

Spring 2019

Synthetic Lectins, a Diagnostic and Prognostic Tool for Detecting Glycans in Breast Cancer

Daniel James Gordon

Follow this and additional works at: <https://scholarcommons.sc.edu/etd>

 Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#)

Recommended Citation

Gordon, D. J. (2019). *Synthetic Lectins, a Diagnostic and Prognostic Tool for Detecting Glycans in Breast Cancer*. (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/etd/5281>

This Open Access Dissertation is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact dillarda@mailbox.sc.edu.

Synthetic Lectins, a Diagnostic and Prognostic Tool for Detecting Glycans
in Breast Cancer

by

Daniel James Gordon

Bachelor of Arts
Boston University, 2012

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Biochemistry

College of Arts and Sciences

University of South Carolina

2019

Accepted by:

John J. Lavigne, Major Professor

Thomas Makris, Committee Member

Maksymilian Chruszcz, Committee Member

Hexin Chen, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

© Copyright by Daniel James Gordon, 2019

All Rights Reserved.

ABSTRACT

Cancer diagnostic tools have been pushed to the forefront of medical research because a person's chance of survival is directly correlated with how early the tumor is identified and treatment is begun. In searching for the subtle differences between healthy and tumor cells, almost every type of cancer has been shown to under, over, or neo express glycans, and these changes in the glycan fingerprint can continue as the disease progresses. This provides a powerful diagnostic opportunity that's works by screening for glycoproteins and glycosylation patterns that deviate from normal cells.

Boronic acids have a useful and tunable property in that they can selectively, but cross-reactively, bind to carbohydrates. They do so by forming reversible covalent bonds with 1,2 and 1,3 cis-diols, which are found all over saccharides. The Lavigne laboratory has created bead-based sensors, called Synthetic Lectins, which can preferentially bind to different glycoproteins in biological samples, using modified boronic acids bound to amino acids conjugated to polystyrene beads.

This dissertation will first cover instrumentation and procedural developments that have advanced this project, to allow it to be translated into the clinical setting. Following this will be a deeper analysis and validation of this glycoprotein sensor array, using hydrophobic interaction liquid chromatography-electrospray ionization-mass spectrometry (HILIC-ESI-MS) to analyze N-linked glycans extracted from breast cell line secretions using the Synthetic Lectin array. Finally, new ways of using this Synthetic Lectin array as a prognostic tool and a research tool will be explored.

TABLE OF CONTENTS

ABSTRACT.....	iii
LIST OF TABLES.....	v
LIST OF FIGURES	vii
CHAPTER 1: CANCER, GLYCOMICS, AND SYNTHETIC LECTINS	1
CHAPTER 2: FLOW CYTOMETRY AS A PLATFORM FOR USING SYNTHETIC LECTINS	13
CHAPTER 3: USING SYNTHETIC LECTINS FOR SOLID PHASE EXTRACTION	68
CHAPTER 4: USING THE SL ARRAY FOR SOLID PHASE EXTRACTION TO ANALYZE N-LINKED GLYCANS IN BIOLOGICAL SAMPLES	85
CHAPTER 5: FUTURE WORK AND EXPERIMENTS	129
BIBLIOGRAPHY.....	131
APPENDIX A: FLOW CYTOMETRY MATLAB ALGORITHM.....	134
APPENDIX B: N-LINKED GLYCANS STRUCTURES	140

LIST OF TABLES

Table 1.1 SL1-9 amino acid sequence	9
Table 2.1 Resin chart by Rapp Polymer	19
Table 2.2 Standard deviation in seconds of sample collection by FC500	39
Table 2.3 Comparing the Amino Acid Sequences of SL1, 2, 4 and 5	55
Table 4.1 Relative amount of A2G2, FA2G2, and A2G2S2 per sample.....	100
Table 4.2 Peak intensity and color scale.....	101
Table 4.3 Properties of monosaccharides found in N-linked glycans	102
Table 4.4 High mannose N-linked glycans from cell lines.....	104
Table 4.5 High mannose N-linked glycans from cell lines eluted with SLs.....	105
Table 4.6 Bi-antennary N-linked glycans from cell lines.....	107
Table 4.7 Bi-antennary N-linked glycans from cell lines eluted with SLs.....	108
Table 4.8 Tri-antennary N-linked glycans from cell lines	110
Table 4.9 Tri-antennary N-linked glycans from cell lines eluted with SLs.....	113
Table 4.10 Tetra-antennary N-linked glycans from cell lines	114
Table 4.11 Fucosylated tetra-antennary N-linked glycans from cell lines	115
Table 4.12 Tetra-antennary N-linked glycans from cell lines eluted with SLs	117
Table 4.13 Hybrid N-linked glycans from cell lines.....	120
Table 4.14 Hybrid N-linked glycans from cell lines eluted with SLs	122
Table B.1 High mannose glycan structures	140

Table B.2 Bi-antennary glycan structures.....	141
Table B.3 Tri-antennary glycan structures.....	142
Table B.4 Tetra-antennary glycan structures	143
Table B.5 Fucosylated tetra-antennary glycan structures.....	144

LIST OF FIGURES

Figure 1.1 Phenylboronic acid and 1,2 or 1,3 cis-diol bonding interaction.....	5
Figure 1.2 Diagram of a Synthetic Lectin.....	6
Figure 1.3 Microscope protocol using the SL array	10
Figure 1.4 Diagram of a flow cytometer analyzing a sample.....	12
Figure 2.1 Registering an event by a flow cytometer	20
Figure 2.2 PBS vs. new blocking buffer	24
Figure 2.3 SKBR3 cell line with SL1 data collected at various PMT voltages.....	29
Figure 2.4 Effects of PMT voltage on LDA classification accuracy	31
Figure 2.5 Bead number vs. RFU	32
Figure 2.6 Assessing the reproducibility of the flow cytometry platform.....	34
Figure 2.7 Protein concentration can increase signal standard deviation.....	35
Figure 2.8 Large differences between SLs results in higher LDA accuracy	37
Figure 2.9 Reproducibility of flow cytometry system after bead counting	43
Figure 2.10 Comparison study between two platforms	46
Figure 2.11 Microscope: SLs1-9 by five breast cell lines	48
Figure 2.12 FC500: SLs1-9 by five breast cell lines labeled at 10:1, FITC:lysine.....	50
Figure 2.13 TentaGel beads: SL1-9 by four breast cell lines at 10x FITC.....	52
Figure 2.14 TentaGel beads: SL1-9 by four breast cell lines at 1x FITC.....	52
Figure 2.15 Polyethyleneglycol and divinylbenzene repeat units in beads	54
Figure 2.16 Ranking SLs based on fluorescent intensity for both bead arrays.....	56
Figure 2.17 Discriminating NLs by SLs and SLs by NLs	59
Figure 2.18 Discriminating NLs by SLs and SLs by NLs, De-N-linked.....	63

Figure 2.19 Discriminating cell line extracts by SLs and NLs	65
Figure 2.20 Discriminating cell line extracts by SLs and NLs, De-N-linked.....	67
Figure 3.1 Protein % Recovery from using PS and TentaGel beads as SPE agents.....	72
Figure 3.2 Images of 10µm PS and 10µm TentaGel beads	73
Figure 3.3 Purified protein % recovery using SLs on 10µm PS beads.....	76
Figure 3.4 BSA (µg) bound to 51.28mg of SLs on 10µm PS beads.....	77
Figure 3.5 OVA and BSM (µg) recovered after using SLs as SPE	79
Figure 3.6 Protein recovery (µg) from SLs used as SPE with breast cell lines MCF10A and MDA-MB-231.....	81
Figure 3.7 Total protein (µg) eluted from SLs using elution buffer, pH = 4.5	83
Figure 4.1 The PNGase F reaction with an N-linked glycan	88
Figure 4.2 The RapiFluor-MS molecule used to tag N-linked glycans	90
Figure 4.3 Solvent gradient used for the HILIC LC-ESI-MS protocol	93
Figure 4.4 The human N-linked glycan biosynthesis pathway.....	94
Figure A.1 Matlab algorithm for flow cytometry protocol on PC.....	134

CHAPTER 1
CANCER, GLYCOMICS, AND SYNTHETIC LECTINS

INTRODUCTION

Cancer is the second leading cause of death in the United States, and is projected to surpass heart disease as the leading killer in many places by 2020.¹ As the average lifespan of people continues to rise, it is more likely that each person, whether directly or indirectly, will be impacted by cancer at some point in their lives. Cancer diagnostic tools have been pushed to the forefront of medical research because a person's chance of survival is directly correlated with how early the tumor is identified and treatment is begun.² A late stage diagnosis of any cancer type, once the cancer has spread throughout the body, forces systemic treatment options, such as chemotherapy agents, which are highly toxic to the body.³

In terms of breast cancer, the most common cancer in women, the mammogram is currently the most successful method for detecting breast tumors. However, it still has a 10% false positive rate and a 1% false negative rate, both often due to the varying breast tissue densities within and between individuals.⁴ Even if it was a perfect diagnostic technique, the mammogram is still a large and expensive instrument that requires trained personnel to operate and radiologists to interpret the results. Unfortunately, accessibility and economics are still important factors to considerations when developing a diagnostic tool, which was demonstrated in a study that correlated low socioeconomic status to late stage cancer diagnosis in the Detroit metropolitan area.²

The metastasis of a tumor is an involved process that requires a number of events to take place in order for it to successfully invade into another tissue of the body. These obstacles for the tumor include gaining access to the circulatory system, evading anoikis, resisting the shear force of the bloodstream, and finding a distal tissue that is compatible for colonization.⁵ Therefore, there is a window of time in which tumor cells and tumor secretions are circulating in the body, but the cancer has not yet been able to successfully colonize to a new location. This is a critical window for diagnosis because the circulating tumor cells, or more likely their secretions and debris, can be detected in the blood⁶, while the disease is still in its early stages and the prognosis is much more favorable.

Cancer is uniquely challenging to detect and treat because it is the own cells of the body slowly changing over time, and rarely causes any noticeable effects until the later stages.⁷ Cells that undergo tumorigenesis have to play a balancing act in order to survive. If they change too dramatically, then the cell itself will internally trigger one of its programmed cell death protocols. Or, if those fail-safes are already disabled within the cell, then the dramatically changed tumorigenic cells are usually detected and destroyed by the natural killer cells of the immune system.⁸ As a result of staying within these phenotypical parameters, surviving tumor cells have often disguised themselves as healthy cells in order to remain undetected by the immune system, consequentially making clinical detection and treatment more challenging as well.

BREAST CANCER

As personalized medicine continues to become the standard in every hospital, breast cancer has been further classified into characteristic subtypes, which are often

indicative of the prognosis and treatment plan as well. Since the breast is a reproductive organ, breast cancer subtypes are first classified by the overexpression of their hormone receptors.⁹ The three main hormone receptors found on breast tissue are the estrogen receptor (ER), the progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2). Generally speaking, breast cancer subtypes that overexpress one or more of these hormone receptors have more treatment options available because targeting these hormone receptors can slow down the growth of the tumor.¹⁰ Therefore, the cancer subtype with the worst prognosis is the triple-negative breast cancer (TNBC) subtype, which does not have the overexpression of any of the three aforementioned hormone receptors. In addition to TNBC having limited treatment options, it has the unfortunate reputation of being a fast spreading cancer that effects the youngest demographic of women of any other breast cancer subtype.¹¹ TNBC is the subtype that has seen the least amount of progress in treatment and survivability, and a favorable prognosis of TNBC relies much more heavily on early detection.

GLYCOMICS

The field of glycobiology has quickly made huge advancements in the past several decades, and has helped to explain the many roles of complex carbohydrates in cell-to-cell communication and the immune system. In a few examples, cells of different tissues types express different complex carbohydrates as their own tissue signature. Other cells can express or secrete different glycoproteins, during an inflammation or infection, to signal or alert the immune system.¹² Glycans are complex carbohydrates, or polymers of saccharides, with the general molecular formula of $(CH_2O)_n$. They are ubiquitously

used throughout the eukaryotic system and are found on all types of proteins and lipids. It is estimated that over 70% of all proteins are modified with carbohydrates, which add several benefits to the protein, including aqueous solubility and maintaining proper protein folding. The importance of glycosylation in the clinical setting is best exemplified by the history of blood transfusion and the appreciation that different glycan expression determines a person's blood type, and whether or not their body will accept or reject foreign blood. Glycans, however, are far less stable than proteins and are often isomeric, which are two reasons why they have been so challenging to study up until recent decades. Technological advancements have now made it easier to investigate carbohydrate expression by different cells. Several of these techniques will be discussed in detail in later chapters.

CANCER AND GLYCOMICS

In searching for the subtle differences between healthy and tumor cells, almost every type of cancer has been shown to under, over, or neo express glycans, and these changes in the glycan fingerprint can continue as the disease progresses.¹³ This provides a powerful diagnostic opportunity that's works by screening for glycoproteins and glycosylation patterns that deviate from normal cells. Bringing this method into the clinical setting has been attempted many times, however one major drawback has left this tool from becoming globally applicable. The biggest challenge facing this technology is that each individual person has his or her own unique glycoprotein expression profile, or glycan fingerprint. Furthermore, each person's glycan fingerprint can be dynamic, and often changes with a person's age and health.^{14,15} Since, two individuals with the same

disease and prognosis could express totally different glycan fingerprints, individual glycan markers are not reliable enough for diagnosing or assessing cancer alone.

Currently, Cancer antigen (CA) 15-3 and CA27-29 are the only carbohydrate markers recommended by the FDA to be used in clinical practice, and are both used for monitoring how successful a therapy is working.¹⁶ Both of these antigens are products of the MUC1 gene, which is an extensively O-linked glycoprotein that has shown to be overexpressed in some breast cancers.¹⁷ There is a great need for a reliable carbohydrate based test for the diagnosing and monitoring of breast cancer. However, this technique may only reliably work by widening the scope of the carbohydrate targets, and looking at the global changes in the glycosylation fingerprint, instead of only focusing on one or a few glycoproteins.

BORONIC ACIDS AND SACCHARIDES

Boronic acids have a useful and tunable property in that they can selectively, but cross-reactively, bind of carbohydrates. They can do this by forming reversible covalent bonds with 1,2 and 1,3 cis-diols, as shown below in Figure 1.1, which are found all over saccharides. This property has been extensively exploited by varying the substituents on

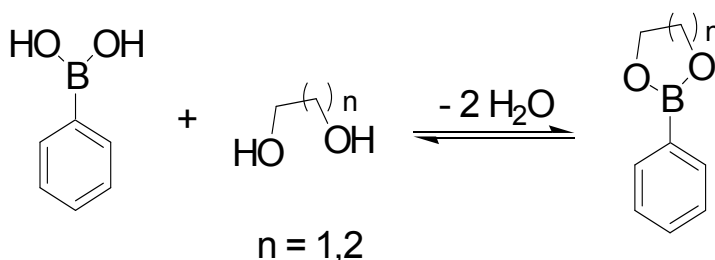


Figure 1.1 Phenylboronic acid covalently bonding to a 1,2 or 1,3 cis-diol to form a cyclic boronate ester.

the boronic acids, which can cause them to preferentially bind to different monosaccharides and disaccharides.¹⁸ The usefulness of phenylboronic acids in discriminating between different carbohydrates is the basis of this research project and will be examined in great detail, for both its utility as a blood-based diagnostic tool and later as a prognostic and research tool.

BEAD-BASED SYNTHETIC LECTINS

The Lavigne laboratory has created a bead-based system that can preferentially bind to different glycoproteins. This tool, shown in Figure 1.2, is made by incorporating phenylboronic acids onto amino acid chains, which are covalently bound to cross-linked polystyrene beads. These are referred to as Synthetic Lectins (SLs)¹⁹ because they



Figure 1.2 Diagram of a Synthetic Lectin. An amino-functionalized polystyrene bead bound with 12mers of amino acids. D* represents diaminobutanoic acid, which is modified with 2-formylphenylboronic acid. B represents β -alanine. X represents a random amino acid. The terminal arginine is acylated.

resemble lectins (NLs), which are proteins found throughout nature that recognize and bind saccharide epitopes, and are a large component of the immune system in plants and animals.²⁰ NLs are useful because they have varying degrees of epitope specificity, however, they can be expensive and challenging to work with because they need to be purified and require proper protein folding. Furthermore, only a relatively small amount

of NLs have been discovered and are commercially available. The SLs created in this laboratory behave similarly to NLs, but have the added benefits of cross-reactivity and the reversibility of the cyclic ester bond. Another benefit the SLs have over NLs is that the amino acid sequences on the polystyrene beads can be varied, which tailors the binding affinities of the SLs to different glycans. These properties make SLs ideal for detecting larger scope changes in the glycosylation fingerprint, and can be used to analyze aberrant structures, or deviations from the normal glycosylation profiles.

An individual SL does not have any added benefit over NLs, and in fact, is probably less useful due to its lack of epitope specificity. However, when combining many of these SLs in an array format, each with a different peptide sequence, patterns can start to emerge based on the different glycans present in a sample. This is a facet of the cross-reactive array format, which works by each component in the array having different, but overlapping binding properties. These patterns, or glycosylation fingerprints, can be used to distinguish between cancer and non-cancerous cells, and more specifically, between cancers with different metastatic potential.²⁰ The reversibility of the covalent bonds formed between boronic acids and cis-diols is beneficial for this purpose because it allows the strongest binding interaction to exist in solution. Therefore, if one glycan has a stronger affinity for a particular SL, then it will out compete any other weaker binding glycans, limiting the amount of nonspecific binding interactions.

Cancer is unpredictable, and can theoretically lead to an infinite arrangement of saccharides because cancer progression governed by random genetic mutations and the cell's subsequent compromises necessary for survival. The knowledge of the existing NLs is small and cannot account for all the possible changes that could occur during

tumorigenesis. Therefore, using an array of NLs is still too narrowly focused, whereas using an array of SLs has the potential to detect any subtle differences and unprecedented changes of tumorigenic cells.

PREVIOUS WORK IN THE LAVIGNE LABORATORY

This project began when previous members of the laboratory synthesized Synthetic Lectins 1-9, shown in Table 1.1, onto polyethyleneglycol-crosslinked polystyrene beads (TentaGel[®] beads by Rapp Polymer) using standard amino acid coupling reagents: N,N-dimethylformamide, piperidine, 4-methylmorpholine, and HBTU to deprotect and activate the amino acids for sequential addition. In a brief synopsis of a lot of preliminary work, the 12-mer on the SLs was chosen because it is the longest an amino acid chain can be before secondary folding interactions begin to occur within the peptide chain.

