the plate and left to dry at room temperature. The following parameters were used in the analysis of neutral glycans: Positive and Linear Modes, 22,000V Accelerating Voltage, 93% Grid Voltage, 0.15% Guide Wire, 150 nsec Delay. Proposed glycan compositions for each peak were determined by imputing the peak masses into the GlycoMod software (http://expasy.org), which calculates all mathematically possible glycan compositions for a given mass.

4.3 RESULTS

4.3.1 RBC Glycan Analysis

The RBC cell surface glycan extraction and purification procedures were successfully completed, and MALDI-TOF spectra showing the neutral and acidic glycans of chicken and turkey red blood cells are shown in Figures 4.1 and 4.2, respectively. Each cell type appears to have a broad range of both neutral and acidic glycans present on the cell surface, and no two spectra look similar. At first glance, it is evident that cRBCs contain a few acidic glycans that are quite prominent; most have intensities that are at least 30% of that of the most prominent peak. The spectrum for tRBC glycans, on the other hand, show many more acidic glycans, with the majority of peaks having an intensity less that 30% of that of the most prominent peak. Both cell types appear to have few glycans above an m/z of 3,000 and below an m/z of 1700.

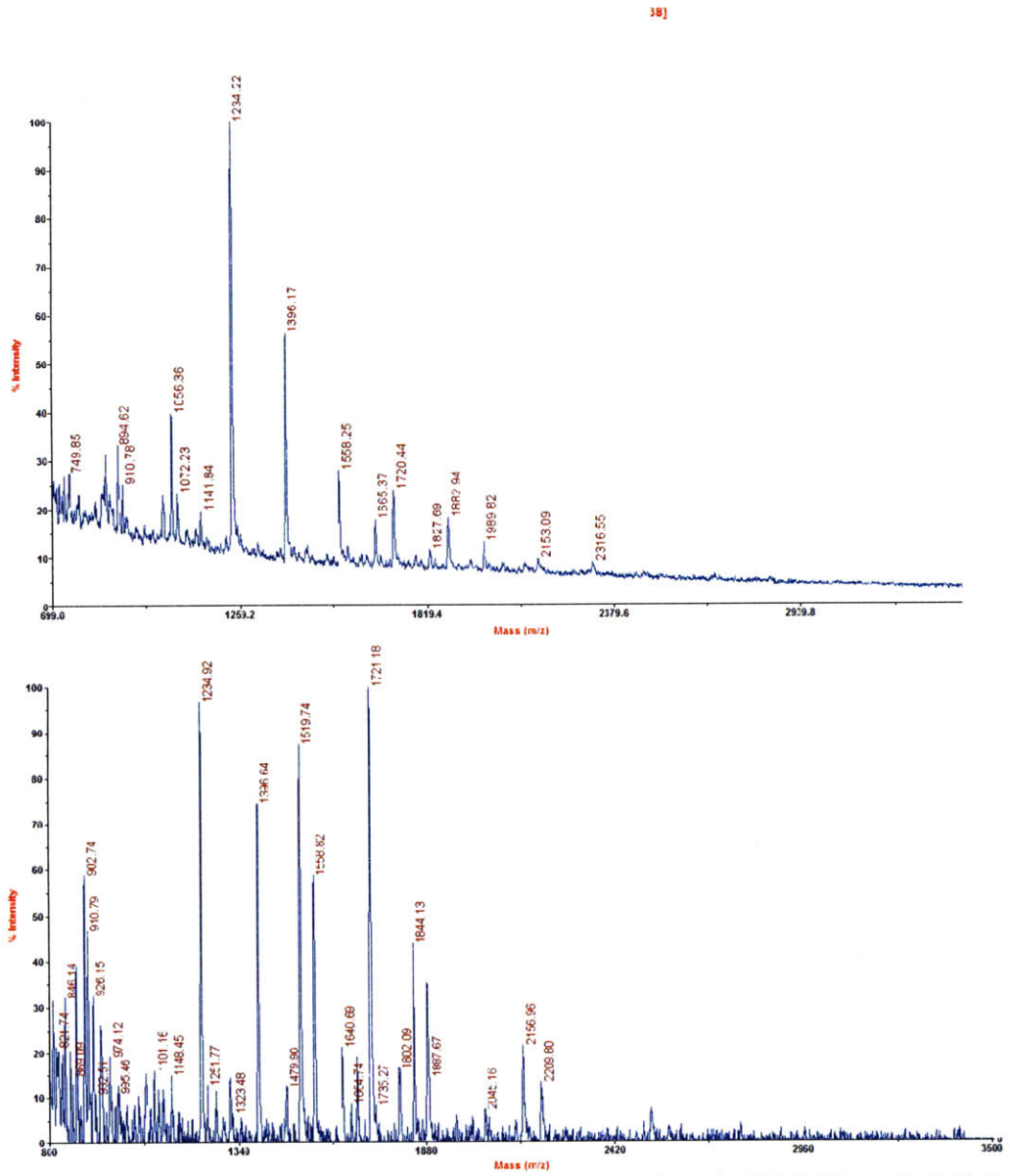


Figure 4.1. Voyager MALDI-TOF MS Spectra of cRBC and tRBC Neutral Cell Surface Glycans. tRBC (top figure) and cRBC (bottom figure) neutral glycans were purified by way of Supelco and analyzed by MALDI-MS using 20mg/mL DHB in 0.01%TFA/ACN (50:50,v/v).

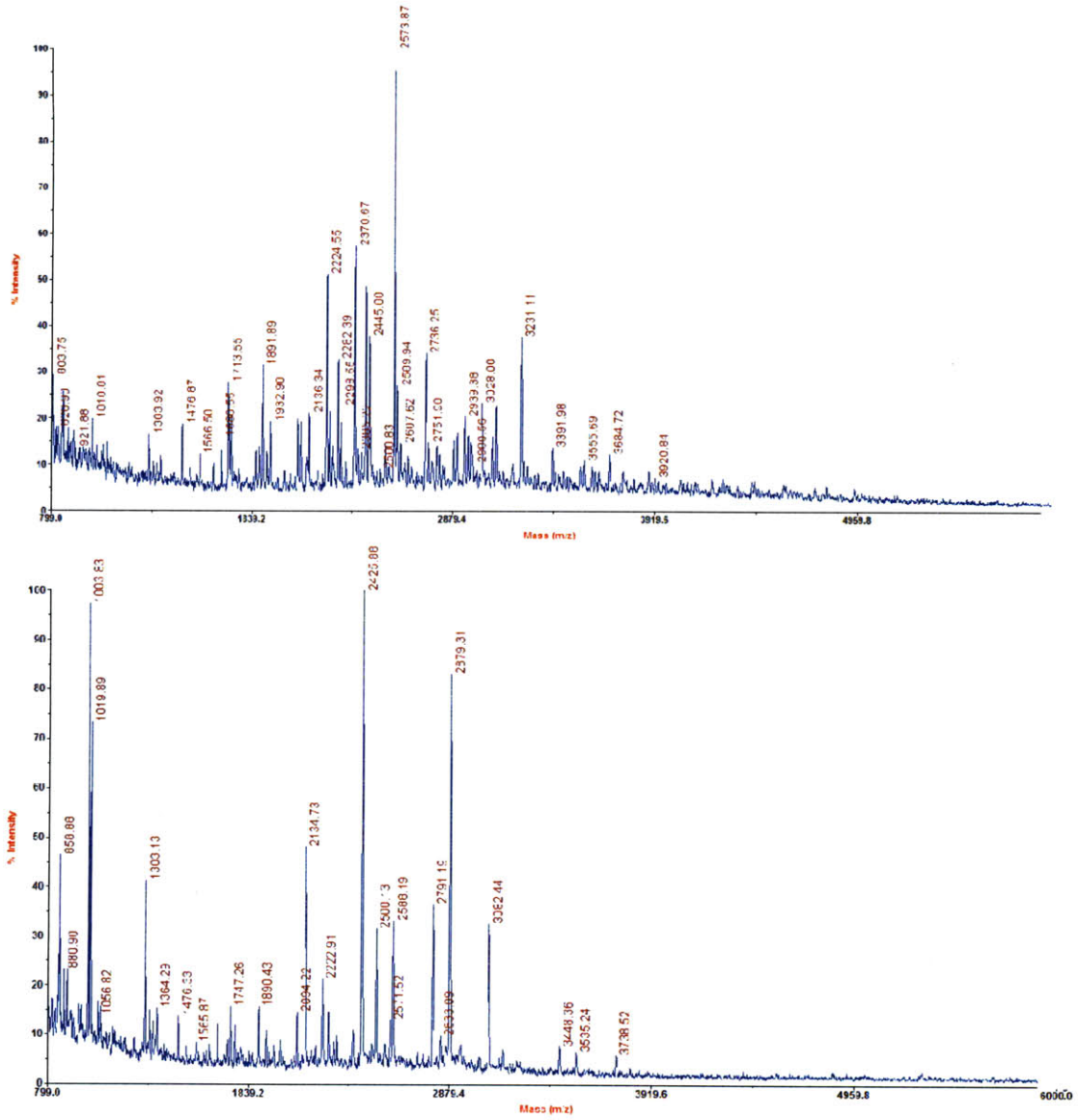


Figure 4.2. Voyager MALDI-TOF MS Spectra of Acidic cRBC and tRBC Cell Surface Glycans. tRBC (top figure) and cRBC (bottom figure) acidic glycans were purified by way of Supelco and analyzed by MALDI-MS using 10mg/mL ATT in ETOH.

Since one of the aims of this project was to characterize and compare the glycans from each of these cell types, a list of the masses and glycan compositions of the 10 most prominent peaks in each acidic and cRBC and tRBC neutral glycan spectra was made (Tables 4.1 and 4.2). The proposed compositions were

picked from a list generated by the GlycoMod software, which takes in an array of mass values and outputs all of the glycan compositions that are mathematically possible. This "list" was then further scrutinized by omitting all of the compositions that appeared to disobey the N-glycan biosynthetic rules. Therefore, the compositions listed in Tables 4.1 and 4.2 are the most reasonable for any given mass.

Table 4.1. Most Prominent Neutral Glycan Peaks of Turkey and Chicken Red Blood Cells.

Mass	**Proposed Glycan Composition**	**tRBC**	**cRBC**
894.62	$(Deoxyhexose)_1 + (Man)_2\ (GlcNAc)_2$		x
1056.36	$(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		x
1072.23	$(Hex)_1 + (Man)_3\ (HexNAc)_2$		x
1234.92	$(Hex)_2 + (Man)_3(GlcNAc)_2$	x	x
1396.64	$(Hex)_3 + (Man)_3(GlcNAc)_2$	x	x
1519.74	$(HexNAc)_3 + (Man)_3(GlcNAc)_2$	x	
1558.82	$(Hex)_4 + (Man)_3(GlcNAc)_2$	x	x
1665.37	$(HexNAc)_3\ (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		x
1721.18	$(Hex)_5 + (Man)_3(GlcNAc)_2$	x	x
1844.13	$(Hex)_2\ (HexNAc)_3 + (Man)_3(GlcNAc)_2$	x	
1884.06	$(Hex)_6 + (Man)_3(GlcNAc)_2$	x	x
1989.82	$(Hex)_2\ (HexNAc)_3\ (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		x
2156.96	$(Hex)_3\ (HexNAc)_3\ (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$	x	x
2045.16	$(Hex)_7 + (Man)_3(GlcNAc)_2$	x	
2209.8	$(Hex)_8 + (Man)_3(GlcNAc)_2$	x	
2316.55	$(Hex)_4\ (HexNAc)_3\ (Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$		x

Table 4.2. Ten Most Prominent Acidic Glycan Peaks of Turkey and Chicken Red Blood Cells.

Mass	Proposed Compositions	tRBC	cRBC
1713.55	$(Hex)_1$ $(HexNAc)_1$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$	x	
1747.26	$(Hex)_3$ $(HexNAc)_1$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		x
	$(Hex)_2$ $(HexNAc)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
1890.43	$(Hex)_3$ $(HexNAc)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$	x	x
	$(Hex)_1$ $(HexNAc)_1$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_2$ $(HexNAc)_1$ $(Deoxyhexose)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_3$ $(HexNAc)_1$ $(Deoxyhexose)_2$ + $(Man)_3(GlcNAc)_2$		
2134.73	$(Hex)_2$ $(HexNAc)_3$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		x
	$(Hex)_1$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
2222.91	$(Hex)_2$ $(HexNAc)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$	x	x
	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_2$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_1$ $(HexNAc)_2$ $(Deoxyhexose)_1$ $(NeuAc)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
2282.39	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$	x	
	$(Hex)_1$ $(HexNAc)_3$ $(NeuAc)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_1$ $(HexNAc)_3$ $(Deoxyhexose)_2$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
2370.67	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$	x	
	$(Hex)_1$ $(HexNAc)_2$ $(NeuAc)_2$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_9$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_4$ $(HexNAc)_4$ + $(Man)_3(GlcNAc)_2$		
2425.88	$(Hex)_2$ $(HexNAc)_3$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$	x	x
	$(HexNAc)_3$ $(NeuAc)_1$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_1$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_2$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		
2445.00	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$	x	
	$(Hex)_1$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_2$ $(HexNAc)_3$ $(NeuAc)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_2$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_3$ + $(Man)_3(GlcNAc)_2$		
2500.13	$(Hex)_3$ $(HexNAc)_4$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		x
	$(Hex)_1$ $(HexNAc)_4$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_2$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_3$ $(HexNAc)_4$ $(Deoxyhexose)_2$ + $(Man)_3(GlcNAc)_2$		
2573.87	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$	x	