Table 1.1 The amino acid sequences of Synthetic Lectins 1-9

SL Name	Sequence
SL1	Ac-RGD*VTFD*RBBRM
SL2	Ac-RTD*RFLD*VBBRM
SL3	Ac-RSD*VTTD*RBBRM
SL4	Ac-RRD*TQTD*QBBRM
SL5	Ac-RAD*TRVD*VBBRM
SL6	Ac-RTD*NRND*FBBRM
SL7	Ac-RSD*YFTD*QBBRM
SL8	Ac-RTD*YGND*NBBRM
SL9	Ac-RTD*YQVD*ABBRM

The first methionine is conserved in the SLs because it allows easy cleavage of the 12-mer peptide, when using cyanogen bromide to form a methyl thiocyanate. The following arginine is to ensure good positive ionization when analyzing the peptide by mass spectrometry. The following two β -alanines are used to extend the 12mer away from the

bead. The diaminobutanoic acids, which are subsequently modified at the R-chain primary amine with 2-formylphenylboronic acid, were selected over lysines because the side chains are two carbons shorter, and are therefore less cross-reactive overall because the boronic acids are less exposed and flexible. The final arginine of the 12-mer is capped with an acyl group to prevent any unwanted side reactions. The X positions on the SLs were randomly selected amino acids (Excluding isoleucine, aspartic acid, glutamic acid, and histidine) to ensure good diversity in the SL array.

The nine SLs, in Table 1.1, were the most commonly used SLs in the laboratory, and were included in the array because they demonstrated an ability to differentially bind to the purified proteins, bovine serum albumin (BSA), ovalbumin (OVA), bovine submaxillary mucin (BSM), and porcine submaxillary mucin (PSM). The five variable amino acid positions of these 12mer were randomly generated by a split-and-pool method, which was incubated with two different purified proteins labeled with two different fluorophores. Since these 12-mers share at least 7 common amino acids in the same positions, SLs1-9 are all at least 58.3% identical in sequence, which is an idea that will be revisited in future chapters.

USING THE SL ARRAY TO DISCRIMINATE BETWEEN HEALTHY AND CANCEROUS CELLS

The first five SLs were initially synthesized on 300 μ m TentaGel beads, and were used to discriminate between different cell line glycoprotein extracts from tissue culture. The workflow of these experiments is outlined in Figure 1.3. Briefly, 2mg of each SL in the array was blocked with PBS, 1%BSA, 10%glycerol to prevent non-specific binding,

and then individually incubated with fluorescently labeled cell line glycoprotein extracts (Previously determined to be 0.0002mg/mL) overnight, and were washed with phosphate buffer saline three times the next day. The beads were then spread out onto a microscope slide and photographed with a fluorescent microscope (Leica6000 with CCD camera). These pictures were then processed through a MatLab algorithm to assign each bead a single value of fluorescent intensity on an 8-bit scale. This data was then processed through a Systat program where linear discriminant analysis was performed on the data sets, to see if the SL array could reliably classify the glycoprotein extracts by their respective cell lines, which in turn classified the cell lines as either healthy or cancerous.

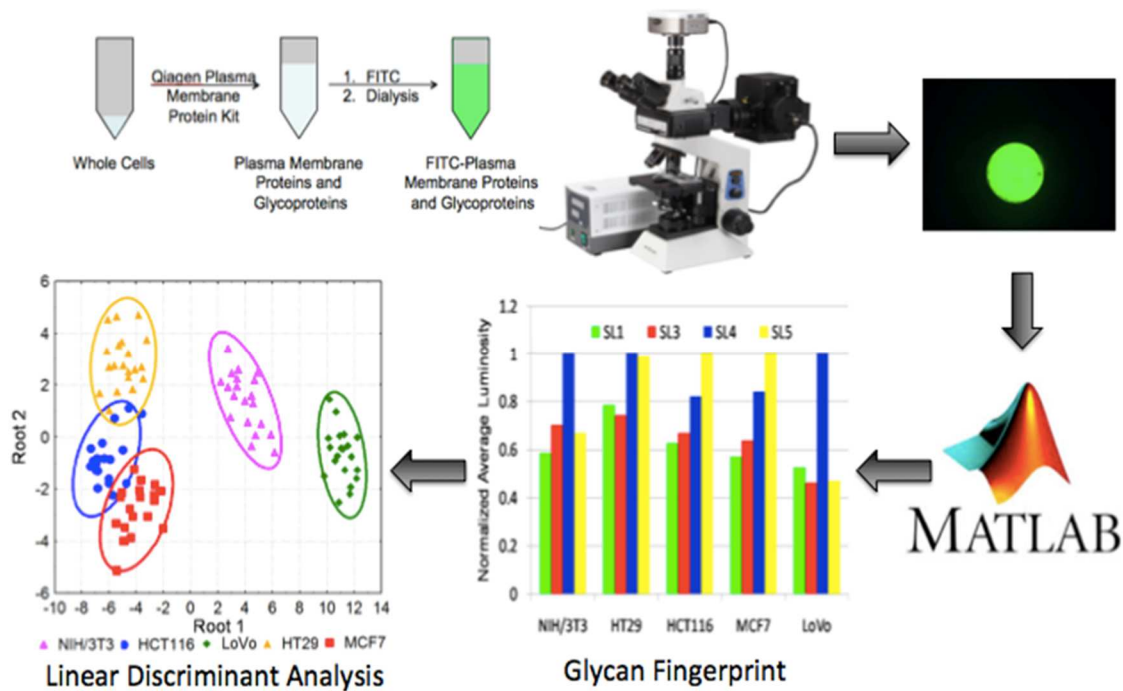


Figure 1.3 The Microscope Protocol used to discriminate between glycoprotein extracts of different cell lines from tissue culture. Extracts are fluorescently labeled with FITC and individually incubated with each SL in the array. All incubations, washes, and analysis take place in PBS_L. The results are processed through a MatLab algorithm before being analyzed by linear discriminant analysis.

This method was the foundation and proof-of-concept for this diagnostic tool, however the microscope procedure described above had inherent impediments that needed correcting before translating into clinical practice. These drawbacks were apparent by the small data sets generated after laborious fluorescent imaging and time consuming data processing. In addition, there was a large amount of influence the individual users had on the data sets because of lighting, normalization of exposure time, and picture taking styles. Essentially, the fluorescent microscope method was slow, cumbersome, and inconsistent between samples and users.

MOVING PLATFORMS FROM THE MICROSCOPE TO THE FLOW CYTOMETER

Keeping the limitations of the microscope protocol in mind, a new platform with better sensitivity and higher throughput was needed to process the bead samples. The flow cytometer was the ideal candidate for this because bead-based flow cytometry continues to become more common for research purposes, and flow cytometers are usually found in the hospital setting.

Originally designed for analyzing cells and particles in bodily fluid samples, flow cytometers work by injecting a very small amount of sample through a thin tube of 50-75 μm in diameter, where it is hydrodynamically focused by sheath fluid, and analyzed by a series of lasers and detectors. Because the tube is so thin, and the flow rate is generally 0.1-1 μl /second, each cell or particle in the sample can be analyzed independently for size, granularity, and fluorescent intensity, as shown in Figure 1.4.

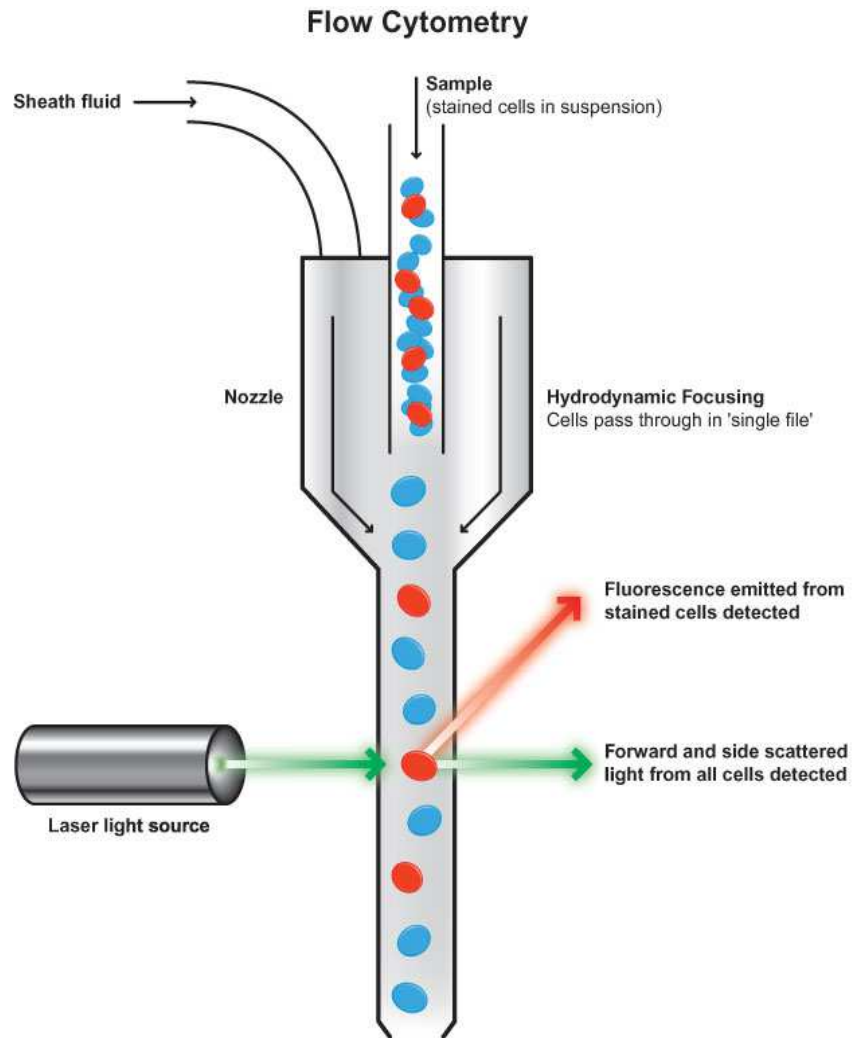


Figure 1.4 Diagram of a flow cytometer analyzing a cell sample.
(Image from Abcam)

The microscope protocol used 300 μm beads, which is much too large to fit into a flow cytometer, so the first step in transitioning from one platform to the next was to synthesize the Synthetic Lectin array onto polystyrene beads that can fit inside of the flow cytometer tube. Since the flow cytometer was designed to analyze human cells, which are on average 4-6 μm in diameter, a bead size of 10 μm was chosen.

CHAPTER 2

FLOW CYTOMETRY AS A PLATFORM FOR USING SYNTHETIC LECTINS

INTRODUCTION

I began by synthesizing SLs 1-9 onto 10 μ m PEGylated Polystyrene (TentaGel) beads, using the same peptide-building conditions described above, on an automated peptide synthesizer. The purpose was to incubate the SL array with labeled breast cell line membrane protein extracts, and then analyze the beads with a flow cytometer, to see if this diagnostic tool could be seamlessly transitioned to a new platform. The resins chosen for this project (Products of Peptide International and Rapp Polymer) are amino-functionalized beads (Specifications of +/-10%) with loading capacity of 0.25mmol/g.

The first experiment was run with four different breast cell line membrane protein extracts from the cell lines MCF10A, SKBR3, MCF7, and MDA-MB-231. These cell lines were chosen because they each represent a different breast cancer subtype. The MCF10A cells are healthy control cells, while SKBR3, MCF7, and MDA-MB-231 represented HER2+, Luminal A (ER+, PR+), and Triple Negative Breast Cancer (TNBC) subtypes, respectively. The cell line extracts used were courtesy of another member of the laboratory, who had used these specific extracts previously with the microscope protocol. Therefore, a direct comparison could be done between the two platforms to see if it was translatable. After the beads and fluorescently labeled extracts were incubated overnight, and washed 3x with PBS the following day, they were analyzed with the BD LSRII flow cytometer.

Having no prior knowledge of bead-based flow cytometry, these samples seemed to be clogging, or at least slowing down the flow rate of the flow cytometer. This revealed that the TentaGel beads have a tendency to stick together, which was also apparent by the huge standard deviations observed within each sample. However, the results looked very promising for platform transitioning because it accomplished both of the goals intended, higher throughput and data automation. Over 1000 data points, or individual beads, were collected from each sample in the array, whereas the microscope protocol only collects 20-30 data points per sample. Furthermore, the bead analysis was automated and internal, which alleviated any concern of changes in background/ambient light that can effect the brightness of the beads, any image distortion created on the peripherals of the microscope lens, and the risk of duplicating data points by accidentally capturing the same bead in multiple images.

The next few experiments were run with two additional breast cell line plasma membrane extracts, T47D and BT474. T47D is another Luminal A breast cancer subtype, which, like MCF7, is characterized by an overexpression of the estrogen receptors and progesterone receptors. The BT474 cell line is a Luminal B subtype, which is characterized by an overexpression of ER, HER2, and possibly the PR receptors. These different expressions, or lack of expressions, of hormone receptors have grave impacts on the aggressiveness and treatability of the patient's breast cancer.²¹ In keeping these subtypes in mind, another goal of the project was to further tailor the SL array, so that it could also be used as a prognostic tool, by being able to discriminate between breast cancer subtypes, based on how they each interact with the SL array.

CELL CULTURE CONDITIONS

Since glycosylation is a transient and dynamic process, there are many external factors that can have an influence on the glycan expression of a cell line. For this reason, extra consideration was put into the laboratory cell culture procedure to ensure consistency between cell lines and biological replicates. For example, the six breast cell lines used for these first experiments were cultured in different medias, including DMEM and RPMI 1640, with and without phenol red. To ensure that the different medias did not have an influence on the cell line glycan profiles, all cell lines were transitioned into RPMI 1640 (+1% penicillin/streptomycin, 10% bovine serum albumin (BSA)) without phenol red. Phenol red is used as a media indicator in cell culture, but it is also known to act as an estrogen analog that can bind weakly to the estrogen receptor.²² Therefore, it was not used since only some of the cell lines are ER+ and was an external interference. Also, prior to all glycoprotein extractions, all cells were serum-starved (RPMI 1640 +1% penicillin/streptomycin) for 48 hours, and were only collected if the culture flasks were 70-90% confluent.

Note: All PBS discussed is not the standard, commercially available PBS²³, but rather PBS with the following materials. Lavigne Lab PBS. This PBS will be referred to as PBS_L, and is essentially 5-10X more concentrated in salt content in order to further stabilize the charge separation created by the boronate ester formation. For this reason, commercially available PBS from Thermo Fisher Scientific is referred to as PBS 1X.

The glycoprotein extracts collected and analyzed during the first years of this project, were all membrane proteins collected, isolated, and purified from cell lines in culture using the Qproteome plasma membrane protein extraction kit (Qiagen). Briefly,

the cells were scraped from the cell culture flask, lysed in a hypotonic solution, and centrifuged to collect all of the cell membranes and debris. The cytosolic material was then removed by decanting the supernatant, and the peripheral and integral membrane proteins were separated from the lipid content using a proprietary Qiagen ligand, which bound to the membrane proteins and then later interacted with magnetic beads. The membrane proteins were then eluted from the ligand and magnetic beads, resuspended in carbonate buffer, and fluorescently labeled with fluorescein isothiocyanate (FITC).

SHIFTING BIOLOGICAL SAMPLES FROM PLASMA MEMBRANE PROTEINS TO SECRETED PROTEINS

In thinking of the future direction of this project, it was understood that there is little use in developing a cancer diagnostic tool that only analyzes plasma membrane proteins. This is because if a membrane sample is being taken, then the diagnosis of cancer has already been made prior to taking the biopsy of that tumor. This renders the SL array useless as a diagnostic test, but still could be used as a prognostic tool, in order to further characterize the tumor type and its aggressiveness. This use of the SL array as a prognostic tool will be discussed in later chapters, but the first focus of this project was to develop the SL array as an early stage diagnostic tool.

Since cancer secretions and debris can be detected in the blood as early as Stage 1⁶, the primary objective shifted to develop the SL array as a blood-based diagnostic tool. Therefore, a new procedure was developed by a previous member of the laboratory to collect the media from the cell lines in culture, in order to analyze the secretome, or the secreted glycoproteins from the cells. This transition from probing for membrane proteins

to secreted proteins was in its final stages as work on the flow cytometer platform was just beginning, and there was a brief overlap for the first few experiments. However, all subsequent data following the technical optimizations of the flow cytometry protocol can be assumed to be secreted proteins extracted from cell line media.

Briefly describing the secreted protein extraction protocol, the media of cells of 70-90% confluence was siphoned from the cell culture flasks after the cells were serum starved (RPMI 1640, +1% penicillin/streptomycin) for 48 hours. The media was centrifuged to remove any cellular debris, and the supernatant was filtered through 3kDa MWCO centrifuge tubes (Amicon) to remove any remaining smaller molecules and to concentrate the secreted proteins. This also served as a buffer exchange step to transfer the secreted proteins into PBS_L. The concentrated secreted protein extract was then precipitated with -20 degree Celsius acetone in a 4:1 ratio. Under these buffer conditions, the acetone precipitation works efficiently because of the extra salinity of the PBS_L, which helps crash the protein out of solution in higher percentage.²⁴

The protein pellet was then washed and resuspended into carbonate buffer, pH = 9.8, which has the basicity necessary for labeling proteins with FITC. Under these conditions, FITC binds to any primary amines present in the samples, which only includes the N-terminus of a polypeptide and any exposed R-chains of lysines. In an attempt to keep the protein labeling consistent, the amount of FITC used to label the protein extract was based off of the amount of primary amines (Lysines +1) on an average sized protein. This calculation was done by taking the protein concentration, and dividing it by the average size of a protein, which was assumed to be 60kDa, and then

multiplying by the number of lysines found on that 60kDa protein, assuming that lysines appear on a protein with a frequency of around 7%.

In this way, the protein extracts were labeled in either a 1:1 molar ratio or a 10:1 molar ratio of FITC:lysine, which summated to about 40 theoretical primary amines per 60kDa protein. Once this calculation was done, the protein extracts were incubated with FITC for 2 hours at 37°C, and then filtered again through a 3kDa MWCO centrifuge tube to remove any excess FITC, and buffer exchange the samples into PBS_L.

OPTIMIZING THE BEAD-BASED FLOW SYSTEM

The next experiment on the BD LSR II was not successful and clogged the line of the flow cytometer. So, the flow cytometry sample preparation protocol had to be carefully examined to ensure that the beads would not sticking together in future experiments. Since an array format was being used, including nine SLs by six cell line extracts, the number of samples to be analyzed quickly became a large number per experiment. Therefore, extra considerations also had to be made in order to ensure the flow cytometer would not clog halfway through a data set.