	$(Hex)_1$ $(HexNAc)_3$ $(NeuAc)_2$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_9$ $(HexNAc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_4$ $(HexNAc)_5$ + $(Man)_3(GlcNAc)_2$		
2588.19	$(Hex)_3$ $(HexNAc)_3$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		x
	$(Hex)_1$ $(HexNAc)_3$ $(NeuAc)_1$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_1$ $(HexNAc)_3$ $(Deoxyhexose)_2$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_2$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		
2736.25	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$	x	
	$(Hex)_2$ $(HexNAc)_3$ $(NeuAc)_2$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_{10}$ $(HexNAc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_5$ $(HexNAc)_5$ + $(Man)_3(GlcNAc)_2$		
2791.19	$(Hex)_3$ $(HexNAc)_4$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		x
	$(Hex)_1$ $(HexNAc)_4$ $(NeuAc)_1$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_2$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_3$ $(HexNAc)_4$ $(Deoxyhexose)_2$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		
2879.31	$(Hex)_3$ $(HexNAc)_3$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$		x
	$(Hex)_1$ $(HexNAc)_3$ $(NeuAc)_2$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_9$ $(HexNAc)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_2$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_4$ $(HexNAc)_5$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		
3082.44	$(Hex)_3$ $(HexNAc)_4$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$		x
	$(Hex)_1$ $(HexNAc)_4$ $(NeuAc)_2$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_2$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_2$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_3$ $(HexNAc)_4$ $(Deoxyhexose)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_4$ $(HexNAc)_6$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
3231.11	$(Hex)_3$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$	x	
	$(Hex)_2$ $(HexNAc)_4$ $(Deoxyhexose)_2$ $(NeuAc)_2$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_3$ $(HexNAc)_6$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_4$ $(HexNAc)_6$ $(Deoxyhexose)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$		
	$(Hex)_5$ $(HexNAc)_6$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		

Even though this project seeks to determine the structural characteristics of acidic glycans in order to elucidate their potential effect on HA binding, neutral glycan analysis is also just as important because it can provide insight on the significance and abundance, if any, of the non-sialylated forms of the acidic glycans. Furthermore, a comparison can be done between these glycans and the neutral glycans that arise from Sialidase treatment of acidic glycans. The neutral tRBC and cRBC glycans shown in Table 4.1 are mostly high mannose in character, suggesting that either these N-glycans had not entered the final stage of N-glycan processing that occurs in the Golgi Body or that these neutral glycans naturally exist on the cell surface of these cells. Interestingly, tRBCs and cRBCs only share the high mannose type neutral glycans, containing either 2, 3, 4, 5, and 6 mannose residues. The neutral glycans that they do not share appear to have been processed because they contain additional HexNAc, Hexose, and Deoxyhexose residues. Furthermore, these cell-type specific neutral glycan species appear to be the de-sialylated versions of some of the prominent acidic glycans listed in Table 4.2.

According to Table 4.2, chicken and turkey red blood cells do not share similar acidic glycan distributions. There is only an overlap of 3 glycans between the two cell types, meaning that 17 of the glycans are unique to a specific cell type. Furthermore, the three glycan peaks that these cell types share are not present at the same intensities. For example, the peak at m/z 2222.91 is the 3rd most prominent peak for tRBCs and only the 8th most prominent peak for cRBCs. It is also important to note that while these cells may share the same glycan peaks, they do not necessarily share the same glycan, since one m/z value can yield multiple potential glycan sequences, as shown in Tables 4.1 and 4.2. Since the binding properties of each type of RBC are conferred by their glycans, it is possible that these RBCs agglutinate HA using different glycan motifs, especially since the most prominent glycans on the surface of each cell type are different.

In order to determine the most relevant glycan compositions for each m/z value and ultimately draw conclusions about their relevance to the glycans that are present in the human tracheal and bronchial epithelium, a sialic acid linkage analysis experiment was performed. In a direct way, this experiment

would be able to provide information on the type of sialic acid linkages present on the cell surface. Sialidase S only cleaves α2-3 and α2-8 linked terminal sialic acid moieties, while Sialidase A cleaves α2-3, α2-6, and α2-8 linked sialic acids. Comparing samples treated with both enzymes provides information about the relative abundance of each linkage, which is important for influenza studies, since humanized HA molecules tend to bind α2-6 linked sialic acids, while avian subtypes bind α2-3 linked sialic acids. Indirectly, sialic acid linkage analysis provides clarity on the most relevant structural compositions for each mass, since the number or type of sialic acid present in each composition may be different. For example, the possible glycan compositions for m/z 2588.19 include structures with either two NeuAcs, two NeuGcs, one NeuAc and two NeuGcs, one NeuAc and one NeuGc, or one NeuAc only. Treatment of this sample with Sialidase A will yield a resultant neutral species that is unique to each composition, since the number and size of the sialic acids in each composition are different. Using this information, one can potentially rule out irrelevant structural compositions.

Before analyzing the results of this analysis, it is important to acknowledge the specific methods that were used in the experiment, as well as the information that each method can provide. Typically, the acidic fraction of the cRBC and tRBC glycans were subject to the Sialidase treatment, and were subsequently purified by Supelco column and separated into neutral and acidic fractions. The neutral fraction will contain neutral glycans only if a Sialidase (either A or S) is successful at cleaving *all* sialic acids off of the particular glycan. In this project, since Sialidase A cleaves every sialic acid, only the information obtained from the Sialidase S treatment can provide information about the sialic acids present in a glycan population. Table 4.3 shows a summary of some of the common experimental outcomes of Sialidase treatment of different glycan subpopulations. Note that both the neutral MS and acidic MS analysis provide critical information about the types of moieties present on each glycan. For example, if a mixture only contains glycans with both α 2-3 and α2-6 linked sialic acids (i.e. every glycan has both species), then the experimental outcome will be different from one in which the mixture contains some

glycans that are only α2-3 linked and some that are α2-3 and α2-6 linked. These outcomes were used to draw conclusions about the cell surface glycan populations in tRBCs and cRBCs.

Table 4.3. Experimental MALDI-MS Outcomes Used to Characterize the Sialic Acid Linkages within a Glycan Population after Sialidase S Treatment

Sialic Acid Linkages Present in Mixture	**Neutral Species Will Appear?**	**Acidic Species Will Appear?**
2-3 only	Yes, all species should now be neutral	No, all species lost their 2-3 moiety.
2-6 only	No, Sialidase cannot cleave 2-6 linked moieties.	Yes, the spectra should be the same as the original
2-3 and 2-6	No, Sialidase S cannot cleave 2-6 moieties present on each and every glycan	Yes, every glycan present in the initial spectra will appear as a shifted mass due to the loss of the 2-3 moiety.
2-3 only, 2-6 only, and 2-3 and 2-6*	Yes, only those within the mixture that have 2-3 linkages	Yes, only those that have 2-6 only or 2-3 and 2-6 linkages. The latter species will have a shifted mass due to a loss of the 2-3 linked moiety.

* Please note that the linkages stated above are for *each* glycan species. In other words, a mixture can have glycans that *only* have 2-3 sialic acids, *only* have 2-6 linkages, *only* have both, or have a mixture of the abovementioned glycans.

Figures 4.3 and 4.4 show the comparisons between Sialidase A, Sialidase S, and non-Sialidase treated samples for each type of RBC. The non-sialidase treated sample spectra has been shifted by 5,000 counts so that clear-cut differences or similarities between the three spectra may be observed. Overall, the results of the sialic acid linkage analysis show that turkey and chicken RBC surface glycans contain a mixture of α2-3 and α2-6 linked sialic acids. More specifically, cRBC glycans probably contain mixtures of α2-3 only, α2-6 only, *and* α2-3 and α2-6 linked sialic acids, since treatment of these samples with Sialidase S results in the appearance of new neutral and acidic glycan species. Furthermore, in the spectra showing acidic residues, some peaks completely disappear with Sialidase S treatment, while others remain. This indicates that there are at least some species within the cRBC acidic glycan population that have sialic

acids that are only α2-6 linked and some that are either only α2-3 linked or both α2-3 and α2-6 linked. The neutral cRBC MS spectra also supports these conclusions since the Sialidase A and Sialidase S treated neutral glycan fractions show the same glycan peaks, but with different relative signal intensities.

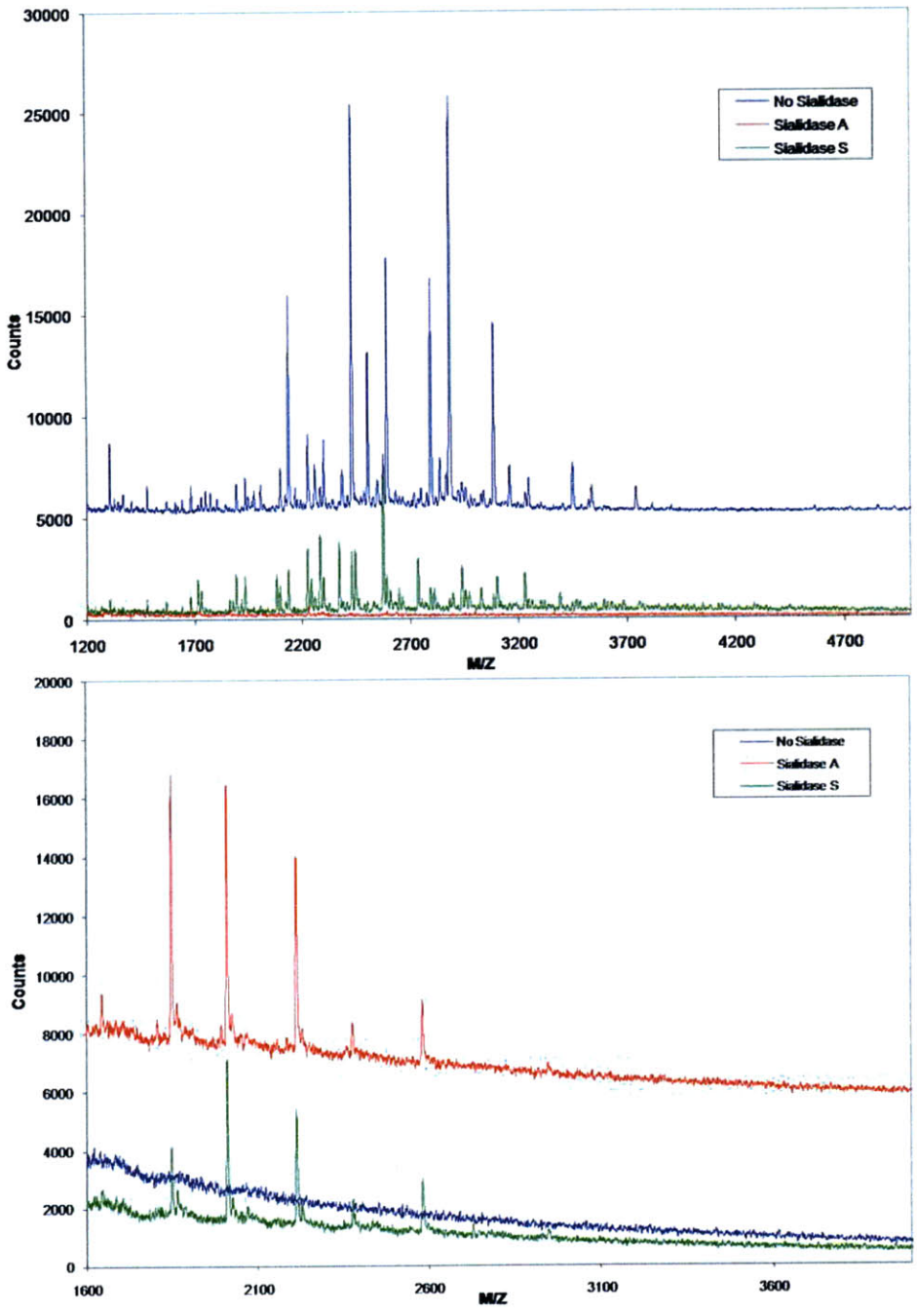


Figure 4.3. Sialic Acid Linkage Analysis of cRBC Acidic Glycans. Acidic cRBC glycans were treated either with No Sialidase (blue), Sialidase A(red), or Sialidase S (green) and incubated at 37C for 18 hours. The top spectra are the acidic glycans purified after Sialidase treatment. The spectra for the three incubation conditions were overlaid, and that for the sample treated with No Sialidase was shifted up by 5,000 counts to allow for a clear comparison of the three spectra. The bottom spectra show the neutral glycans purified after treatment conditions. The Sialidase A labeled spectra was shifted up by 5000 counts to allow for an easy comparison between the three spectra.

The sialic acid linkage analysis also provided information about the most relevant glycan conformations for each glycan peak. By subtracting off the masses of all of the relevant sialic acids for each proposed glycan conformation, and then crosschecking this mass with the masses present in the Sialidase A treated neutral glycan spectra, the most probable glycan conformation for each mass was identified. Table 4.5 shows the most relevant glycan conformations for the ten most prominent acidic glycans of cRBCs. Of the 39 potential glycan compositions listed for the ten most prominent cRBC acidic glycans, 27 were eliminated because of the results of the sialic acid linkage analysis. The resultant glycan compositions appear to consist of very little if any fucose, and mostly NeuAc sialic acid residues. There don't appear to be many lactosamine units (GlcNac-Gal) because of the low number of GlcNac and Gal residues per glycan, but the results of this experiment don't absolutely confirm or deny the presence of these moieties. Finally, Table 4.5 shows a glycan population that contains species with mostly 1-2 sialic acid residues, although one of the most prominent peaks at m/z 2879.31 contains 3 sialic acid residues.

Table 4.4. Comparison of Prominent Glycan Species Present in Sialidase A and S Treated cRBC samples.

Mass	Present in Original Spectra?	Present in Sialidase A Spectra?	Present in Sialidase S Spectra?
1747.26	Yes	No	No
1890.43	Yes	No	No
2134.73	Yes	No	Yes
2222.91	Yes	No	Yes (significantly reduced)
2425.88	Yes	No	Yes (significantly reduced)
2500.13	Yes	No	Yes (significantly increased)
2588.19	Yes	No	Yes (significantly reduced)
2791.19	Yes	No	No
2879.31	Yes	No	No
3082.44	Yes	No	No

Table 4.5. Most Probable Glycan Compositions of the Ten Most Prominent Acidic cRBC Surface Glycans.

Mass	Proposed Compositions
1747.26	$(Hex)_2$ $(HexNAc)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$
1890.43	$(Hex)_3$ $(HexNAc)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_1$ $(HexNAc)_1$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_2$ $(HexNAc)_1$ $(Deoxyhexose)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$
2134.73	$(Hex)_2$ $(HexNAc)_3$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
2222.91	$(Hex)_2$ $(HexNAc)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2425.88	$(Hex)_2$ $(HexNAc)_3$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2500.13	$(Hex)_3$ $(HexNAc)_4$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
2588.19	$(Hex)_3$ $(HexNAc)_3$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2791.19	$(Hex)_3$ $(HexNAc)_4$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2879.31	$(Hex)_3$ $(HexNAc)_3$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$
3082.44	$(Hex)_3$ $(HexNAc)_4$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$

The results of the tRBC sialic acid linkage analysis, shown in Figure 4.4, are unlike those of cRBC in a very interesting way. Firstly, since the MS spectra for acidic glycans illustrates that nearly every major peak has shifted or disappeared after Sialidase S treatment, it is apparent that tRBC glycans have a mixture of α2-3 and α2-6 linked sialic acids. Furthermore, the resultant neutral glycan spectrum shows no neutral glycans for Sialidase S treated tRBC. This data suggests that not a single glycan had all of its sialic acid residues removed, and more interestingly, that every glycan was at least altered by Sialidase S because none of the peaks of the treated and untreated spectra overlap. In other words, every glycan composition in the tRBC mixture has both α2-3 and α2-6 linked sialic acids.