The first change made was with the solid phase resin being used. The TentaGel beads that had been used in all previous experiments on this project, swell 3-4x their size in aqueous solution, shown in Table 2.1 below, which makes them considerably larger than their counterpart polystyrene cross-linked polystyrene (PS) beads of the same size, which do not swell in aqueous solution.²⁵ This property makes PS beads a more ideal resin for use in the flow cytometer because they are smaller. In addition to the bead swelling, the PEGylated PS beads were observed to have a higher tendency to aggregate

or stick together, than the polystyrene cross-linked polystyrene (PS) beads. For these two reasons, SLs 1-9 were re-synthesized onto polystyrene cross-linked polystyrene (PS) beads, and were used in the majority of the experiments subsequently discussed. Another big difference between TentaGel beads and PS beads is that PS beads have 4x more functionalization, or primary amines available, than the TentaGel beads. This property was taken advantage of in Chapter 3, when the SLs on PS beads were used for solid phase extraction, but that discussion will be reserved for the next chapter. The differences between these two types of resins will be further analyzed in later chapters as well.

Table 2.1 Resin chart by Rapp Polymer. TentaGel vs Polystyrene resin relative size increase (swelling) in different solvent.

resin \ solvent	Water	MeOH	EtOH	CH ₂ Cl ₂	Toluene	DMF
Polystyrene 1% DVB	--	1.6	1.7	7.5	7.5	4.1
TentaGel XV 0.2 - 0.4 mmol/g	3.6	6.2	2.2	18.0	12.6	13.2
TentaGel R 0.18 - 0.23 mmol/g	2.9	3.7	1.8	6.8	3.4	5.0
TentaGel S 0.25 - 0.3 mmol/g	3.6	3.6	2.9	6.3	4.8	4.7

The PS beads, being hydrophobic in an aqueous environment, still have a slight tendency to clump, and therefore possess a potential to clog the flow cytometer. For this reason, it became necessary to sonicate all SLs prior to incubating them with glycoprotein samples. This was also done to ensure a more homogeneous interaction between single beads and glycoproteins. The samples were then sonicated a second time, directly before injecting them into the flow cytometer, to break up any bead clumps so the beads would register as single events. The latter reason is necessary because each data point must be

an individual bead with its own fluorescent intensity. Otherwise, it cannot be used as a data point in subsequent analysis because it doesn't accurately reflect the binding relationship between the SLs and the glycoprotein extract, since each bead in the sample is technically a replicate of the next.

Any multiple beads stuck together could be later identified and removed from the data set by comparing the front scattering and side scattering information of that event provided by the flow cytometer. Shown in Figure 2.1 below, if the front scattering to side scattering ratio was not roughly equal to 1, then either two beads were stuck together and showed a deviation from the average particle size, or two beads were stuck together and showed a deviation from the granularity of the average bead.

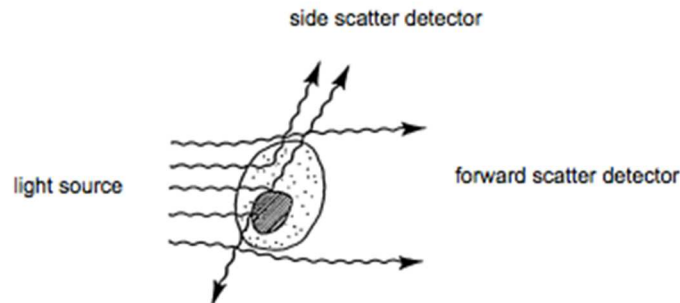


Figure 2.1 Registering an event by a flow cytometer. Front scattering indicated size of particle, side scattering indicates granularity of particle.

The benefits of sonication showed an immediate payoff as the number of events registered as single beads went from 60-70% to 80-90%. This in turn, ended up being another way to ensure that the data set collected weren't unintentionally biased, because beads that stick together may have a physiological reason as to why they do so, namely that they have glycoproteins bound to them, and should therefore not be excluded from data sets. Since, the larger 300 μ m beads could not be sonicated due to their high tendency

to shred apart, any beads that were touching one another during the imaging step of the fluorescent microscope protocol, were generally excluded from subsequent data analysis because they were rejected as data points by the MatLab algorithm. This added the potential of excluding important data points, which could be used to gain a better understanding of the overall binding interactions between SLs and cell line extracts.

OPTIMIZING THE SAMPLE BUFFER FOR THE FLOW SYSTEM

Sonicated the samples prior to incubation and injection helped the flow rates tremendously, but clogging the flow cytometer was still a major concern. When consulting other flow cytometry practices, the biotech company BD sold polystyrene size calibration beads that were suspended in PBS 1X with 2% Bovine Serum Albumin (BSA). Albumin is the main blood protein found in mammals and has high aqueous solubility for a non-glycosylated protein, and was being used as a surfactant for these polystyrene bead suspensions. The addition of 2% BSA to the sample buffer seemed entirely reasonable and of little consequence because the typical range of human albumin in human blood is around 3-5g/dL, which is twice as high as this suggested 2% BSA in the buffer. Furthermore, BSA was used as a component in the blocking buffer (PBS_L, 2% BSA, 10% glycerol) to reduce any nonspecific protein-peptide interactions, prior to incubating the SLs with the glycoprotein extracts. Therefore, it was decided to add 2% BSA to all sample buffers, including the incubation buffer and washing buffer for consistency, to help as a surfactant. This addition of 2% BSA lowered the overall fluorescent intensities of the samples, but blocking these nonspecific interactions was a way to ensure that the SL array was more focused on binding to the glycan moieties of

the glycoproteins, instead of relying on the hydrophobic interactions that could occur between the polystyrene beads, or 12-mers on the beads, and hydrophobic protein moieties.

Another additional precaution was taken to reduce the potential of clogging the flow cytometer, which was the use of 40 μ m filters specifically designed for samples about to be injected into a flow cytometer. These 1mL pipette tip filters, provided by FlowMi, retained any clumps, or aggregated cell debris, larger than 40 μ m that may clog the lines of the flow cytometer. These filter tips were used for all samples for the next dozen experiments, however the cost of using these with the large number of samples in each array essentially doubled the cost of each experiment, and eventually had to be abandoned, despite the consistent 98% of beads registering as single events. The flow cytometry sample preparation filter tips will be highly beneficial when the SL array is ready for the clinical setting. However, during the research and development stages, the cost of these filters does not outweigh the benefit of only adding 10% more single events per sample.

Having made one change to the sample buffer that drastically improved the flow rate of the samples, a further investigation into adding other components to the buffer system was conducted to see if the protocol could be further optimized without any drawbacks. To do this, PS beads were suspended in different solutions and looked at under a microscope to visually assessing whether certain solutions promoted or prevented bead aggregation. For buffers that are compatible within a flow cytometer, the beads appeared the most homogeneously suspended in PBS_L with 2%BSA and 10%ethanol, and in PBS_L with 2%BSA and 10%glycerol. However, glycerol is the other component

used in the blocking buffer for the SLs, which is used to limit nonspecific or weak binding interactions occurring between the SLs and glycans in the sample. The addition of glycerol to the incubation buffer has been shown to drastically reduce the fluorescent intensities of the samples, and was therefore not suitable to be included in the sample buffer. As a result, ethanol was selected as the additive to the buffer system. It was speculated that the hydroxyl group present on ethanol helped to reduce the stickiness between beads that are coated in carbohydrates, which also include hydroxyl groups, and therefore resulted in a more homogeneous suspension and better sample flow rate.

NOTE: Due to the lower potential of clogging, and the lower workload at the University of South Carolina, the flow cytometer instrument used on this project was changed from the BD LSRII to the Beckmann-Coulter FC500. The FC500 is an older and more robust instrument, and is less likely to clog than the BD LSRII because the sample tube runs vertically rather than horizontally.

The next goal was to determine ideal percentage of ethanol in the sample buffer. For this, a small array of two cell lines by three SLs was used, while varying the percentage of ethanol, in PBS with 2%BSA, from 0% to 10%. The best buffer condition was determined by both the flow rate, in events/sec, and by the sharpness of the fluorescent peaks, which is representative of the standard deviation of all the beads in each sample. The sharpest signals with the best flow rates were the samples with 2%EtOH and 5%EtOH, so the smaller concentration of ethanol that was chosen.

This new sample buffer and then compared to the old sample buffer of PBS_L alone, as shown in Figure 2.2 below. This comparative study was used to demonstrate that the sample buffer did not have a major impact on the data sets, but actually

demonstrated that the PBS_L, 2%BSA, 2%EtOH buffer resulted in a lower percent standard deviation of the signals within the samples. This lower %ST DEV was thought to be attributed to the consistent use of 2%BSA throughout the entire assay, which limited any dynamic interactions that could still be occurring during the analysis step, when only using BSA in the blocking buffer and not in the incubation and wash buffers. From this point onward, all samples were incubated, washed, and injected into the flow cytometer in PBS_L, 2%BSA, 2%EtOH.

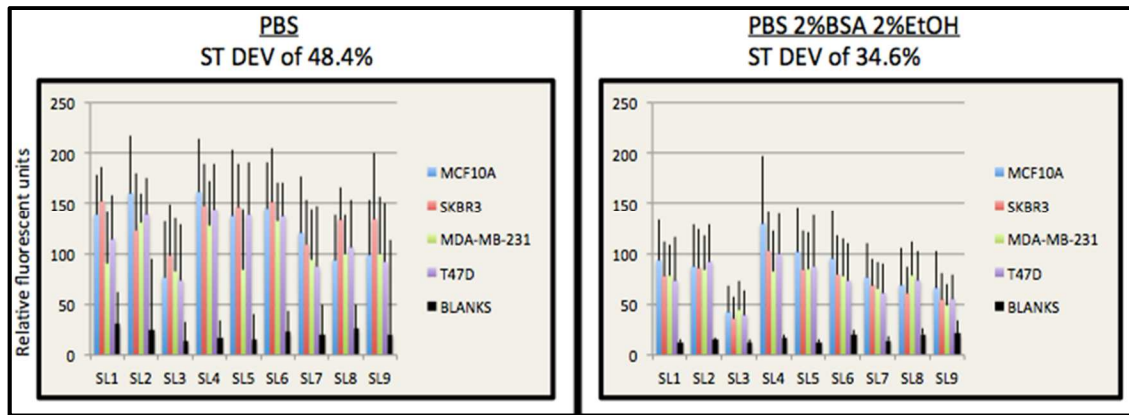


Figure 2.2 The percent standard deviations of samples in PBS (left) is higher than the percent standard deviation of the sample in PBS 2%BSA, 2%EtOH (right).

Once a suitable buffer and sample preparation protocol was in place, data was collected on the 4 breast cell lines, MCF10A, SKBR3, MDA-MB-231, and T47D, under these conditions. The flow cytometer was programmed to collect 250 beads per sample, which was determined to be a sufficient size data set by a collaborating statistician. The cell line MCF7 was discontinued for this research project due to the ambiguity in the publications regarding the nature of that cell line. The cell line BT474 was temporarily discontinued due to difficulties in cell culture, but was reintroduced in later experiments.

PROCESSING DATA WITH THE MATLAB ALGORITHM

Initially, the large data sets acquired from the flow cytometer generated a lot of time consuming and laborious data processing. For the previous microscope protocol, a MatLab algorithm had been written to analyze the photographs taken of the beads under the fluorescent microscope. This algorithm selected single circular beads in the region-of-interest and gave them an average intensity in the red, green, blue, and rgb channels. This code could not be used for the flow cytometer data because the inputs were not images and the numbers (Fluorescent intensity, side scattering, and front scattering) were on a 10-bit scale (0-1023) instead of an 8-bit scale. Furthermore, the raw flow cytometry data from the FC500 could only be extracted by plotting all the beads in a sample (n=250) in a histogram, with fluorescent intensity, from 0 to 1023, on the x-axis and a count of the beads, in that integer, on the y-axis. From here, the data points could be extracted and imported into Microsoft Excel, however, the data extracted from these graphs was *binned* (Integer from 0-1023) data because it came from a histogram.

Therefore, between the unbinning of the data, the outlier removal, and the array formatting, it became clear that a new MatLab algorithm was needed to process these large data sets. I wrote a new MatLab code that was able to process the data from each cell line, including all nine SLs in the array, in less than a minute. This was a vast improvement in the data throughput of the flow cytometry protocol, which previously took 45 minutes to manually process the data from each cell line. The code for this MatLab program is provided in Appendix A, in Figure A.1.

NOTE: The MatLab code was written in MatLab R2014b. Certain functions, such as the XLR_WRITE function, are not supported by Apple users for this version. Therefore, the code is only compatible with Microsoft Excel supported on a PC.

REMOVING OUTLIERS BY INTERQUARTILE DISTANCE RELATION

Prior to analyzing the data sets, each data point was subjected to an outlier removal step using interquartile distance (IQD) computation²⁶, which reveals which data points are most likely to be outliers, and therefore removable from the data set. This is achieved by computing the interquartile range (IQR) of a data set by subtracting the Q1 from Q3, which represent the first quarter (25th percentile) and the third quarter (75th percentile). From there, all data points outside the range of either, $(Q1-IQD*IQR)$ or $(Q3+IQD*IQR)$ are considered to be a probable outlier and are removed from the data set. The IQD value is the multiplier of the IQR, and was previously used at 1.5. However, once the data sets became much larger for the flow cytometry protocol and the FlowMi filters became too expensive, the IQD value was lowered to 1.0, to tighten the data set and effectively remove those data points that would have been removed by filtration.

USING LINEAR DISCRIMINANT ANALYSIS IN ASSESSING THE SL ARRAY

The final step in the data analysis process was determining whether the SL array was useful in discriminating between cancerous and noncancerous glycoprotein extracts. In order to accomplish this, the program Systat was used for its linear discriminant analysis (LDA)²⁷ capabilities, which acts to statistically maximize the differences between data sets, while minimizing the differences within a data set. This statistical method only

works in an array format, meaning the data must be in the shape of a rectangle, or each column (SL) must have the same number of data points (beads), which was also done by the MatLab algorithm previously described. However, if one SL data set was missing for a cell line, then that SL has to be excluded from all the other cell lines as well to keep the data in an array format. By redefining the variables (Cell lines) of the data set, the LDA could also be used to discriminate between disease states of different metastatic potential, or the aggressiveness of the cancers. This was often done with the data sets to highlight the differences between the healthy cell line (MCF10A), the lowly metastatic cell lines (SKBR3 and T47D), and the highly metastatic (MDA-MB-231).

OPTIMIZING THE DATA COLLECTION PARAMETERS OF THE FLOW CYTOMETER

One added benefit of using a more sensitive platform was being able to detect the autofluorescence of the SLs alone, or have a negative, unincubated, control for each SL, which was never done using the microscope protocol. This allowed for a more direct comparison of the overall binding capacity of each SL, because each SL inherently has a slightly different autofluorescent intensity. However, upon analyzing the flow cytometer data sets, it was noted that the fluorescent intensities of the blank SLs were registering in the 500-600 RFU range, which is 8-bits on a 10-bit scale. These negative controls still had an overall RFU signal less than the positive samples, but the autofluorescence of the blank SLs were far too bright for polystyrene beads with amino acids on them. This peculiarity unveiled a preset on the flow cytometer that was set to convert the data into a logarithmic scale, which is a commonly used preset that emphasizes the first decade of

the 10-bit scale, and can help highlight minute changes on the lower end of the detection threshold. Fortunately, whether the data was collected in linear or logarithmic form, it did not have an impact on the LDA discrimination. However, the logarithmic data was far less informative than the raw data because it scaled the data in each sample, making it challenging to compare the individual SLs to each other. Therefore, all data subsequently collected was done so in linear form. Once that was corrected, the blank SLs had an appropriately dim autofluorescence of around 4-20 RFU, and became useful negative controls, particularly in determining the ideal incubation conditions and the threshold of detection for the assay.

OPTIMIZING THE PMT DETECTOR VOLTAGE AND PROTEIN CONCENTRATION

One parameter of the flow cytometer that had not been considered up until this point was the voltages of the photomultiplier tubes (PMT) of the fluorescence detectors on the flow cytometer. The voltage of the PMT detector acts similarly to the gain of a signal, which means higher voltage results in more sensitive fluorescent detection. Since fluorocein was the only fluorophore used in this project, the FL1 detector on the FC500 (A band pass filter centered at 525nm +/- 25nm) was the only channel used.

Up until this point, the FL1 detector had been defaulted to the lowest setting of 250 voltage units, but the potential usefulness of this parameter was explored by performing a standard assay of SL1-9 by the four breast cell lines (0.0002mg of glycoprotein extract), but collecting the data using three different PMT voltages, FL1(250), FL1(325), and FL1(425), all in linear form, as shown in Figure 2.3.

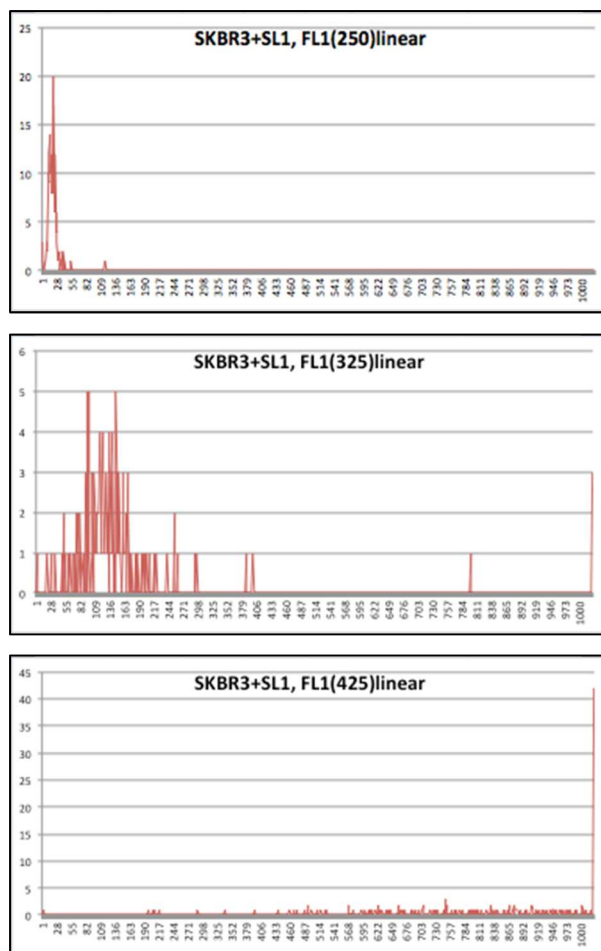


Figure 2.3 SL1 and SKBR3, data collected with various PMT voltages on FL1 channel

The conclusion of this experiment was that the voltage of the PMT detector has a huge impact on the signal intensities. The FL1(250) setting seemed to be too low of a setting for such a low protein concentration, because the positive signals were barely above the blank SL signals. However, the FL1(425) setting was too high of a voltage setting because the signal was almost completely saturated, meaning most of the beads registered at 1023, and would have to be excluded from the data set. This assay suggested that the FL1(325) voltage setting was ideal for this protein concentration of 0.0002mg/mL, which had been used throughout the project. However, when assessing the

signal at the lowest PMT voltage setting, it became apparent that the fluorescence signals were hovering around the limit of detection, which for biological samples is generally accepted to be three times the background, or noise, of the system.

Therefore, a new experiment was performed in order to examine the relationship between the PMT voltage and the protein concentration, and to see if a higher protein concentration would aid in the ability of the SL array to discriminate the extracts. For this, the same SLs and cell line extracts were used as the previous assay, but a much higher protein concentration of 0.01mg/ml was used, which was 50x higher than the laboratory protocol. At this concentration, the sample signals quickly saturated at a PMT voltage of 275 or higher. This demonstrated that the FL1 PMT voltage parameter had to be carefully balanced with the protein concentration incubated with the SLs, to ensure the best signal.

The next experiment to further explore this relationship between PMT voltage and protein concentration, was performed using a more reasonable protein concentration of 0.001mg/ml, which was only 5x higher than the laboratory protocol. The same data set was recorded at PMT voltage settings of 250 and 275, but even this incremental increase in voltage resulted in a 5% drop in the overall classification accuracy, as shown in Figure 2.4. This phenomenon was similar to what was seen when the protein concentration within a sample was too high, namely that the standard deviation of the signal began to increase, which also led to a lower overall percent discrimination. This is shown in Figure 2.5 below, where the protein concentration of two cell lines SKBR3 and MDA-MB-231, labeled in a 2:1, FITC:lysine ratio as previously described, was varied from 0.002nmol/ml to 0.5nmol/ml, while keeping the bead number constant. This assay also

demonstrated that too high of a protein concentration correlating with a larger percent standard deviation in the signal intensities. In fact, any protein concentration higher than 0.001mg/ml showed a significant increase in the standard deviation of the beads. For these reasons, a protein concentration of 0.0005mg/ml was chosen because it consistently held a signal significantly above the blank beads while maintaining a low standard deviation within the sample.

After a few more experiments with the new protein concentration of 0.0005mg/mL, an appropriate FL1 PMT voltage of 300 was chosen. These parameters were chosen for all future experiments because the positive signals were above the limit of detection, defined as $LOD = \text{blank SL} + 3\sigma$, whereas σ is the average standard deviation of the blanks SLs, or the noise of the negative controls. Furthermore, the system parameters needed to remain constant in order to compare data sets to each other.

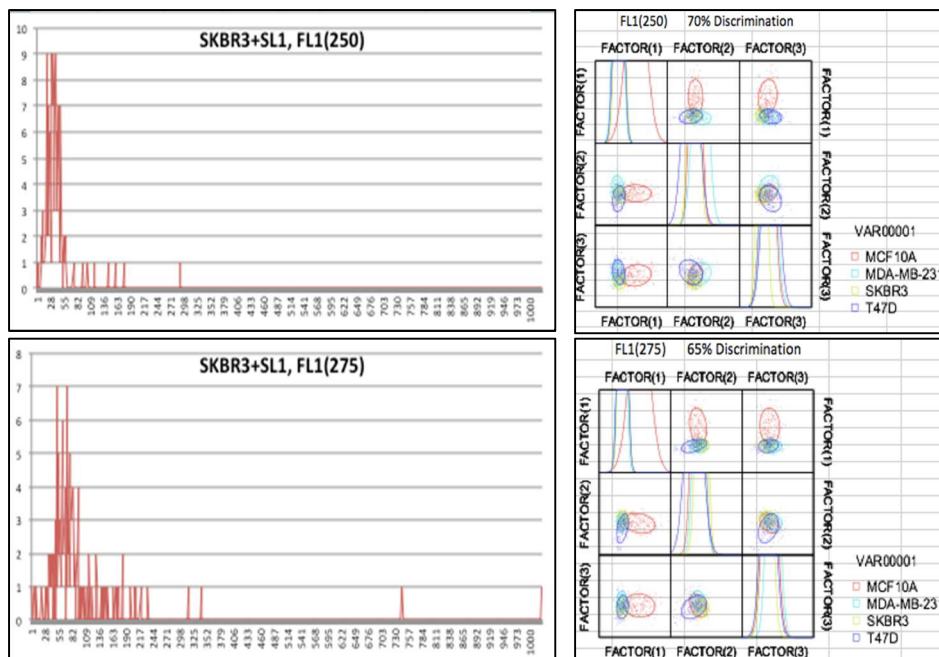


Figure 2.4 Flow cytometry histograms (left), increasing PMT voltage causes LDA classification accuracy (right) to drops from 70% to 65%.

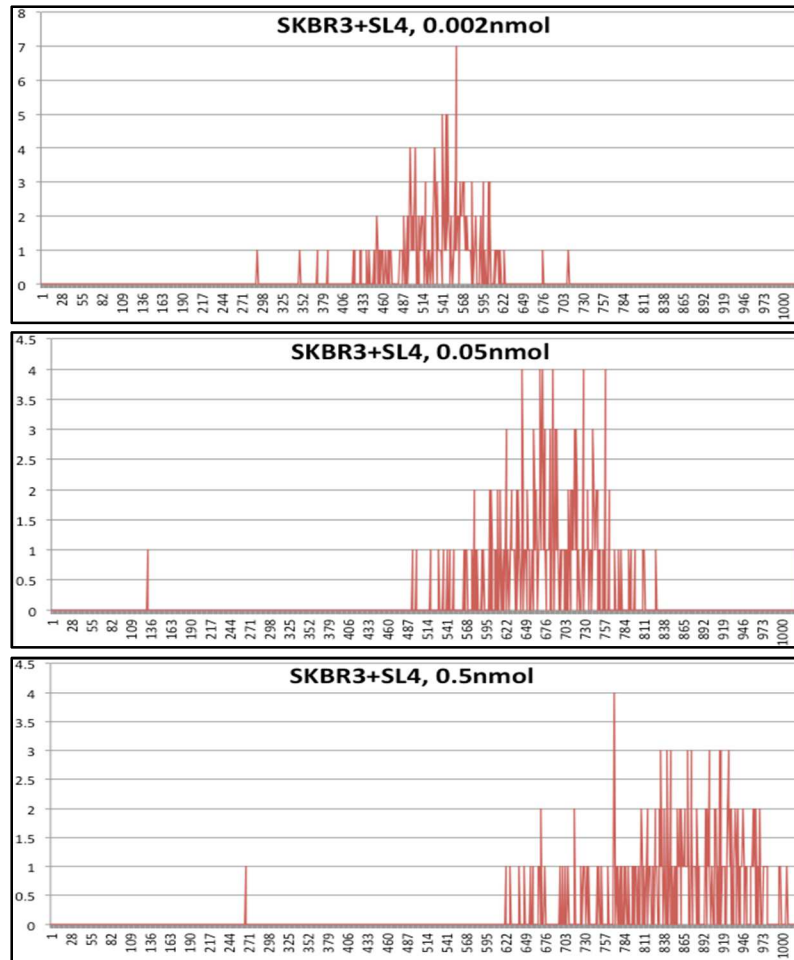


Figure 2.5 Larger standard deviation of signal can be caused by increase in protein concentration

ASSESSING THE REPRODUCIBILITY OF THE FLOW SYSTEM

Having recently changed several variables in the protocol, the next step in validating the flow cytometer platform and all the adjustments made thus far, was to pick a condition and perform technical replicates of SLs1-9 by the four breast cell lines, and to compare the data sets to each other for reproducibility. As mentioned before, the original SLs1-9 were chosen for their ability to differentially bind purified glycoproteins. This, along with data that will be presented at the end of this chapter, are strong affirmations

for the reproducibility of the SLs array when binding to purified proteins. However, the story changes when dealing with complex biological samples that include a plethora of glycoproteins of all different shapes and sizes, all of which interact differently with the cross-reactive sensor array. Therefore, reproducibility of the sensor array has been a major concern since the beginning of the project, and will continue to be as the project moves towards using more complex biological serums, such as patient blood.

Forgetting the instrumentation, the buffers, and the data analysis for a minute, the foundation, and most important relationship is between the amount of labeled glycoprotein and the number of beads in the sample during the incubation. In thinking about this important relationship, the last variable that had not yet been taken into account was the amount of beads in the samples. Taking a step back and looking at the microscope protocol, the first step of the established protocol was to weigh out 2mg of each SL, (Synthesized on 300 μ m TentaGel beads), which are then blocked with PBS_L, 1%BSA, 10%glycerol, before being individually incubated with each cell line glycoprotein extract overnight. However, the act of distributing 2mg of beads among the array was actually much more challenging in practice because the resins, when dry, are very electrostatic and difficult to contain. Therefore, weighing and distributing the 300 μ m TentaGel beads among the array as a wet slurry became common practice. Since the water content of a 2mg mass is hard to account for, and the TentaGel beads swell 3-4x in aqueous solution, this practice likely introduced error into the array format because the amount of beads across the array was assumed to be the same. This is of the utmost importance in an array format because any unaccounted for inconsistencies among the array, could be reflected in overly optimistic LDA classification accuracy.

For example, if Cell Line A is incubated with an aliquot of SL1, which has a certain amount of beads. While Cell Line B, is incubated with an aliquot of SL1 that has half as many beads. Then SL1 will appear to have a stronger preference for Cell Line B (The beads will appear brighter) because the ratio of glycoprotein to bead binding capacity has shifted. Therefore, the amount of beads in a sample, or the overall binding surface area available on the beads within the sample will have a direct impact on the fluorescent intensity of the individual beads.

This was demonstrated in an assay, shown in Figure 2.6, where the fluorescent intensity of two cell line extracts, MCF10A and SKBR3, was inversely related with the amount of polystyrene beads in the sample when keeping the protein concentration constant (0.0002mg/ml). The signal of the samples were no longer significantly above the blank SLs when ~600,000 beads were incubated with 0.0002mg of cell line extract. This data combined with the following experiment were evidence that a more stringent bead distribution protocol was needed for this project to ensure a higher level of confidence.

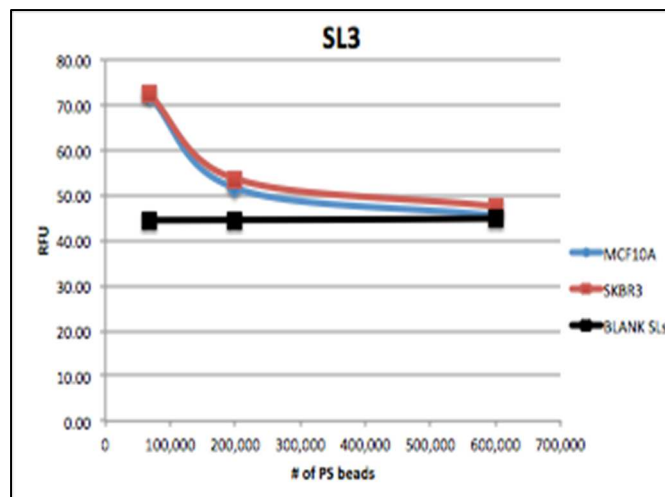


Figure 2.6 SL3 with MCF10A and SKBR3, bead number vs. fluorescent intensity.

In taking another step back, the initial bead distribution method for the flow cytometry protocol was to pipette 1mg of wet slurry of PS beads across the array. The first indication that this practice was not a consistent method, was seen when comparing the reproducibility of three technical replicate data sets. These three data sets are shown in Figure 2.7 below, where the 4-class LDA results of the three technical replicate were 86%, 69%, and 80%.

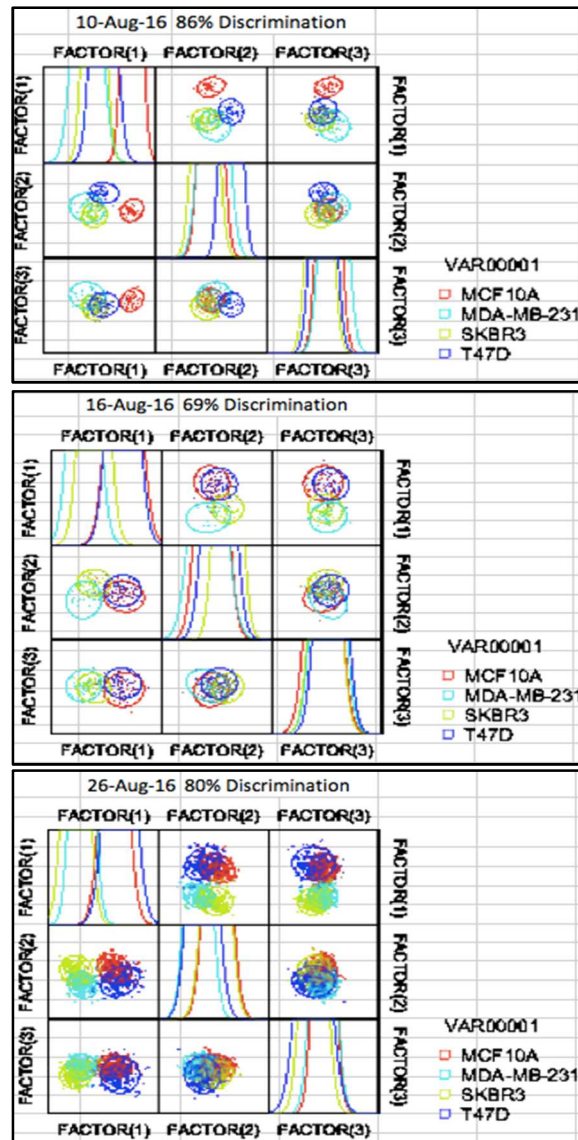


Figure 2.7 LDA results of three technical replicates of SL1-9 by 4 breast cell lines

This triple replicates showed promising results in terms of discriminating the four cell lines, or breast cancer subtypes, in relation to each other, especially since a 4-class LDA of randomly generated data points would result in ~25% accuracy. The cell lines MCF10A and T47D overlapped more in two of the three data sets, while the cell lines SKBR3 and MDA-MB-231 overlapped more significantly in all three runs. However, despite the accuracy being 3x better than random, the three data sets had 17% variability in discrimination accuracy, which was not as consistent as what had been previously seen when using the microscope protocol. Furthermore, these discrimination accuracies were lower than previously members of the laboratory had observed using the microscope protocol, whom reported discrimination accuracies >90%.

In addition to the overall inconsistencies in the LDA classification accuracies, there were also some reservations about the reproducibility of the individual SLs. For instance, certain cell lines were observed to have a high affinity for a certain SL on one day, but did not show that same affinity on a different day. These SL anomalies led to large impacts on the percent discrimination results from LDA, as shown in Figure 2.8. Here, in an assay that was performed earlier while still discerning of the limit of detection, the cell lines SKBR3, T47D, and MDA-MB-231 have disproportionately high affinities for SL2, SL3, and SL5, respectively. Drastic disparities between SLs, such as these, were not uncommon in previous assays and usually resulted in a very high LDA discrimination accuracy, such as 94%. The F-to-remove factors shown in the bottom right of Figure 2.8, reveal that SL2, SL3, and SL5 had the biggest impact on the discrimination accuracy. Therefore, there was still an important variable that had not been taken into account yet, which was likely causing these inconsistencies.

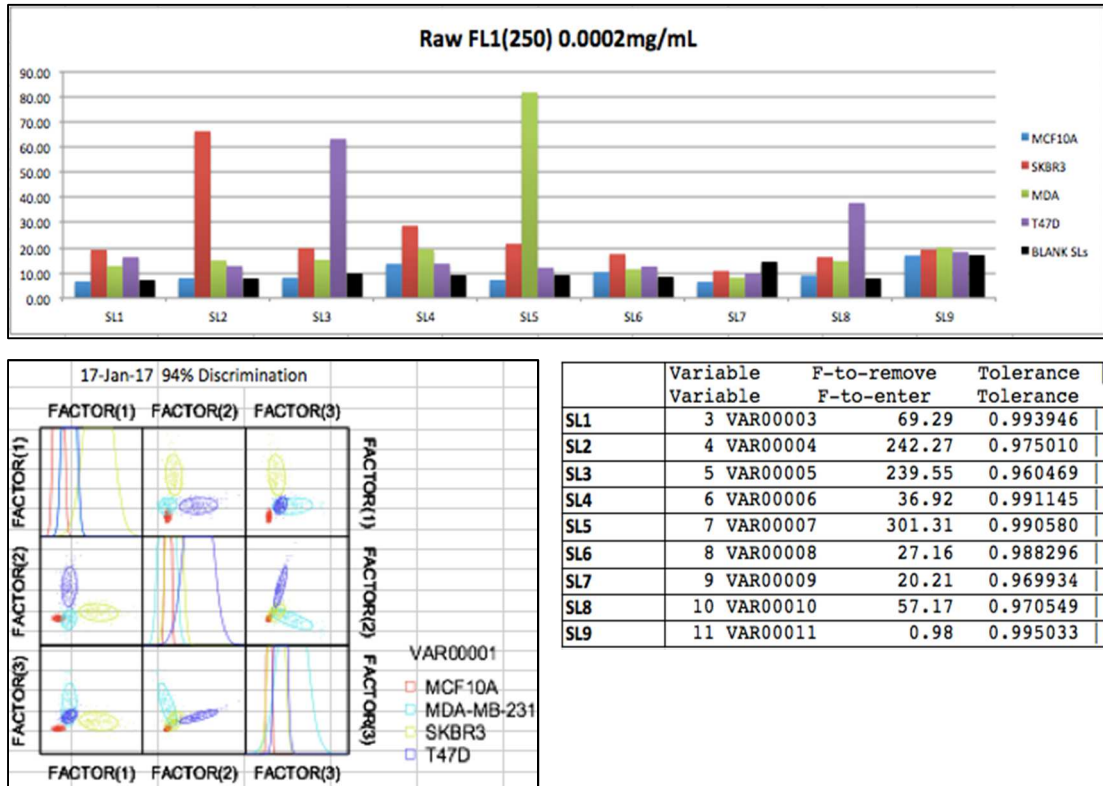


Figure 2.8 Large differences in SLs have an impact on LDA discrimination accuracy.

Having focused heavily on the protein labeling technique, the protein concentration, and the PMT detector voltage, the last variable to consider was the number of beads used throughout the array. The microscope protocol used aliquots of ~2mg of 300µm TentaGel beads, which were distributed as a slurry among the array from a bead stock suspended in PBS_L, due to the electrostatic difficulties previously alluded to. Since the beads are large enough to be visible to the eye, the variability between these estimated ~2mg slurries, and the previously discussed error introduced by this method, were either not taken into account or did not appear to have an impact on the data. However, the variability in the ~1mg of 10µm PS beads being distributing across the array in the flow cytometry protocol, was having a larger impact on the data and had to

be corrected in order to observe a more accurate binding interaction between the SLs and the cell lines extracts.