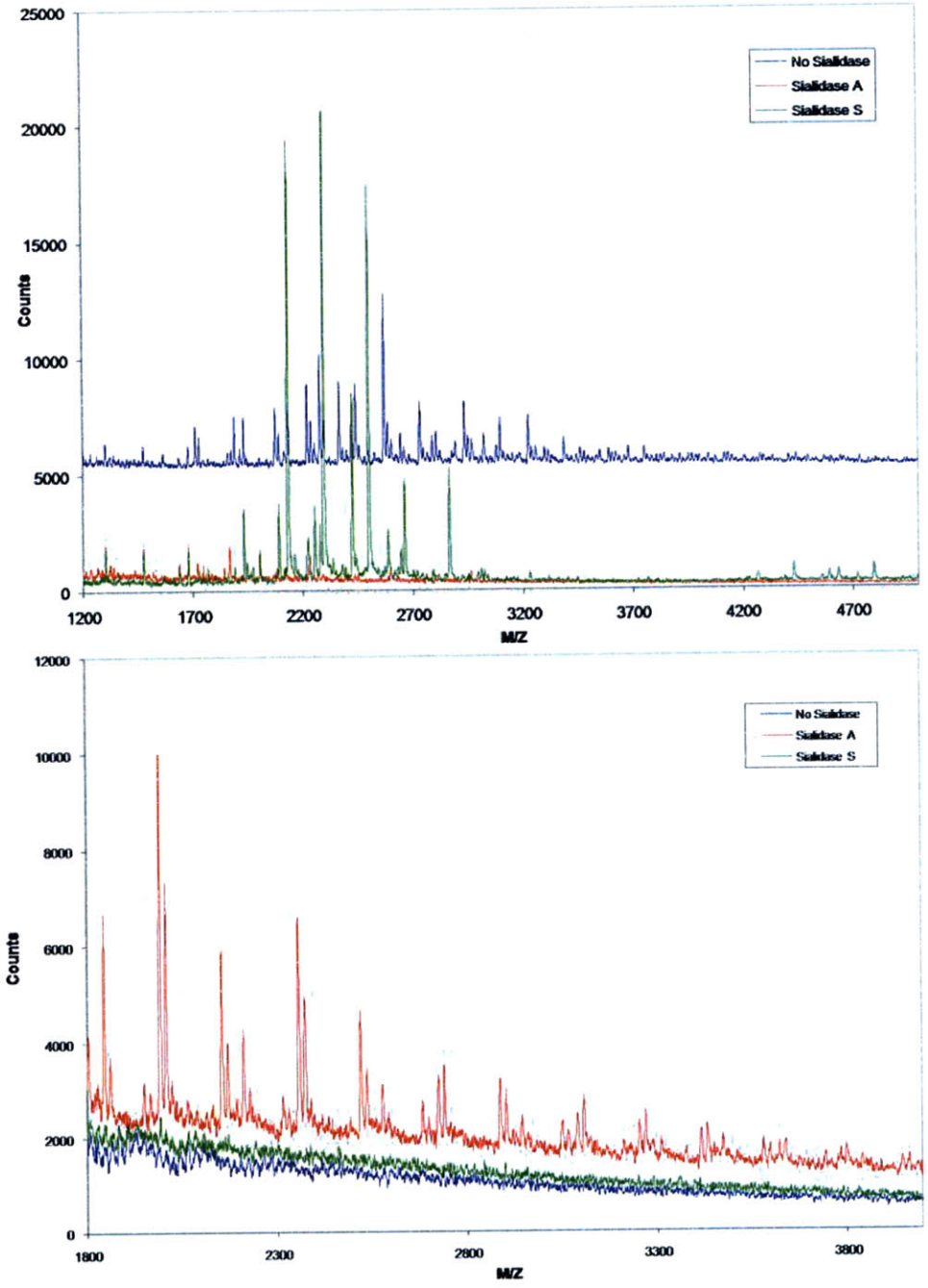


Figure 4.4. Sialic Acid Linkage Analysis of tRBC Acidic Glycans. Acidic tRBC glycans were treated either with No Sialidase (blue), Sialidase A(red), or Sialidase S (green) and incubated at 37C for 18 hours. The top spectra are the acidic glycans purified after Sialidase treatment. The spectra for the three incubation conditions were overlaid, and that for the sample treated with No Sialidase was shifted up by 5,000 counts to allow for a clear comparison of the three spectra. The bottom spectra show the neutral glycans purified after treatment conditions.

As with cRBC, the number of relevant tRBC glycan compositions was also reduced because of the sialic acid linkage results. Table 4.7 shows the most relevant glycan conformations for the ten most prominent acidic glycans of tRBCs. Of the 37 potential glycan compositions listed, 24 were eliminated. Unlike cRBC glycan compositions, those of tRBC contain fucose residues in most of the species. Furthermore, analysis on all of the tRBC peaks (not just the top 10) shows that unlike cRBC glycans, tRBC glycans contain both NeuAc and NeuGc sialic acid residues. These results are consistent with previously reported information about tRBC cell surface glycans. There don't appear to be many lactosamine units (GlcNac-Gal) because the majority of the prominent species contain 2-3 HexNac moieties, but again, the results of this experiment don't absolutely confirm or deny the presence of these moieties. Finally, Table 4.7 shows a glycan population that contains species with mostly 1-2 sialic acid residues.

Table 4.6. Comparison of Prominent Glycan Species Present in Sialidase A and S Treated tRBC samples.

Mass	Present in Original Spectra?	Present in Sialidase A Spectra?	Present in Sialidase S Spectra?
1713.55	Yes	No	No
1890.43	Yes	No	No
2222.91	Yes	No	Yes (significantly reduced)
2282.39	Yes	No	Yes (significantly reduced)
2370.67	Yes	No	No
2425.88	Yes	No	Yes
2445.00	Yes	No	No
2573.87	Yes	No	No
2736.25	Yes	No	No
3231.11	Yes	No	Yes (significantly reduced)

Table 4.7. Most Probable Glycan Compositions of the Ten Most Prominent Acidic tRBC Surface Glycans Based on Sialic Acid Linkage Analysis

Mass	Proposed Compositions
1713.55	$(Hex)_1$ $(HexNAc)_1$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
1890.43	$(Hex)_3$ $(HexNAc)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_1$ $(HexNAc)_1$ $(NeuGc)_2$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_2$ $(HexNAc)_1$ $(Deoxyhexose)_1$ $(NeuGc)_1$ + $(Man)_3(GlcNAc)_2$
2222.91	$(Hex)_2$ $(HexNAc)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2282.39	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
2370.67	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2425.88	$(Hex)_2$ $(HexNAc)_3$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2445.00	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
2573.87	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2736.25	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
3231.11	$(Hex)_5$ $(HexNAc)_6$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$

4.3.2 HTE Glycan Analysis

The HTE cell surface glycan extraction and purification procedures were successfully completed, and MALDI-TOF spectra showing the acidic and neutral glycans of HTE cells are shown in Figure 4.5. The HTE acidic glycan distribution shows three prominent peaks sitting amongst a range of peaks with lower but similar intensities ranging from m/z 1475.77 to m/z 3683.76.