In addition to these inconsistent SL anomalies, the bead concentration problem was also apparent by the large variability in the time it took the flow cytometer to collect 250 data points from each sample. Since the flow cytometer is set to collect 1 μ l of sample per second, and each SL has the same relative density, the time it takes to analyze each sample should be relatively constant across the array. Therefore, the variability in the time it took to process 250 data points for the array was a good indication of the variability in the bead population across the array, which also correlated to artificial discrimination potential due to experimental error. This bead distribution problem was corrected by creating homogenous SL stocks of the 10 μ m PS beads, so that the aliquots could be distributed among the array more accurately by pipet.

OPTIMIZING THE BEAD CONCENTRATION FOR THE FLOW SYSTEM

In order to overcome this problem of even bead distribution, each SL stocks was dried using vacuum centrifugation, washed thoroughly with several aliquots of acetone to remove any remaining buffer, and re-dried again in the vacuum centrifuge. The theoretical mass of each SL, fully functionalized with the modified 12-mers, was calculated. Due to the extensive functionalization throughout the PS beads, 1.25mmol/g, the SLs, each with a slightly different mass due to the differences in amino acid content, were roughly 3.5x the mass of the blank PS beads. Since the PS beads swell in N,N-dimethylformamide and the porosity of the resin is large enough for a single amino acid to penetrate into the center of the bead, it was assumed that the PS beads were fully

functionalized throughout, as an 8x excess of each amino acid was used during the synthesis process.

Using these numbers, the mass of 19.5 million beads (1mg of blank 10µm beads) of each SL was calculated, weighed out from the dried stocks, and resuspended in 1ml of 75% glycerol/25% PBS. The high percentage of glycerol content was used to get the beads into a not-to-viscous suspension that was roughly equivalent to the density of the amino-acid functionalized polystyrene resin. Roughly matching this density with glycerol was a crucial to ensure even distribution, so that upon vortexing the SL stocks, the suspension would remain homogenous, and the beads would not fall out of solution during the distribution of the aliquots among the array. In order to get the 10µm PS bead SL stocks homogenous enough to pipet, they needed to be heavily vortexed and sonicated until no clumps were visible. After this, the beads were visualized under a microscope afterwards to ensure they had not fractured by this step, which they never did.

After creating this homogeneous suspension for each SL stocks, the standard deviation in the time it took to collect 250 data points from each sample (1µl/sec collection setting on the FC500), dropped dramatically, as shown in Table 2.2 below in another triple replicate of SLs1-9 and the same cell line extracts.

Table 2.2 Percent standard deviations of 250 data points collected (sec) in three assays by FC500 flow cytometer, set to collect at a rate of 1µl/sec.

	Average	% ST DEV		Average	% ST DEV		Average	% ST DEV
SL1	2.5	8.0	SL1	2.5	8.2	SL1	2.2	19.2
SL2	2.7	16.9	SL2	2.6	5.8	SL2	2.8	21.6
SL3	9.6	6.8	SL3	9.0	8.0	SL3	7.5	16.9
SL4	2.7	7.8	SL4	2.2	22.5	SL4	2.2	27.4
SL5	2.5	57.5	SL5	2.8	11.0	SL5	2.8	8.3
SL6	3.0	8.8	SL6	2.1	12.6	SL6	2.3	19.0
SL7	3.3	7.7	SL7	3.6	45.7	SL7	2.8	11.3
SL8	3.4	12.4	SL8	2.0	5.7	SL8	2.1	2.7
SL9	4.1	14.6	SL9	3.8	29.7	SL9	3.3	33.7
	3.1	15.6		3.4	16.6		3.1	17.8

These SL stocks proved to be a vast improvement in the consistency of the bead populations across the array, and allowed for a more confident and direct comparison between cell lines and SLs. These more accurate representations of the SLs also allowed more confidence in determining which SLs were important and which SLs seemed redundant or overly cross-reactive. This feedback is important in developing the best array, which also includes eliminating useless or convoluted SLs from the array.

CALCULATING THE BINDING CAPACITY OF THE BEADS

The 10 μ m PS beads SL stock concentration was made to roughly approximate the binding capacity of the 2mg of 300 μ m TentaGel beads from the microscope protocol. This was calculated based off of the size, binding capacity, and surface area to volume ratio of the resins used. The surface area to volume ratio is important because despite the resin's overall size, it has the same roughly uniform porosity. This porosity, for both TentaGel and PS beads, is large enough for single amino acids to pass through, but not for entire glycoproteins to pass through. Therefore, only the surface areas of the beads are available to bind with the cell line glycoprotein extracts.

The 300 μ m TentaGel beads had a binding capacity of 0.28mmol/g. The protocol called for 2mg of beads, which was really only 1.43mg of beads after the mass added by functionalization. The 1.43mg of TentaGel beads has a binding capacity of 0.56 μ mol, only 2% of which is located on the surface of a 300 μ m sphere ($SA/V = 4\pi r^2 / (4/3\pi r^3) = 3/r$). Therefore, the overall binding capacity of the TentaGel beads, per sample, in the microscope protocol was 0.008 μ mol. Comparing this to TentaGel beads of 10 μ m

diameter ($1.95E^9$ beads/g determined by manufacturer), which are much smaller spheres and therefore have a larger surface area to volume ratio of 60%, 100,000 beads would equate to $0.0086\mu\text{mol}$ of binding capacity.

Initially though, these calculations were not done and were only taken into account when it was noticed that the overall fluorescent intensity of the beads in the microscope platform were much brighter than the beads registered by the flow cytometer platform. The beads in the microscope protocol were averaging around 150 RFU, or 7 bits on an 8-bit scale, whereas the beads in the flow cytometry protocol were registering around 50 RFU, or 6 bits on a 10-bit scale.

At this time, $10\mu\text{l}$ of the $10\mu\text{m}$ PS bead SL stocks was being used per sample in the flow cytometry protocol, which equated to $\sim 200,000$ beads per incubation. PS beads are also 4x more amino functionalized than TentaGel beads (1mmol/g compared to 0.28mmol/g), and therefore the total binding capacity available in the flow cytometry samples was $0.0615\mu\text{mol}$, which is about 7.5x higher than the binding capacity available ($0.008\mu\text{mol}$) in the microscope samples. This explains, at least in part, why the individual beads appeared dimmer when using the flow cytometry protocol because the same amount of glycoprotein had 7.5x as much surface area to interact with, which in turn, spread out the glycoprotein more sparsely among the larger number of beads.

RE-ASSESSING THE REPRODUCIBILITY OF THE FLOW SYSTEM

Once the SL stocks were made and the 200,000beads/sample method was still in place, another three technical replicate assays of SL1-9 by 4 breast cell lines glycoprotein extracts (0.0005mg , data collected at voltage 300 on FL1 PMT detector) were performed

to reassess the reproducibility of the newer and more stringent protocol. The results, shown in Figure 2.9 on the following page, were much more consistent and had a deviation in the classification accuracy of only 1% in the 4-class LDA, and only 0.33% in the 3-class LDA. This accomplished the goal of signal reproducibility with the sensor array, however, the average 4-class LDA result was only 45%, and the average 3-class LDA was only 53%. As a reference again, the classification accuracies of a random data in a 4-class LDA and 3-class LDA would be 25% and 33%, respectively. So the results, although low, were almost twice what would be observed randomly. This was encouraging that the SL array worked as it should, but also alluded to the fact that the SL array may need to be ameliorated and further tailored if it is to be used as a breast cancer diagnostic tool.

These numbers were dramatically lower from what had been previously reported on this project, but were the most consistent and reproducible data sets observed since the flow cytometry platform was introduced, and possibly since the beginning of the project. However, these results were not that shocking because they followed the general trend that every time a variable of the project was scrutinized, quantified, and adjusted, the overall classification accuracy of the data set was lowered. This was the case when calculating and keeping constant the FITC:lysine ratio, when fluorescently labeling the cell line extracts. This was also the case when the adjusting the protein levels, from 0.0002mg to 0.0005mg, so that the signals remained significantly above the limit of detection. Finally, counting the beads and consistently distributing them among the array led to overall lower discrimination accuracy as well.

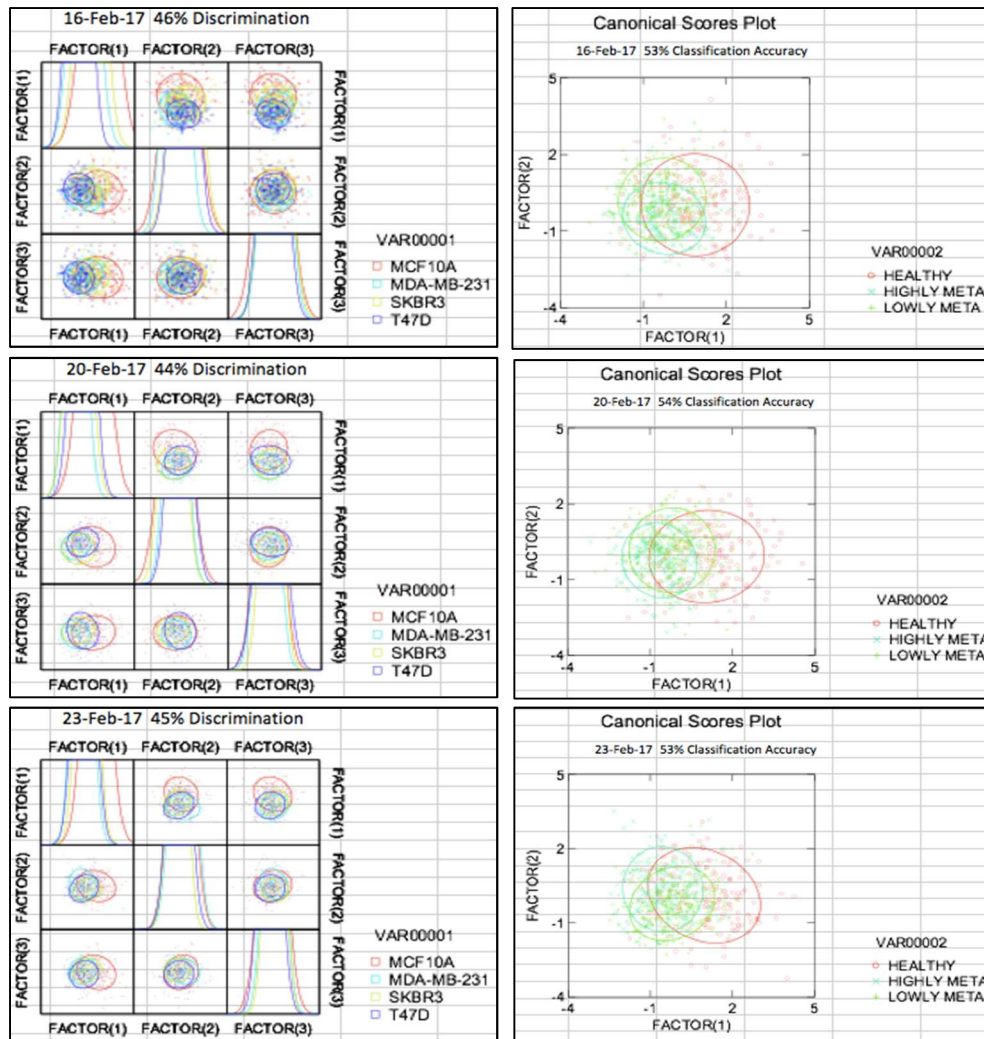


Figure 2.9 Triplicate data shows consistent signal, but lower overall classification accuracy. 4-class LDA, between 4 breast cell lines is on the left, and 3-class LDA, between 3 disease states is on the right

The conclusion of all this is that the adjustments made for the flow cytometer platform were done in order to accommodate a more sensitive and automated system. Switching instrumentation highlighted some of the weaknesses in the existing protocol, which came at a cost of confidence in the SL array. However, the results from this more advanced platform represent, at least in terms of breast cancer diagnostics, a stronger foundation for the true capacity of the SL array. This will allow for a better development of the SL array, and was necessary in order to understand the limitations and expectations

when incorporating this diagnostic tool into the clinical setting. Having made the assertions above, a few more experiments were necessary in order to back up this claim. The most direct of which, was a comparison of the same data set between both platforms.

COMPARING THE FLOW CYTOMETRY DATA TO THE MICROSCOPE DATA

The results from the most recent triplicate experiments were promising because the problem of reproducibility was no longer an issue, but were discouraging because they showed that the SL array's ability to discriminate between breast cell line glycoprotein extracts was not as favorable as previously thought. To make sure this was not caused by some other unknown factor in the instrumentation, a comparison study was done between the flow cytometer and the microscope protocol, using the same samples and having the same user collect the data and perform the analysis. This direct comparison had been avoided until this point because of the difficulty in imaging the 10 μ m hydrophobic beads in aqueous solution, particularly in focusing on the beads suspended in solution. This is much more difficult than focusing on the 300 μ m beads typically used with this platform because the larger beads generally rest directly on the glass slide, which allows them to have a more constant reflection off of the glass slide, and places them at the same focal range. However, it became apparent that this assay would be beneficial to this study by providing evidence to rule out that the instrumentation was having an impact on the LDA discrimination accuracies.

The same incubated samples, SL1-9 by 4 breast cell lines, were incubated under the normal conditions, using 0.0005mg of cell line glycoprotein extracts with 200,000beads (PS) of 10 μ m SLs. The samples were analyzed and collected on both

platforms within four days of each other, first collecting 250 events from the flow cytometer, and then analyzing the remainder of the sample under the microscope. The data sets between the microscope and the flow cytometer, in Figure 2.10 below, showed very similar patterns in the raw data, which translated to the LDA graphs as well. Despite the difficulty in imaging the 10 μ m beads, which was apparent by the small number of data points collected per sample (Average n=28 for microscope data, and average n=225 for flow cytometer data) the 4-class LDA classification accuracy was within one percent, with 52% and 53% for the microscope and flow cytometer data, respectively. The cell line glycoprotein extracts discrimination patterns almost completely overlap when one of the data sets is flipped along the horizontal axis, or multiplied by negative one. In both LDA graphs, the SKBR3 cell line was the most distinguished, while the largest misclassification, or overlap in data points, was between MCF10A and T47D.

These results were reassuring primarily because it provided evidence that the two different platforms, along with the protocol adjustments, were detecting the signals comparably. These results were also beneficial because they highlighted which of the SLs in the array have more reproducible signals, and which SLs in the array do not. In particular, SLs 1,2, 4, and 5 showed the most consistent signals across the platforms, which were also the original SLs created for this project. For this reason, these four SLs, in particular, were the focus of all the experiments discussed in Chapters 3 and 4, which include using these SLs as a solid phase extraction material and N-linked glycan analysis tool that can be used for prognostic and research purposes.

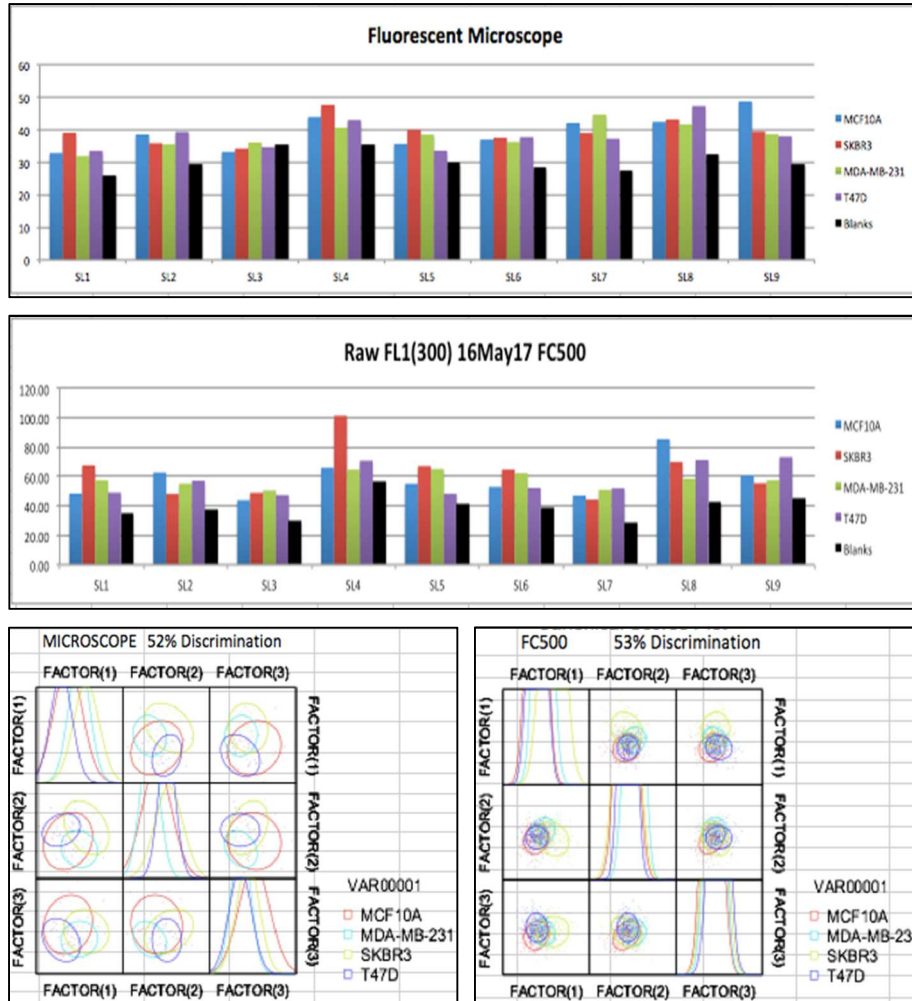


Figure 2.10 Comparison study between microscope and flow cytometer protocols. Raw data(top) and 4-class LDA (bottom).

These consistent results across the platforms left only a few more variables to consider, in searching for the reason why lower discrimination ability was being observed by the SL array. One factor that had to be ruled out, was the error introduced by the user, or the scientist collecting and analyzing the data. Since no other personnel in the laboratory had ever used the flow cytometer protocol, or observed lower classification accuracies, I was the only person who could be held accountable for this user error. Therefore, I performed the microscope protocol using the same exact cell line

glycoprotein extracts that were used in the previous platform comparison study, which were labeled in a 1:1, FITC:lysine ratio, as previously discussed in detail. This was one small difference in this assay from the typical microscope protocol, which does not take the concentration of the glycoprotein extract into account during the fluorescent labeling step, and instead uses a large excess of FITC. The cell line BT474, a Luminal B breast cancer subtype, was included in this study because it was available and could potentially add more insight.

As per the microscope protocol, 2mg of the SLs on 300 μ m TentaGel beads were weighed out as a slurry and incubated under the normal conditions (0.0002mg glycoprotein extract), as previously described. The next day, the beads were washed, and imaged using the fluorescent microscope at an exposure time of 4 seconds. At the time of data collection, it was noted that the 4sec exposure time was a lot longer than what is typically used during imaging. This was likely due to the differences in labeling technique, namely that the prior method in the microscope protocol uses a large excess of FITC in the glycoprotein-labeling step. All other factors were kept constant, including using the same MatLab algorithm that was designed for this protocol.

The results, shown in Figure 2.11, showed that the bead intensities for the samples incubated with fluorescently-labeled glycoprotein extract were barely brighter than the negative controls, or blank SLs. In fact, the blank SLs 1, 5, 6, 7 and 8 all were brighter than some of their corresponding samples incubated with fluorescent glycoprotein. This again, was an indication that the combination of protein concentration and labeling ratio were placing the signal at or below the limit of detection. This anomaly was difficult to compare to past experiments because the autofluorescent signals of the blank SLs were

never collected, or the exposure time was too fast to detect their signals. This extent of glycoprotein labeling will be addressed in the next section. It was also noted that the numbers of data points, or beads, per sample were extremely small for each cell line with an average of 10 data points, which also likely had an impact on the LDA results.

Despite the small data sets and data points having less fluorescence than the blank SLs, the data had a 78% classification accuracy in the 5-class discrimination (5 breast cell lines) and 93% classification accuracy in 3-class discrimination (Healthy, lowly metastatic, highly metastatic). These results were on par with previous experiments run using the fluorescent microscope method, which helped eliminate the user error factor. However, these results were far from promising, especially since the 1:1 FITC:lysine labeling ratio seemed to be way too low of a ratio for the microscope protocol. The next step was to re-examine the labeling ratio to see if that was the last factor that could be having an effect on the LDA discrimination accuracies.

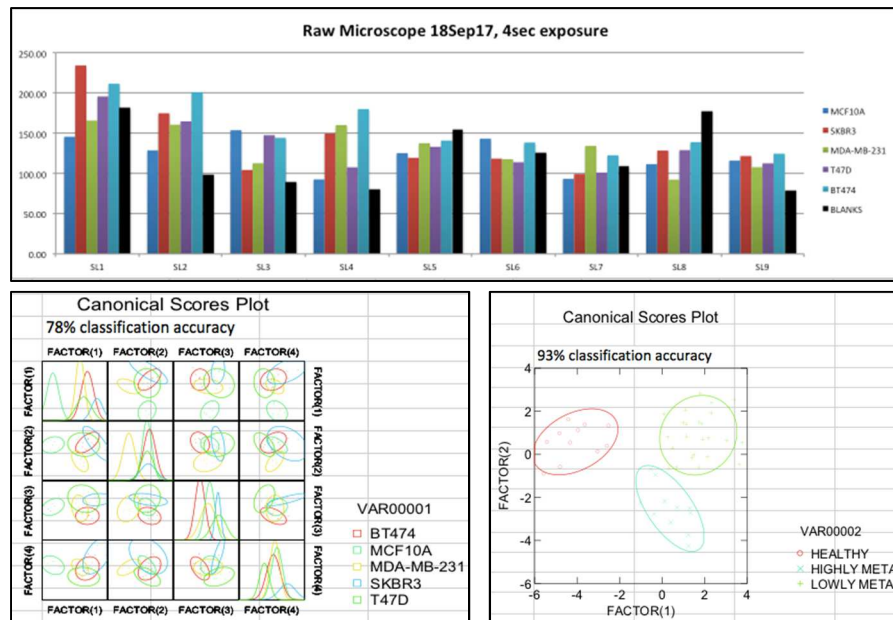


Figure 2.11 Microscope Protocol, SL1-9 by five breast cell line extracts. Raw data (top), 5-class LDA (left) 3-class LDA (right)

RE-EXAMINING THE FITC:LYSINE RATIO DURING EXTRACT LABELING

The extent of fluorescently labeling in the previous microscope experiment demonstrated that the glycoprotein extracts were not sufficiently labeled, as many of the data points had fluorescence dimmer than the blank SLs themselves. To further explore the best FITC:lysine ratio, a few more data sets were run with the cell line extracts fluorescently labeled to different degrees. As mentioned before, the extent of FITC used in the labeling step is based off of a calculation that relies on several assumptions. These assumptions are that the average size of a protein is 60kDa, and that lysine accounts for 7% of the amino acid content on that protein. This calculation was referred to throughout this project in order to keep the ratio between FITC and glycoprotein extracts consistent, which is again very important when trying to discriminate the things that are being individually labeled. Therefore, a follow up experiment was done with 100,000beads/sample of 10 μ m SLs on PS beads with the 5 breast cell line extracts labeled in a 10:1, FITC:lysine ratio to see if the extent of labeling was a big factor to the SL array and LDA discrimination accuracy.

The results, shown in Figure 2.12, were surprising because the raw intensities of the beads were not brighter than the glycoprotein samples labeled with FITC:lysine ratio of 1:1, as anticipated. However, the classification accuracy did become about 10% more accurate than seen before, suggesting that an excess of FITC may help to more completely fluorescently tag the glycoprotein extracts, making the signals easier to detect, but somehow, without changing the overall raw intensity of the signals.

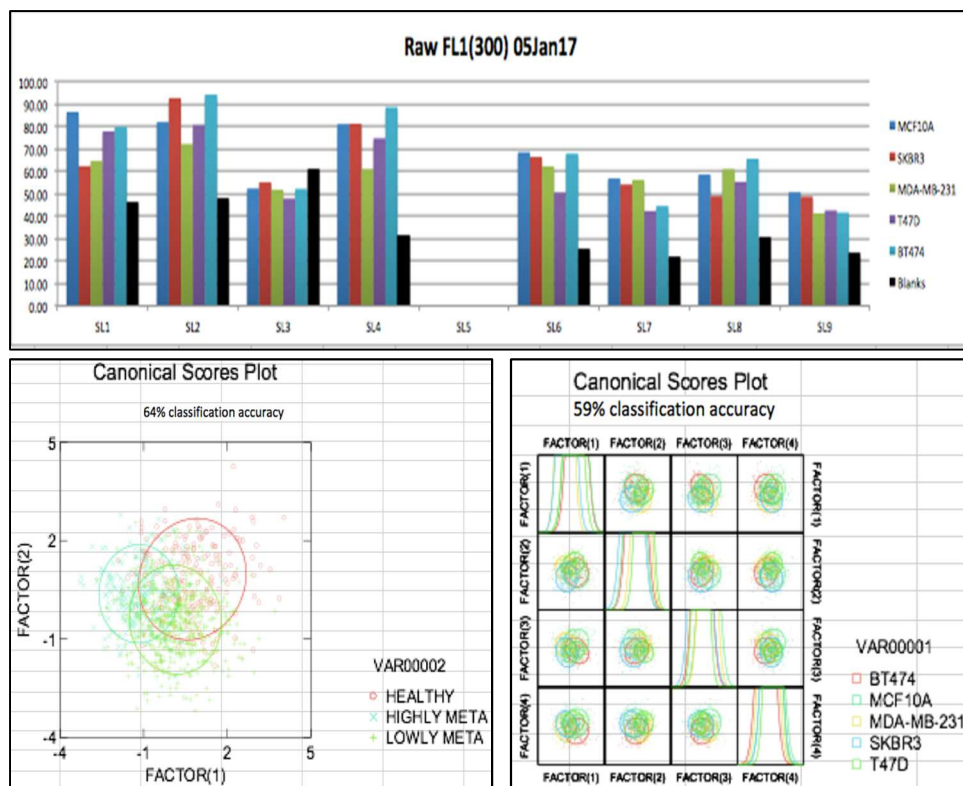


Figure 2.12 SL1-4, 5-9 by 5 breast cell lines labeled at 10:1, FITC:lysine ratio, incubated with 10 μ m SLs on PS beads. SL5 omitted because of lack of resources. Raw data (top), 3-class LDA (left), 5-class LDA (right).

To see if this peculiarity of PS beads was the same with the TentaGel beads, the same experiment was run with SL1-9 on 10 μ m TentaGel beads (100,000beads/sample) with the same glycoprotein extracts labeled in the 10:1, FITC:lysine ratio. The results, as shown below in Figure 2.13, were very similar to what was observed with the PS beads as well. The overall fluorescent intensity of the beads, which averaged around 150-200RFU for most SLs, did not appear to be impacted by the 10x increase in the amount of FITC used during the glycoprotein-labeling step. This data set collected on TentaGel beads was compared to an earlier data set that was performed using the ~130,000beads/sample of the 10 μ m TentaGel beads with 4 breast cell line extracts labeled in a 1:1, FITC:lysine ratio. These results depicted in Figure 2.14, showed that the

overall average fluorescent intensity of the beads were also in the 150-200RFU.

Interestingly though, the 10x increase in FITC used also increased the classification accuracy of both the 3-class and 4-class discrimination accuracies by 20%.

One possible explanation for this increased classification accuracy without seeing an increase of the overall fluorescence intensity of the PS and TentaGel beads, is that the excess FITC is more completely labeling the primary amines available on the glycoproteins in the extracts. This may allude to the fact that the 1:1, FITC:lysine ratio that was established early during the flow cytometer protocol development process, may already be in slight excess in relation to the amount of primary amines exposed in the extracts. This could be because glycosylation helps proteins to maintain a folded structure, shielding some of the theoretically calculated lysines on the proteins, which could possibly occur even after the proteins were extracted from solution by the acetone precipitation.

This claim could be further supported by looking at the overall standard deviations of the beads in the data sets, for which the PS and the TentaGel bead fluorescence, both decrease ~10% standard deviation within the cell lines. The TentaGel beads %SD decreased from 46.25% to 41.5% when increasing the FITC 10x, and the PS beads decreased from 41.25% to 25% when increasing the FITC 10x during labeling. The results from these labeling assays point toward the fact that the 1:1 FITC:lysine ratio may have been too low of a concentration, and caused more deviation in the signal by incomplete labeling of the cell line extracts. These results also revealed some interesting differences between the SLs on PS beads and the SLs on TentaGel beads, which will be the focus of the next section.

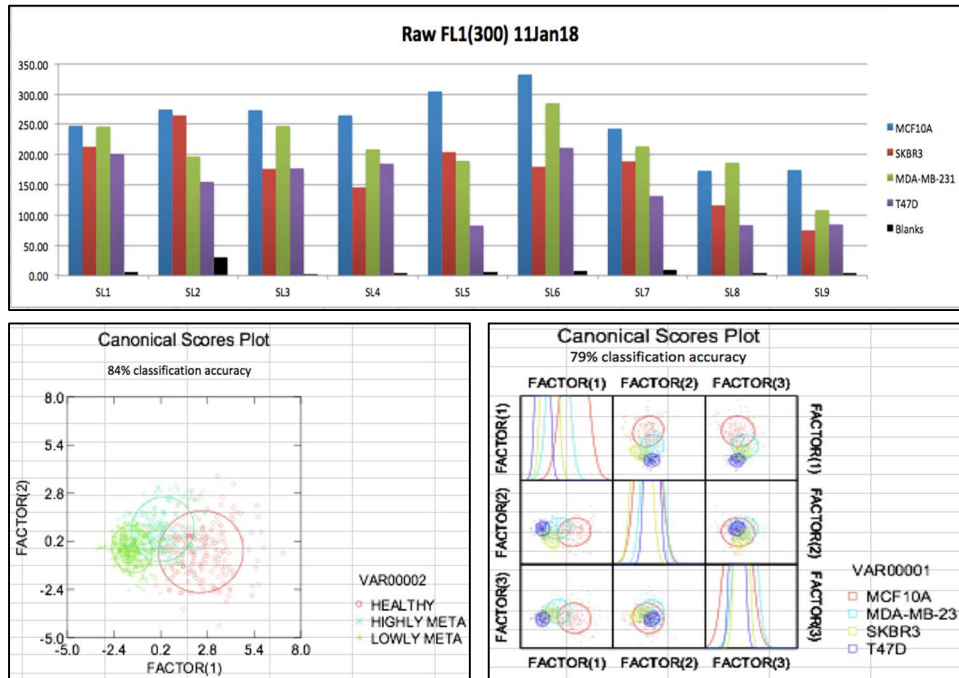


Figure 2.13 SL1-9 by 4 breast cell lines, labeled 10:1, FITC:lysine ratio, incubated with 10 μ m SLs on TentaGel beads. Raw data (top), 3-class LDA (left), 4-class LDA (right)

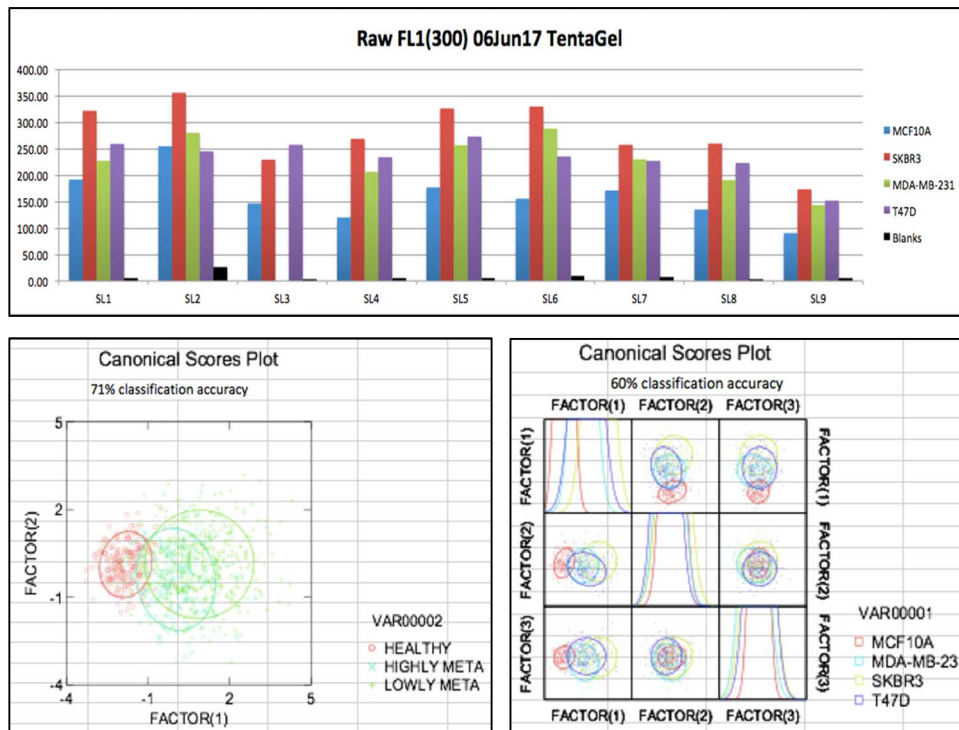


Figure 2.14 SL1-9 by 4 breast cell lines, labeled in 1:1, FITC:lysine ratio, incubated with 10 μ m SLs on TentaGel beads. Raw data (top), 3-class LDA (left), 4-class LDA (right).

EXAMINING THE DIFFERENCES BETWEEN PS AND TENTAGEL BEADS

One interesting pattern that continued to emerge while examining the different properties between the SL array on PS beads and the SL array on TentaGel beads was a repeating fluorescent intensity patterns between the cell lines, when using the TentaGel beads. This pattern is evident in the two data sets described above using the SL array on TentaGel beads. In Figure 2.13, MCF10A was the brightest cell line in 8 out of the 9 SLs used, while MDA-MB-231 was the second brightest cell line in 6 out of 9 SLs, and T47D was the least brightest cell line in 6 out of 9 of the SLs. In Figure 2.14, SKBR3 was the brightest cell line in 8 out of 9 SLs, while T47D is the second brightest cell line in 5 out of 9 SLs, MDA-MB-231 is the third brightest cell line in 5 out of 9 SLs, and MCF10A is the least brightest cell line in 8 out of the 9 SLs.

These repeating patterns between the cell lines within the data set, but not consistent between data sets, seems to suggest that the extent of fluorescent-labeling between the cell lines plays a major role in the way the SL array on TentaGel beads discriminates between the cell lines glycoprotein extracts. The biggest argument for this is that the pattern between the cell lines was completely different between the data sets. The two TentaGel data sets above were incubated with two differently labeled glycoprotein extracts, one in a 1:1 FITC:lysine ratio, and one in a 10:1 ratio. In Figure 2.13, the patterns depicts that SL1-9 are all responding to the four different amounts of fluorocein present in the extracts, from MCF10A, to MDA-MB-231, to SKBR3, to T47D in that order. This pattern is not reproduced in Figure 2.14, where the patterns depicts that

SL1-9 all respond to the cell lines SKBR3, T47D, MDA-MB-231, and MCF10A, in that order from brightest to dimmest.

This inexplicable patterning was not observed when using the SL array on PS beads, where the fluorescent intensities between the cell lines were typically much more similar within each SL. This would stand to suggest that the variability of the SLs in detecting different glycoproteins in a sample, are more unique in PS beads than in the TentaGel beads, which bind more cross-reactively and rely more heavily on the amount of fluorescein present in the cell line extracts. However, the reverse may also be true that the SL array on the PS beads, were binding more cross-reactively to the different glycoprotein extracts, and therefore never make a pattern, which is a bad quality.

One last speculation as to why the SL arrays behave differently is because of the different cross-linking agents used in their resins. TentaGel beads are polyethyleneglycol (PEG) cross-linked, which includes an ether in the repeat unit shown as the structure to the left in Figure 2.15. PEG causes the TentaGel resin to be much more polar and allows the bead to swell to 3.5x in aqueous solution. In contrast, the PS beads are hydrophobic and do not swell in aqueous solution.

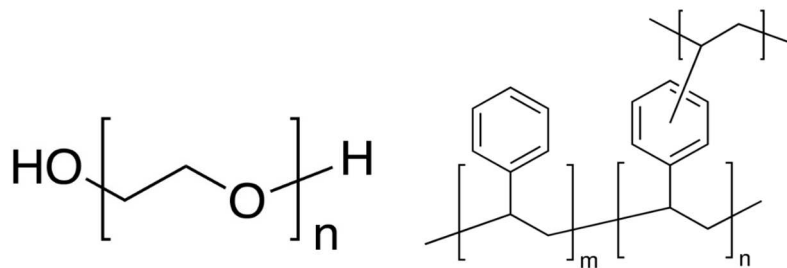


Figure 2.15 Polyethyleneglycol repeat unit used in TentaGel beads (left), added to PS beads, which are comprised of the divinylbenzene repeat unit (right).