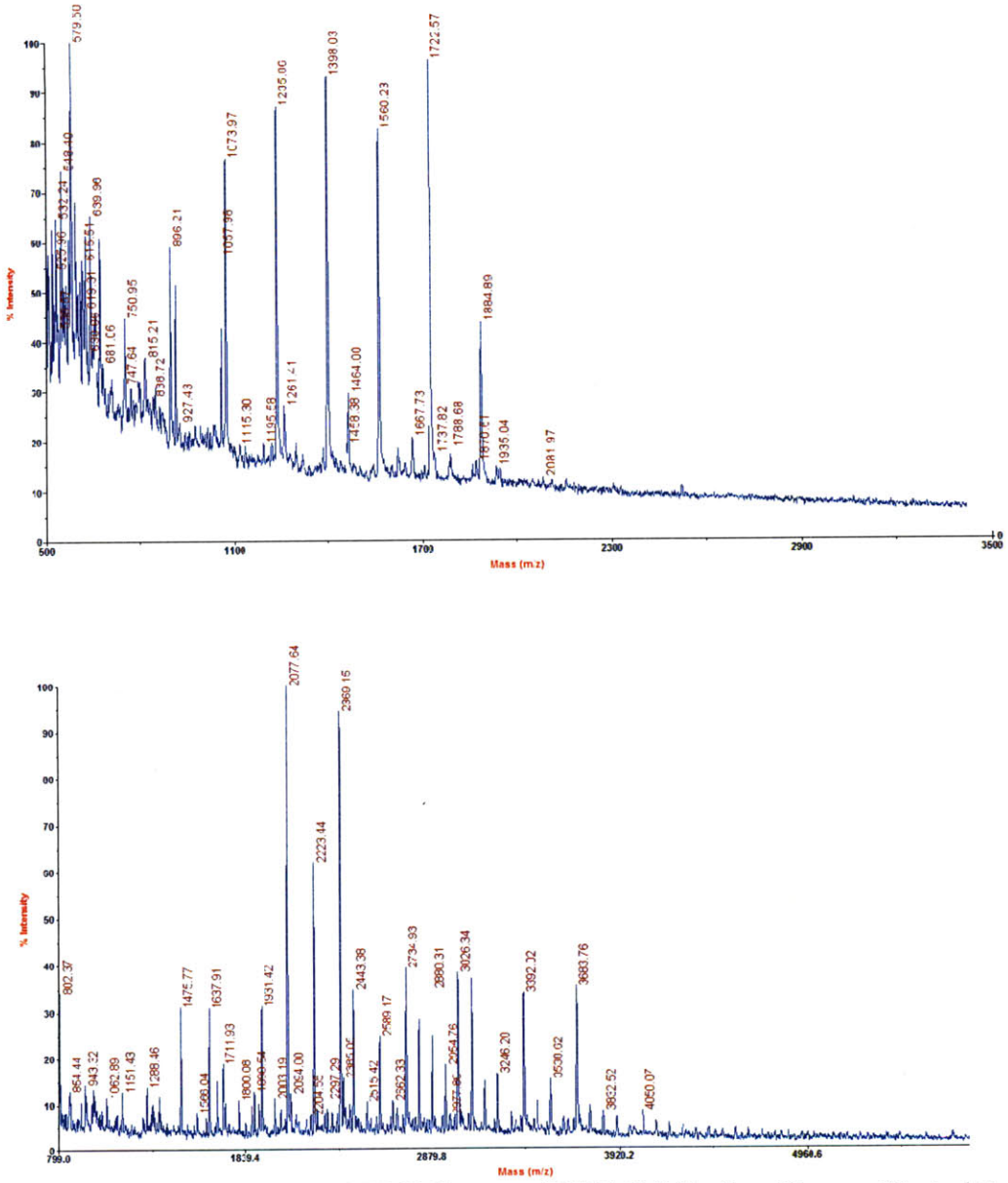


Figure 4.5. Voyager MALDI-TOF MS Spectra of HTE Cell Surface Glycans. Neutral (top figure) and acidic (bottom figure) glycans of HTE were purified by way of Supelco and analyzed by MALDI-MS using 10mg/mL ATT in ETOH and 20mg/mL in 0.01% TFA/ACN (50:50, v/v), respectively.

A list of the masses and glycan compositions of the most prominent peaks in each spectra was made (Tables 4.8 and 4.9). From comparing the data in Table 4.9 with that in Table 4.2, it is evident that 2 and 4 of the ten most prominent acidic glycan peaks are same between HTE and cRBC and tRBC,

respectively. Still, it is important to re-iterate that while these cells may share the same glycan peak masses, they do not necessarily share the same glycan, since one m/z value can yield multiple potential glycan structural compositions. Four of the remaining six prominent HTE glycans have masses above an m/z 3000, and are not observed in tRBC and cRBC MS spectra. These glycans appear to have at least 4 (HexNac)(Hex) units. Even though this MALDI data does not provide any information about the topology of the glycans, just the fact that these 4 glycans have more HexNAc-Hex units that those of tRBC or cRBC means that these HTE glycans either have longer glycan arms on average (compared to cRBCs and tRBCs) or they have a higher degree of branching.

Table 4.8. Most Probable Glycan Compositions of the Neutral HTE Surface Glycans

Mass	Proposed Glycan Composition
896.21	$(Deoxyhexose)_1 + (Man)_2 (GlcNAc)_2$
1073.97	$(Hex)_1 + (Man)_3 (GlcNAc)_2$
1235.86	$(Hex)_2 + (Man)_3(GlcNAc)_2$
1398.03	$(Hex)_3 + (Man)_3(GlcNAc)_2$
1560.23	$(Hex)_4 + (Man)_3(GlcNAc)_2$
1722.57	$(Hex)_5 + (Man)_3(GlcNAc)_2$
1884.89	$(Hex)_6 + (Man)_3(GlcNAc)_2$

Table 4.9. Most Probable Glycan Compositions of the Ten Most Prominent Acidic HTE Surface Glycans

Mass	Proposed Compositions
2077.64	(Hex)2 (HexNAc)2 $(Deoxyhexose)_1$ $(NeuAc)_1 + (Man)_3(GlcNAc)_2$
	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_3 + (Man)_3(GlcNAc)_2$
2223.44	$(Hex)_2$ $(HexNAc)_2$ $(NeuAc)_2 + (Man)_3(GlcNAc)_2$
	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_2$ $(NeuAc)_1 + (Man)_3(GlcNAc)_2$
2369.15	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_1$ $(NeuAc)_2 + (Man)_3(GlcNAc)_2$
	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_3$ $(NeuAc)_1 + (Man)_3(GlcNAc)_2$
	$(Hex)_9 + (Man)_3(GlcNAc)_2$
	$(Hex)_4$ $(HexNAc)_4 + (Man)_3(GlcNAc)_2$
2443.38	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1 + (Man)_3(GlcNAc)_2$
	$(Hex)_1$ $(HexNAc)_1$ $(Deoxyhexose)_2$ $(NeuAc)_3 + (Man)_3(GlcNAc)_2$

2734.93	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_3$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_{10}$ $(HexNAc)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_3$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_5$ $(HexNAc)_5$ + $(Man)_3(GlcNAc)_2$
2880.31	$(Hex)_3$ $(HexNAc)_3$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_5$ $(HexNAc)_5$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_{10}$ $(HexNAc)_1$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
3026.34	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_{10}$ $(HexNAc)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_3$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_5$ $(HexNAc)_5$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_5$ $(HexNAc)_5$ $(Deoxyhexose)_2$ + $(Man)_3(GlcNAc)_2$
3100.42	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_6$ $(HexNAc)_6$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_{11}$ $(HexNAc)_2$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_3$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
3392.02	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_{11}$ $(HexNAc)_2$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_3$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_6$ $(HexNAc)_6$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_6$ $(HexNAc)_6$ $(Deoxyhexose)_2$ + $(Man)_3(GlcNAc)_2$
3683.76	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_4$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_{11}$ $(HexNAc)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_3$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_6$ $(HexNAc)_6$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_{11}$ $(HexNAc)_2$ $(Deoxyhexose)_4$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_6$ $(HexNAc)_6$ $(Deoxyhexose)_2$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$

In order to decipher more information about the most relevant HTE glycan compositions for each m/z value, a sialic acid linkage analysis experiment was performed. The results of the HTE sialic acid linkage analysis, shown in Figure 4.6, indicate that Sialidase S treatment reduces the intensities of the majority of

the prominent acidic glycans, suggesting that these glycans contain at least one α2-3 linked sialic acid residue. Some of the prominent peaks, like those at m/z 2369.15, 3026.34, and 3110.42, either disappear completely or are significantly reduced in intensity with Sialidase S treatment. These peaks represent glycans that contain at least one α2-3 linked sialic acid because every glycan for that specific m/z was modified by Sialidase S. However, it is also possible that these glycans have α2-6 linked sialic acids. This theory is supported by the appearance of several new acidic glycan peaks in the Sialidase S spectra. For example, the peak appearing at m/z 2807.35, which has the second highest intensity, was not in the original spectra, and has a proposed glycan composition of $(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$. This species could arise from the loss of a sialic acid from the original glycan species at m/z 3100.42, 3392.02, and 3683.76, since all of these species have at least 4 lactosamine units at one fucose in their proposed compositions. Nonetheless, the strong appearance of this peak in the Sialidase S spectra suggests that there is at least one glycan antenna with an α2-6 linked sialic acid in each of these three glycans. Furthermore, the fact that very few neutral species show up in the Sialidase S treated samples implies that the majority of Sialidase S treated glycans are still sialylated. As a matter of fact, all of the 6 prominent neutral glycans in the Sialidase S spectra are actually acidic glycans that are ionizing in the presence of DHB. This theory is supported by the presence of sodium adduct peaks, which are a hallmark of the ionization of acidic glycans in the positive mode in MALDI-MS analysis.

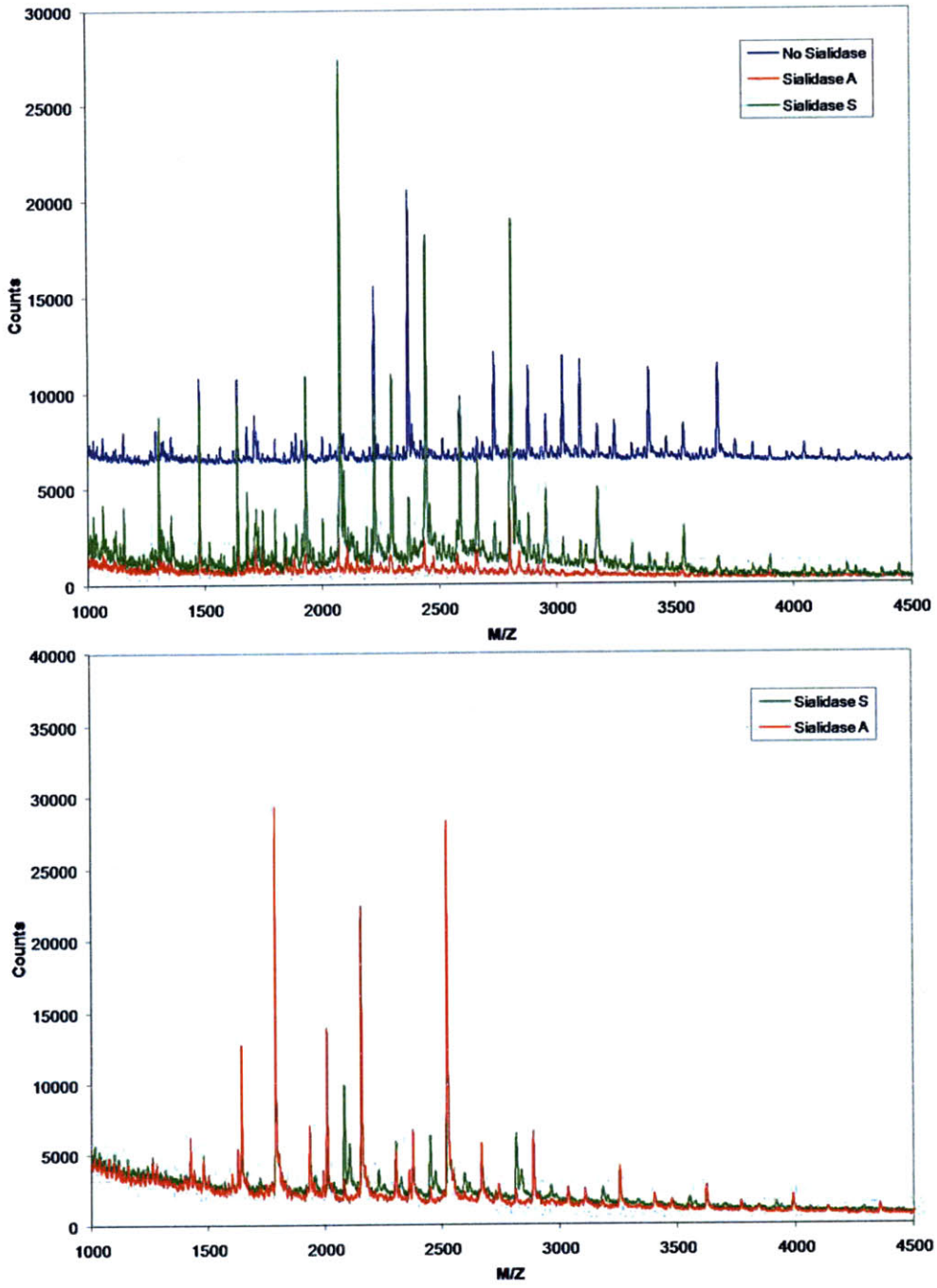


Figure 4.6. Sialic Acid Linkage Analysis of HTE Acidic Glycans. Acidic HTE glycans were treated either with No Sialidase (blue), Sialidase A(red), or Sialidase S (green) and incubated at 37C for 18 hours. The top spectra are the acidic glycans purified after Sialidase treatment. The spectra for the three incubation conditions were overlaid, and that for the sample treated with No Sialidase was shifted up by 6,000 counts to allow for a clear comparison of the three spectra. The bottom spectra show the neutral glycans purified after treatment conditions.