When examining these resins with respect to the hydrophobicity of the SLs in the array with regards to each other, a similar pattern emerges with both SL arrays. This idea will be explored again in greater detail, but briefly, each SL in the array is at least 57% identical because seven of the twelve amino acids in the sequences are conserved (MRBB, Dab, Dab, R). Therefore, by comparing the 5 remaining variable amino acid positions, the SLs can be ranked in order of polarity. SL1, 2, 4, and 5 were ranked in this order as shown in Table 2.3 below, and their relative polarity in order was SL4 > SL5 > SL2 > SL1.

Table 2.3 Comparing the amino acid (AA) sequences of SL1, 2, 4, and 5

SL	SEQUENCE	Charge	Polar AAs	Phenyl Rings AAs	Hydrophobic AAs
SL1	Ac-RGD*VTFD*RBBRM	+3	1	1	2
SL2	Ac-RTD*RFLD*VBBRM	+3	1	1	2
SL4	Ac-RRD*TQTD*QBBRM	+3	4	0	0
SL5	Ac-RAD*TRVD*VBBRM	+3	1	0	3

Interestingly, the consistently brightest SL for the TentaGel array is SL1, which is the most hydrophobic SL on a hydrophilic resin. For the PS array, SL4 is consistently the brightest SL, which is the most hydrophilic SL on a hydrophobic resin. The polarity ranking described above between these four SLs continues along that trend for PS and TentaGel beads, but in opposite directions. For the TentaGel array, the dimmest SL is SL4, which is the most polar SL on a hydrophilic resin. Similarly, the dimmest SL on the PS array is SL1, which is the most hydrophobic SL on a hydrophobic resin. Essentially, the brightest SLs are the amino acid sequences that are most unlike the resins on which they reside, and the dimmest SLs are the amino acid sequences that are most like, in terms of hydrophobicity, to the resin they are conjugated to. This hydrophobicity

relationship is shown below in Figure 2.16, which ranks the SLs in brightness with regards to the two different resins. This concept will be referred to several times throughout this project, especially in Chapter 3 and 4, when using these SLs for solid phase extraction and eluting glycoproteins from them.

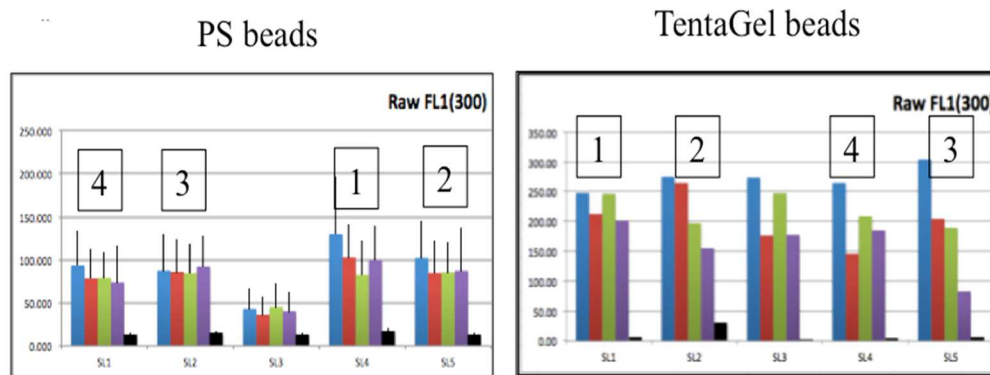


Figure 2.16 Comparing SL fluorescent intensities between PS beads and TentaGel beads. Rank: 1 = brightest, 4 = dimmest fluorescent signal.

Having scrutinized many of variables in this project, a concern that continual arose was whether or not the fluorescent-labeling step was having an impact on the data set and LDA discrimination accuracy. In addition to observing suspicious repeat patterns in the TentaGel array, there were reservations about the whole process from a theoretical perspective. Namely, the cell lines should not be individually labeled if the goal is to discriminate those cell lines from each other. For this reason, an external source of fluorescence was chosen to elucidate the underlying binding interactions between the SLs and glycoproteins, without having to consider the potential of biasing data.

USING NATURAL LECTINS TO DETERMINE THE UTILITY OF THE SL ARRAY

In considering the dilemma regarding the labeling of the glycoprotein extracts, another approach was taken using fluorescently labeled natural lectins (NLs) to bind to

the glycoproteins that are bound to the SLs. These fluorescein-conjugated NLs, provided by Vector Laboratories, are purified proteins of known size and FITC ratios, and were therefore thought to be a better method for quantifying of fluorescent signals between the cell lines. Most importantly, using these NLs in the assay meant that the glycoprotein extracts did not have to be fluorescently labeled, alleviating any potentially biased data set. In addition, these quantifiable signals could allowed for more specific comparisons between the SLs because each of the three NLs have their own binding preferences for different known carbohydrate moieties²⁸, which could provide insight into the biological content bound to each SL.

The three fluorescently-labeled NLs that were used were Griffonia Simplicifolia Lectin II (GSLII), Jacalin, and Sambucus Nigra Bark Lectin (SNA), which are all derived from plants. For these experiments, the four breast cell line extracts from MCF10A, SKBR3, MDA-MB-231, and T47D, were incubated with the four most reproducible SLs 1,2,4, and 5, which were selected based on their reproducibility across platforms. However, instead of analyzing the samples the next day, they were each incubated again overnight with each of the three NLs. These samples were then analyzed on the FC500 the following day.

This binding assay using the Synthetic Lectin-glycoprotein-Natural Lectin, in that order was referred to as a sandwich assay, and will be discussed in few sections. But first, in order to properly introduce these NLs into the sensor array, the relationships between the each of the NLs with each of the SLs had to be established. These were important control experiments because all 3 of the NLs used are each glycoproteins themselves, each with several known N-linked glycans. The NLs extent of glycosylation will be

discussed in greater detail in Chapter 4, when the N-linked glycans were removed from the NLs and analyzed by LC-ESI-MS.

Briefly though, GSLII is a homotetrameric protein of 113kDa with two N-linked glycans at amino acid positions 5 and 18. Jacalin is a 65kDa tetrameric protein made-up of two α -chains and two β -chains, and also has two N-linked glycans on the α -chain at amino acid position 43 and 74. Additionally, Jacalin is also reported to have a biantennary oligomannoside of 1272Da somewhere. SNA is a 140kDa protein and also a tetramer made-up of two α -chains and two β -chains. The α -chain has N-linked glycans at positions 40, 62, 144, and 260, and the β -chains have N-linked glycans at positions 492, and 526. The glycans mentioned above were taken from published literature, however, the N-linked glycans of these three NLs were analyzed in Chapter 4 by LC-ESI-MS, and will be discussed there again.

THE SYNTHETIC LECTIN AND NATURAL LECTIN INTERACTION ASSAY

In using these NLs as a fluorescent detection method, the recommended concentration range provided by the manufacturer is 5-20 μ g/ml. Since each of these NLs were going to be used in a number of assays designed as three-dimensional arrays (4 cell lines x four SLs x 3 NLs), the smallest concentration of 5 μ g/mL was chosen to conserve materials. For these first SL-NL characterization assays, 200,000 beads of each SL on 10 μ m PS beads were incubated directly with 5 μ g of each of the 3 NLs. Three technical replicates of this experiment were performed, all of which were highly reproducible. The LDA graphs shown below in Figure 2.17 are from one set of data, where the four SLs were used to discriminate the 3 NLs on the left, and the 3 NLs were used to discriminate

the four SLs on the right. When using the SLs to discriminate the 3 NLs from each other, the classification accuracies of the three replicates were 93%, 92% and 91%. The only misclassifications of NLs occurred between GSLII and SNA. Jacalin was distinguished from both of these.

The success of these four SLs in their ability to discriminate between different purified glycoproteins was anticipated and is consistent with the proof-of-concept experiments that began this project, and how the 12-mer sequences of the SLs1-9 were selected. In quickly speculating about the classification of these 3 NLs in relation to each

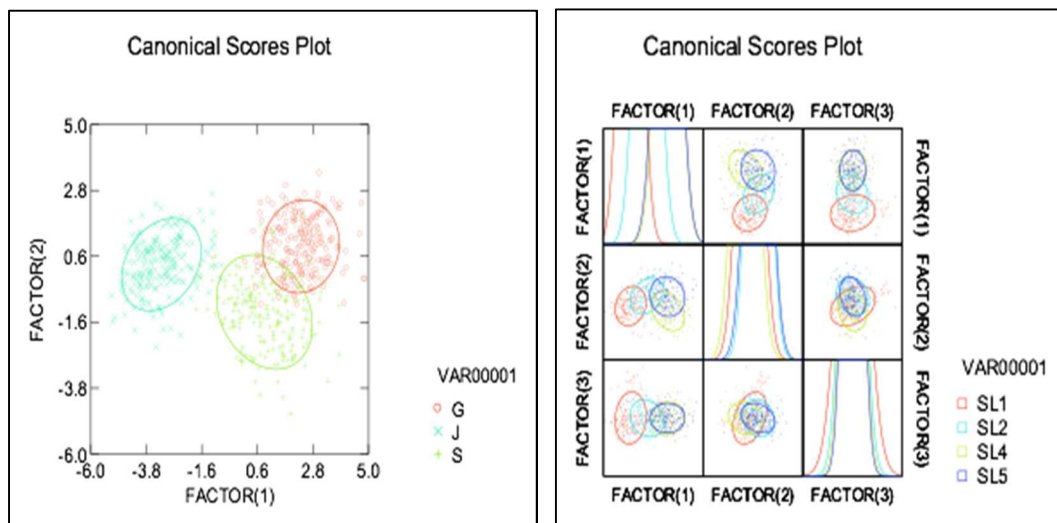


Figure 2.17 Discriminating the three NLs by the four SLs (left), discriminating the four SLs by the three NLs (right)

other, the most common factor between GSLII and SNA is their size, which is about twice as large as Jacalin. Conversely, the 3 NLs had more difficulty in discriminating the four SLs from each other, with classification accuracies of 69%, 57%, and 68%. In all three runs, GSLII had a lower F-to-remove factor by 3-10fold than Jacalin and SNA, meaning that it did not contribute much in discriminating between the SLs.

The lower classification accuracies in distinguishing the SLs from each other was not surprising considering the similarities in the sequences of the 12-mers conjugated on the beads. Each SL is at least 57% identical because seven of the twelve amino acids in the sequences are conserved (MRBB, Dab, Dab, R) among all of them. However, despite the lower classification accuracy, the 4-class LDA results showed very similar patterns for all the technical replicates. SL1 was the most distinguishable SL in each run, but overlapped with SL2, which more significantly overlapped with SL4 and SL5. SL4 and SL5 consistently had the most misclassifications between them.

When comparing this spatial relation of SLs with the similarities they have between their 12-mer amino acid sequences (Shown in Table 2.3), it stands to reason that SL1 and SL2 would resemble each other because of the five variable amino acid positions, both contain an arginine, 1 polar amino acid, one amino acid with a phenyl ring and 2 hydrophobic amino acids. SL2, SL4, and SL5 were all significantly misclassified as each other, and could possibly be explained by the common threonine they share in the last or second to last variable amino acid position at the N-terminus, which is the most exposed region of the 12-mer. SL4 is the most polar of these four SLs, particularly at the N-terminus, and did not overlap with SL1, which arguably has the most hydrophobic sequence at the N-terminus.

Table 2.3 Comparing the amino acid (AA) sequences of SL1,2,4, and 5

SL	SEQUENCE	Charge	Polar AAs	Phenyl Rings AAs	Hydrophobic AAs
SL1	Ac-RGD*VTFD*RBBRM	+3	1	1	2
SL2	Ac-RTD*RFLD*VBBRM	+3	1	1	2
SL4	Ac-RRD*TQTD*QBBRM	+3	4	0	0
SL5	Ac-RAD*TRVD*VBBRM	+3	1	0	3

In fact, when analyzing past and present data sets using the SLs on PS beads, the overall fluorescent intensities between the SLs seems to correlate to the hydrophobicity trend between these four SLs. SL4, the most polar SL, is consistently the brightest SL for all of the cell lines in every data set, followed by SL2 and SL5 with similar intensities, and then SL1, which has the most hydrophobic sequence at the N-terminus and is usually the dimmest between these four SLs. This correlation between bead hydrophilicity and fluorescent intensity was expected because the biological material used in the incubations is proteins precipitated out from material secreted from the cells, which have been sighted throughout the literature to be ~70% glycosylated, making them more polar. Interestingly, this correlation was observed to be the opposite for the SLs on TentaGel beads, where SL1 was usually the brightest, SL2 and SL5 remained comparable in the middle, and SL4 was consistently the dimmest SL. This could again be explained by the increased polarity in the TentaGel resin as a result of the polyethyleneglycol cross-linking agent. This topic will be referred to again in Chapter 4 when these four SLs on PS beads were used as solid phase extraction materials for these samples, and the N-linked glycans were isolated, eluted, and analyzed by LC-ESI-MS.

THE SYNTHETIC LECTIN AND NATURAL LECTIN INTERACTION ASSAY
AFTER DE-N-LINKED GLYCOSYLATING THE NATURAL LECTINS WITH
PNGASE F

To further understanding the interactions between these four SLs and these 3 NLs used, the enzyme PNGase F (Ned England Biolabs) previous described earlier in this

chapter, was used to cleave off all of the N-linked glycans on the NLs. The 3 NLs were glycosylated with PNGase F at 37°C for 4hrs, and then filtered through a 10kDa centrifuge tube (Amicon) and buffer transferred into PBS_L before incubating them with the four SLs, as described above. The mass range of human N-linked glycans is 2-5kDa, therefore filtering the samples through 10kDa centrifuge tubes was sufficient in removing all the freely liberated N-linked glycans from any larger remaining proteins and O-linked glycoproteins in the samples.

When the de-N-linked-glycosylated NLs were incubated with the four SLs, the overall discrimination accuracies did not change drastically, as Shown below in Figure 2.18. Two technical replicates were performed and both 4-class LDA classification accuracies were 63%, while the 3-class LDAs were 87% and 83%, which dropped on average 5% from the previous study using the fully intact NLs. When the four SLs were used to discriminate the 3 NLs, Jacalin remained distinguished as before, but GSLII and SNA began to overlap more significantly than they did when they had their N-linked glycans. Using the same logic as before, this result may be a factor of the size of the NLs used, which became more influential once the NLs were cleaved of their N-linked glycans. The tighter clustering of GSLII and Jacalin may also suggest that those two NLs did not contain any other glycans, such as O-linked glycans, and/or was the most efficiently de-N-linked glycosylated NL.

When using the 3 NLs to discriminate between the four SLs, only SL2 appeared to be significantly impacted by the de-N-linked-glycosylation of the NLs. SL4 and SL5 still remained overlapped as before, as did SL1, which only slightly overlapped with SL5 and not at all with SL4. However SL2 drifted away from SL5 towards SL1, lowering the

classification accuracy of SL1 and SL2. This significant misclassification between SL1 and SL2 may be a result of the similarity of their overall amino acid content in the 12-mers, which only differed in their location of hydrophobicity. SL1 and SL2 have extremely similar 12-mers, however, SL1 is more hydrophobic than SL2 at the N-linked terminus, which appeared to be less of a factor once the NLs were de-N-linked glycosylated, and the NLs became less polar.

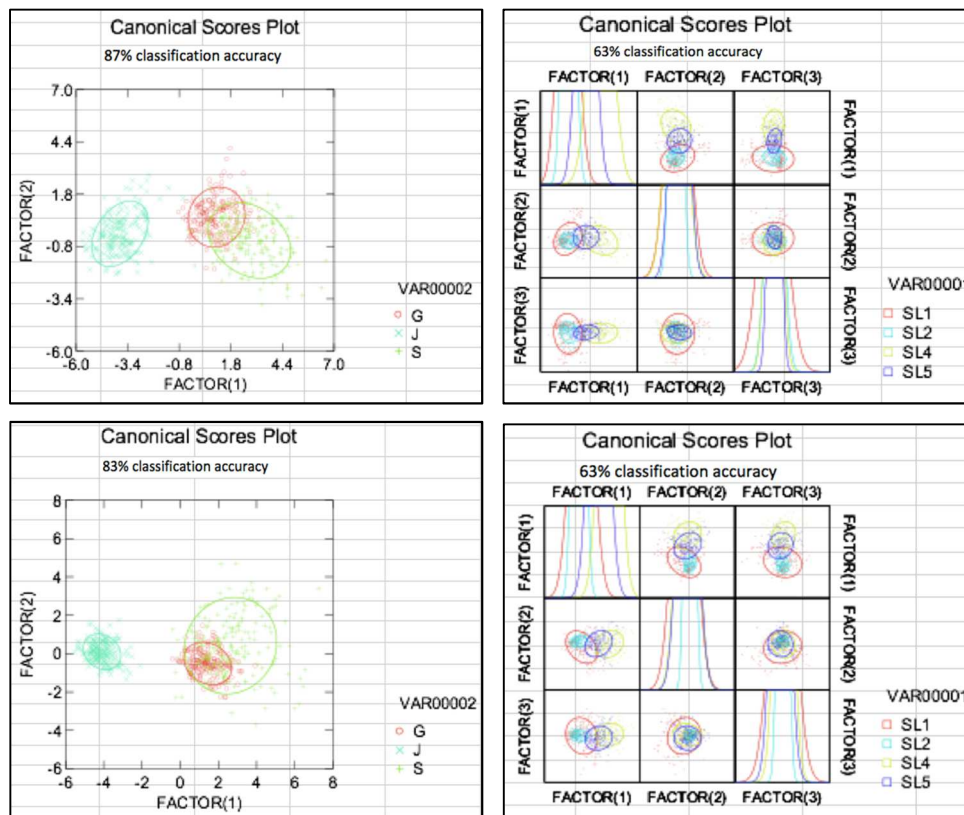


Figure 2.18 Discriminating 3 NLs by four SLs (left), discriminating four SLs by 3 NLs (right), after de-N-linked glycosylating the NLs.

The similarity in the two experiments described above, between the SL-NL interaction assays and the SL-de-N-linked-glycosylated NL interaction assays, demonstrated that the N-linked glycans on the NLs only played a minor role in the ability of the SLs to discriminate between glycoproteins. The affinity of SLs to nonglycosylated

proteins, such as BSA, has been known since the beginning of the project, as BSA was used as one of the purified glycoproteins in the selecting SLs1-9. Therefore, this information of SL affinity to glycosylated proteins over the same nonglycosylated proteins will be a valuable tool in the discerning of future SLs for the sensor array. The observed differences could also be a result of the O-linked glycosylation on the NLs, which was unaddressed in this comparison assay.

Whatever the contributing factors may be, the challenge still remains in the complexity of the biological samples over the purified glycoprotein samples. The tight reproducibility of the purified glycoproteins and the inconsistencies seen in the biological glycoprotein extracts, alludes to the fact that the biological samples have much more complex interactions with the SL array, and therefore have an inherent lack of reproducibility because of the diversity of the biological material. This in turn, reflects on the diversity of the sensor array being used, and suggests that a more diverse group of SLs may be necessary in order to highlight the minute differences seen between healthy complex biological samples and cancerous complex biological samples, each containing over a hundred different types of glycoproteins per sample.