Table 4.11 shows the most relevant glycan conformations for the ten most prominent acidic glycans of HTE, as determined by the sialic acid linkage analysis. Of the 39 potential glycan compositions listed, 28 were eliminated. Like the glycans on the surface of tRBCs, the majority of HTE cell surface glycans appear to contain fucose residues. Furthermore, unlike the most probable glycan compositions of cRBCs and tRBCs, the majority of HTE glycan compositions suggest the presence of at least one "long" glycan arm, meaning that on at least one glycan arm, there are at least 2 lactosamine units. This theory must be confirmed with additional MS studies, but the preliminary analysis seems to indicate HTE cell surface glycans likely contain long glycan arms which *may* by α2-6 sialylated, like those present on the glycans of HBE cells. Finally, Table 4.11 suggests that the HTE glycan population contains species with variable degrees of charge, since these glycans have anywhere from 1 to 4 sialic acid residues.

Table 4.10. Comparison of Prominent Glycan Species Present in Sialidase A and S Treated HTE samples.

Mass	Present in Original Spectra?	Present in Sialidase A Spectra?	Present in Sialidase S Spectra?
2077.64	Yes	No	Yes
2223.44	Yes	No	Yes
2369.15	Yes	No	Yes (significantly reduced)
2443.38	Yes	No	Yes
2734.93	Yes	No	No
2880.31	Yes	No	Yes (significantly reduced)
3026.34	Yes	No	Yes (significantly reduced)
3100.42	Yes	No	Yes (significantly reduced)
3392.02	Yes	No	Yes (significantly reduced)
3683.76	Yes	No	Yes (significantly reduced)

Table 4.11. Most Probable Glycan Compositions of the Ten Most Prominent Acidic HTE Surface Glycans Based on Sialic Acid Linkage Analysis

Mass	Proposed Compositions
2077.64	$(Hex)_2(HexNAc)_2$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
2223.44	$(Hex)_2$ $(HexNAc)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2369.15	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2443.38	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$
2734.93	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
2880.31	$(Hex)_3$ $(HexNAc)_3$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$
	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
3026.34	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$
3100.42	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$
3392.02	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$
3683.76	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_4$ + $(Man)_3(GlcNAc)_2$

Using the aforementioned analytical methodology, the results from the HTE linkage analysis suggest that HTE cell surface glycans contain a mixture of α2-3 and α2-6 linked sialic acids, with the majority of prominent acidic glycans containing at least one α2-6 linked sialic acid. At this point, it is difficult to decipher where the α2-6 linked sialic acid may be on the glycan antennas, or if it is attached to a GlcNAc residue. Of particular interest is whether this sialic acid residue is located on the arm of the glycan with more than one lactosamine unit, mostly because this arm is most susceptible to HA binding. In the future, additional MS-MS and NMR studies can provide further insight on the structural specifics of these glycans.

4.4 DISCUSSION

The ultimate goal of the experiments in this project was to be able to obtain some very basic structural information about cRBC, tRBC, and HTE cell surface glycans so that researchers could evaluate the potential advantages or disadvantages of using chicken and turkey RBCs as an experimental system to classify HA subtypes and perform HA inhibition assays. More specifically, if one of the desired outcomes of influenza research is the development of an effective vaccine or therapy, it will be crucial

that these treatments be specific to humans. Along these lines, it will be necessary to ensure that inhibition assays are performed with cells that express at least the glycans that are predominant on the upper respiratory tracts of humans.

The MS characterization and sialic acid linkage analysis results from this project have suggested that the glycans in tRBCs and cRBCs fail to fully capture the glycan distributions present in the human respiratory tracts. Only 5 of the ten most prominent peaks are same between HTE and cRBC/tRBC; cRBCs and HTEs only share two prominent acidic glycans masses, while tRBCs and HTEs share four. These results also suggested that while HTE may contain glycan antennary arms with more than 2 lactosamine units, those of tRBC and cRBC probably do not; in other words, 80% of the prominent HTE glycans contained at least 3 HexNAc-Hex units, while only 30% of tRBC and 50% of cRBC prominent acidic glycans contain these units. Of course, allegations of the existence of long glycan arms with α2-6 linked sialic acids must be further confirmed with additional MS-MS and NMR studies, and these studies will ultimately validate the results shown and conclusions drawn here by confirming the existence of these entities and providing quantitative estimates.

In conclusion, the results from this project indicate that cRBCs and tRBCs may not be the best experimental system for studying influenza inhibition or binding in humans because of the basic structural differences between the glycans present on the surface of all cells in question. In the future, additional studies should be performed to further validate these conclusions, and researchers should try to look into other more relevant model systems to study these important binding event.

4.5 REFERENCES

1.) "Influenza Research" 2009 www.wikepedia.org. Wikimedia Foundation. 9 May 2009. <http://en.wikipedia.org/wiki/Influenza_research>.

2.) "Influenza A Virus" 2009. www.wikipedia.com. Wikimedia Foundation. 9 May 2009. <http://en.wikipedia.org/wiki/Influenzavirus_A>

3.) "Avian and Pandemic Influenza: Preparedness and Response" 2009. www.usaid.gov US Aid. 9 May 2009. <http://www.usaid.gov/our_work/global_health/home/News/news_items/avian_influenza.html>.

4.) Klausner, R. D. and R. Sitia (1990). "Protein degradation in the endoplasmic reticulum." Cell **62**(4): 611-4.

5.) Chandrasekaran, A., A. Srinivasan, et al. (2008). "Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin." Nat Biotechnol **26**(1): 107-13.

6.) Spackman, Erica Avian influenza virus (2008) Springer.

7.) McNamara, N. A., R. A. Sack, et al. (2000). "Mucin-bacterial binding assays." Methods Mol Biol **125**: 429-37.

8.) Huang, Y., T. Konse, et al. (2002). "Matrix-assisted laser desorption/ionization mass spectrometry compatible beta-elimination of O-linked oligosaccharides." Rapid Commun Mass Spectrom 16(12): 1199-204.

5 Summary and Future Directions

The three projects carried out in this thesis successfully illustrate the importance of the study of glycobiology by showing how both the development and application of tools to study glycans can result in new and interesting information relevant to the pathological processes that affect human health and disease.

The O-glycan analytical method optimization project successfully showed that sound MALDI-MS spectra for bovine mucin and fetuin O-linked glycans could be obtained and that sialic acid linkage information about the glycans on these proteins could be determined. However, researchers can further exploit the benefits of this procedure by performing additional structural studies on the released glycans, such as NMR, permethylation, and MS-MS, to complement the type of preliminary information acquired in this project. Ultimately, this method could be widely beneficial to our understanding of key human physiological and pathological processes because of the omnipresence of O-linked glycans and mucins in the majority of living cells.

In the glycoprotein direct binding project, bovine fetuin and mucin were effectively bound to microtiter plates. This assay could potentially become a powerful tool for comparing and contrasting the binding specificities of two or more glycan-binding proteins on one substrate. It is therefore well suited for influenza studies which seek to determine quantitative binding differences between two or more HA subtypes. When used in conjunction with glycan arrays and glycan characterization experiments, a more complete picture of the binding dynamics can be obtained, and researchers can specifically focus on therapeutic and vaccine targets with particular characteristics, such as glycan topologies and linkage types.

researchers to think about the relevance of the experimental model systems that are currently being used to study influenza inhibition or binding in humans.