THE SL-GLYCOPROTEIN-NL SANDWICH ASSAY WITH THE BREAST CELL LINE EXTRACTS

When using the NLs to detect the SL-glycoprotein extract interactions in a sandwich assay, the results were even more inconsistent than when labeling the glycoprotein extracts directly. This was likely due to the added complexity when introducing a third component into the array. The NLs had already demonstrated a strong

affinity for the SLs, which would make them competitive binders with the cell line glycoprotein extracts. Three technical replicates were performed incubating SL1, 2, 4, and 5 overnight with the four breast cell line secreted extracts. The samples were washed 3x with PBS_L, 2%BSA, 2%etOH, and reincubated overnight with 5µg of the three NLs. The LDA results for the 4-class discrimination were 91%, 60%, and 74%. However, these experiments proved to be useful information when analyzing the relationship between the cell line extracts. As shown below in Figure 2.19, the first data set had 91% classification accuracy with little overlap between the cell lines. The largest misclassifications happened between SKBR3 and MDA-MB-231, MCF10A and SKBR3, and MCF10A and T47D. The next technical replicate showed 30% lower classification accuracy, most of which was attributed to the more significant overlap between SKBR3 and MDA-MB-231, and MCF10A and T47D to a lesser extent. This consistently observed relationship between the cell line extracts was examined in greater detail in Chapter 4, when analyzing the N-linked glycans in each of the cell line extracts.

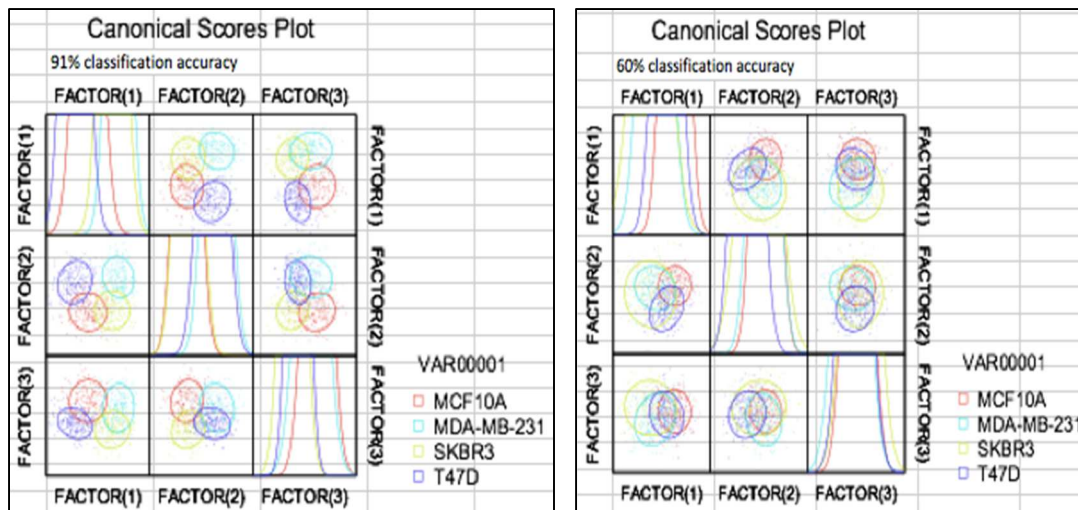


Figure 2.19 Two technical replicates of four breast cell lines incubated with SL1, 2, 4, and 5 overnight. Then washed and incubated with 5µg of three NLs, GSLII, Jacalin, and SNA the following night.

THE SL-GLYCOPROTEIN-NL SANDWICH ASSAY AFTER DE-N-LINKED GLYCOSYLATING THE BREAST CELL LINE EXTRACTS WITH PNGASE F

The SL-glycoprotein-NL sandwich assay also provided additional useful information when comparing this previously discussed data set, with another triplicate data set that used PNGase F (New England Biolabs) to de-N-linked glycosylate the cell line extracts before incubating them with the SLs and the NLs. Shown at the top left in Figure 2.20 below, five breast cell lines were incubated with 4 μ g of PNGase F for 4hr @ 37°C, then filtered through 10kDa MWCO Amicon centrifuge tubes to remove the freely liberated N-linked glycans from the remaining proteins and O-linked proteins. These samples were then run on an SDS-PAGE alongside the parent cell line extracts. The SDS-PAGE shows a mobility shift in the de-N-linked extract samples on the right by the darker bands lower on the gel, which correlates to smaller molecule size than the parent extracts on the left. This demonstrated that the extracts had a smaller average size after using the PNGase F, which implies that N-linked glycans were successfully removed from the glycoproteins in the extracts.

When this was done the 4-class discrimination accuracies of the de-N-linked extracts dropped to 76%, 53%, and 44%. In all three of the LDA plots, MCF10A remained the most distinguished cell line, while the three cancer cell line extracts began to overlap more significantly. This implies that there are other factors beside N-linked glycosylation that allow the SL array to discriminate between healthy and cancer extracts, such as a different protein profile or O-linked glycosylation, but the presence of the N-linked glycans was significant in discriminating between cancerous cell line extracts, and is clearly much more important in discriminating between complex biological samples.

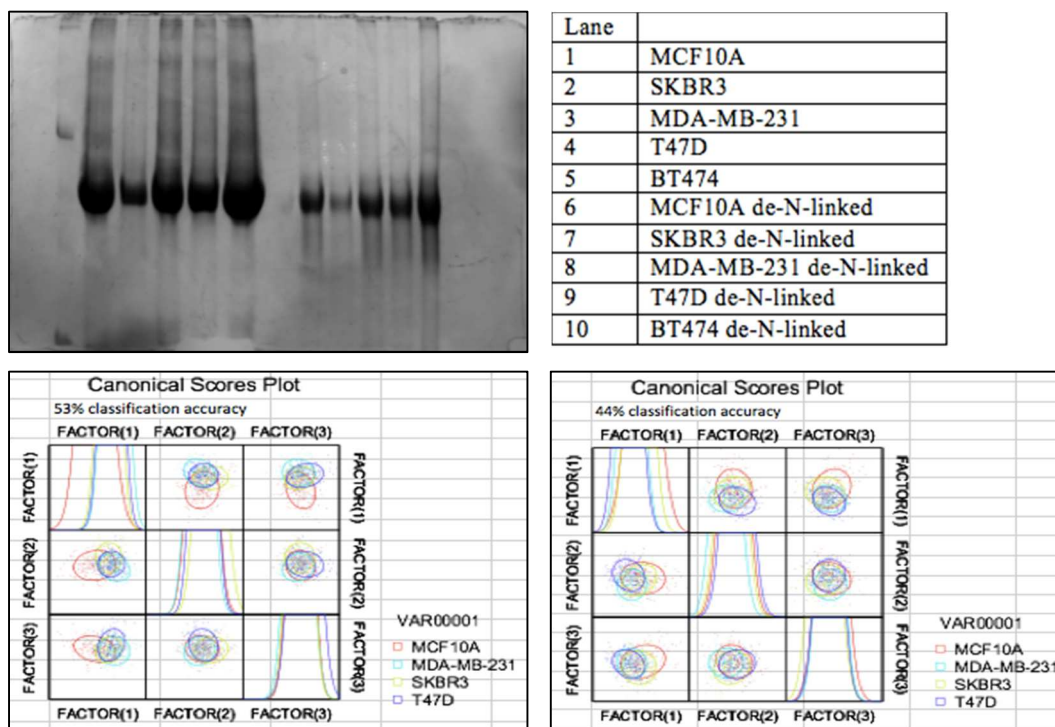


Figure 2.20 Five breast cell line extracts de-N-linked glycosylated with PNGase F (top left), the corresponding samples and lanes (top right), two technical replicates of SL1, 2, 4, and 5 incubated with four of the de-N-linked breast cell line extracts overnight, washed and re-incubated with three NLs the following night (bottom).

OUTRO

In order to better understand the SL interactions with these complex human glycoproteins, it seemed important to have some idea of what these glycoprotein targets are. Therefore an investigation into the N-linked glycans bound to each SL was conducted. This was done by first using the SLs as solid phase extraction (SPE) agents to bind and release their preferred glycoproteins from the samples, which were then further analyzed by LC-ESI-MS. This project also helped to explore the idea of using the SL array as SPE agents for the purpose of being a research tool or a prognostic tool. These topics are the main focus of Chapter 3.

CHAPTER 3
USING SYNTHETIC LECTINS FOR SOLID PHASE EXTRACTION
BACKGROUND

The cross-reactivity of the Synthetic Lectin array provides an added benefit in that it can potentially bind any glycoprotein it comes in contact with, including glycoproteins that have yet to be identified. This wide-scope utility of the SL array coalesces nicely with the ever-growing complexity of cancer diagnostics and personalized medicine. Cancer is a disease unique to the individual, and therefore follows reason to believe each cancer-associated glycan profile is unique to the individual as well, since it is these glycosylation pattern changes that largely drive the tumor behavior in the first place. In this regard, it would be a powerful learning tool to be able to non-selectively bind to any glycoproteins associated with tumors in individuals across populations, to gain insight into the relationship between aberrant glycosylation and cancer progression, and to more deeply understand an individual's disease state and a more accurate prognosis.

In order to have a successful solid phase extraction material for glycoproteins, a few other criteria must be met. First, the material must be non-toxic and non-destructive of the biological molecules it is binding. Secondly, the material must be able to reversibly release the bound biological molecules under conditions that will not be destructive to the biological molecules. If either one of these criteria are not met, then the solid phase extraction material would be of little use because the biological molecules could not be subsequently examined and identified. The above criteria are more easily met when

dealing with peptides or proteins because of their stability, however the challenges become much greater when dealing with complex carbohydrates, which are notoriously unstable, as they serve more transient purposes in the body, such as cell-to-cell communication and immune responses.

Peptides are very stable structures, and require a low pH of 1 along with high heat (100 degrees Celsius) for hours, such as in the Edman's Degredation process, just to break the peptide bond between amino acids. This is not the case with carbohydrates, which are liable to hydrolyze at temperature in excess of 37°C or pH conditions outside of the 4-9 range. This is particularly true for the hydrolysis of N-acetylneuraminic acids (Sialic acid), which when present are always terminal non-reducing saccharides on glycan molecules and the first to break off. Sialic acid is of great interest in the fields of cancer biology and immunology because it has been observed to play roles in the disguising of cancer cells from immune cells, as well as aid in the detachment of cells from the primary tumor and metastasize to distal locations.²⁷ For this reason, extra care had to be taken into account in using the SL array as solid phase extraction material because any conditions too harsh for the preservation of the intact carbohydrate molecules would be detrimental to the utility of the SL array as a learning tool.

The SL array met the first condition of a good solid phase extraction material because the beads are essentially plastic cells that are coated with 12-mer peptides, with nothing unnatural save for the phenylboronic acids. The two phenylboronic acids on each 12-mer were covalently bound with diaminobutanoic acids through a stable imine formation, and has never been observed to break-off and contaminate any of the bound biological material as an adduct. The reversibility of the bond between 1,2 and 1,3 cis-

diols with boronic acids is known to occur, and has been previously viewed as a way to ensure the strongest binding affinity between the SLs and different, and very numerous amount of glycoproteins. However, the unbinding of the SLs with its target glycoprotein material has never been examined in detail, and warrants thorough consideration because of the aforementioned instability of complex carbohydrates. For samples simpler than biological samples, the reversible binding of glycans with boronic acids has been known for a long time and has been thoroughly elucidated, in particularly by E.V. Anslyn, the former boss of Dr. Lavigne. Experiments have shown that the bond is most stable in neutral and basic conditions, but reversible under acid conditions of $\text{pH} < 5$. In a monosaccharide and disaccharide competition assay, it was also demonstrated that d-fructose has the highest binding affinity for boronic acids, but at a concentration of 1M. Therefore, in choosing an elution buffer for reversing the SL-glycan bonds, 1M d-fructose in 10% acetic acid in dH_2O , pH adjusted to 4.25, seemed to be a good candidate. Initially however, the pH was left at $\text{pH}=2$ to give some better context into the stability and reversibility of these bonds.

CHOOSING THE BEAD-TYPE FOR SPE

The SL array has already demonstrated its ability to cross-reactively bind to a number of different targets, including non-glycosylated proteins such as bovine serum albumin, but the total binding capacity, or total amount of protein able to bind to 1mg of each SL, had not yet been quantified. As previously mentioned, the functionality of the polystyrene beads is 4x that of the TentaGel beads, which allows for more 12-mer peptides to be assembled from the primary amines bound to the matrix of the beads. For

this reason, the polystyrene beads were assumed to have a higher binding capacity for glycoproteins over the TentaGel beads, but this theory had to be confirmed because the TentaGel beads swell significantly more in aqueous solution, so the opposite could be true as well. Or, in contrast, the higher functionality of the polystyrene beads could lead to steric hindrance of the glycoproteins, and therefore effectively lowering the overall binding capacity for glycoproteins. The goal of binding the most amount of glycoprotein was to ensure that there would be enough biological material so that there would not be any sensitivity issues in subsequent analysis, particularly with mass spectrometry. For this reason, the first question asked was whether to use polystyrene or TentaGel beads as the SPE material.

The first solid phase extraction assay was to determine which of the larger 300 μ m beads, TentaGel or polystyrene, could be used to bind and elute the most amount of glycoprotein. The previous work and publication regarding this SL Array was done using the 300 μ m TentaGel beads. The 300 μ m beads are large enough to be individually visible to the eye, and therefore initially thought of to be easier to work with, in terms of distributing them across the array and separating them from liquid suspensions and washes. Briefly, the method of incubation included an overnight incubation of each SL with a sample of glycoprotein extract, in PBS_L at room temperature on a rocker. The following day, the samples were washed 3x with PBS_L to remove any excess and weakly bound material, and then incubated 3x for an hour each in 1mL aliquots of the elution buffer described above. Those three aliquots of elution buffer were collected, and concentrated and buffer exchanged with deionized water in Amicon 3kDa MWCO centrifuge tubes. The samples were then concentrated by vacuum centrifugation, and the

protein concentrations of those samples were determined using the Bradford Assay. The BCA assay could not be used for these experiments because the presence of the d-fructose and the acetic acid of the elution buffer both interfere with the BCA protein concentration assay, even in dilute amounts. This was readily apparent when comparing the BSA standard curves in PBS_L and in the elution buffer. Therefore, the Bradford Assay had to be used for determining the protein concentrations of all the following solid phase extraction experiments.

For the incubations, shown in Figure 3.1 below, 70-80mg of SL1 and SL2, synthesized on both polystyrene and TentaGel beads, were incubated with 50µg of glycoprotein from MCF10A and MDA-MB-231 cell line extracts. Three technical replicate were performed and consistently showed that the TentaGel samples contained almost all of the total protein in the wash aliquots, which meant that more glycoprotein was binding with weaker affinity to the TentaGel beads. The polystyrene SLs also had roughly 2x the amount of total protein eluted compared to the corresponding TentaGel SLs. SL2 eluted more total protein than SL1 in both polystyrene and TentaGel.

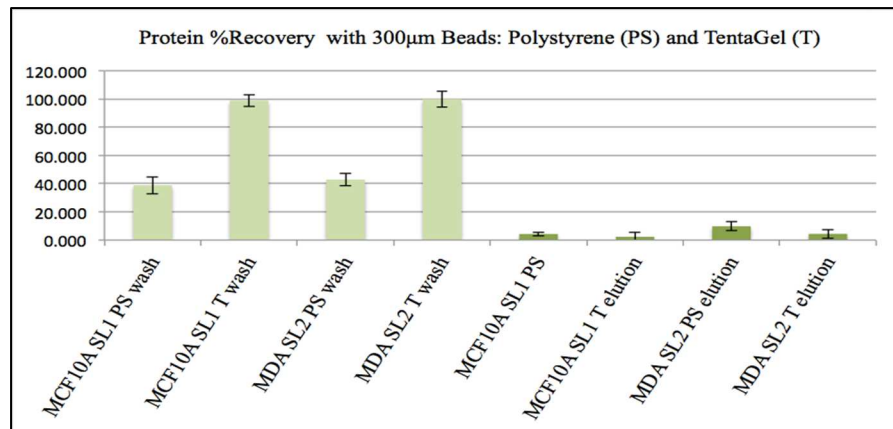


Figure 3.1 Protein % Recovery from using 300µm PS and TentaGel beads as SPE agents with MCF10A and MDA-MB-231 cell line

However, the total amount of protein recovered from every sample was only about 2-10% of the starting glycoprotein used. Additionally, the wash and elution buffer for the TentaGel beads accounted for almost all of the protein, whereas around 50% of the glycoprotein was assumed to be still bound to the polystyrene beads because it was unaccounted for.

Upon examining the beads under a microscope, a large portion of the 300 μ m polystyrene beads were fractured and broken into smaller, non-spherical pieces, whereas the TentaGel beads were still completely intact. SL2 of the polystyrene beads, which eluted the most amount of protein, was much more significantly fractured than the SL1 polystyrene beads, which was evident by the bright halos around the beads in the first and third pictures of Figure 3.2 shown below. This gave the impression that the fracturing of the beads had an impact on the overall total protein binding capacity due to the increase in surface area.

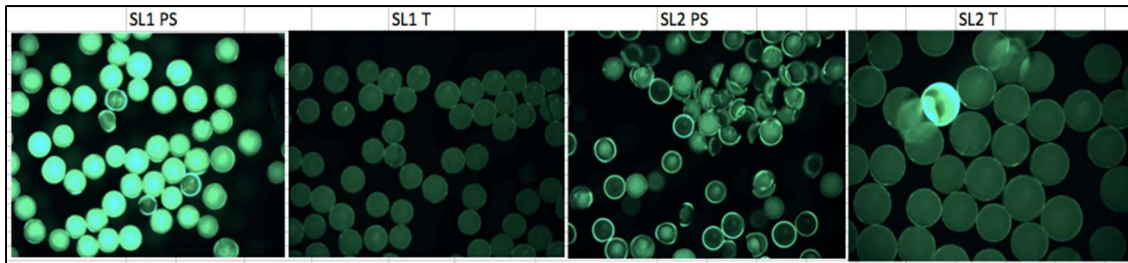


Figure 3.2 Images of 300 μ m PS TentaGel (T) beads taken with fluorescent microscope. Beads with bright halo rings around the circumference are fractured. SL2 PS was the most fractured and had the most eluted glycoprotein material.

As previously mentioned, the surface area to volume ratio is of importance when considering the size of bead resin used because the matrix pore size of the resins are too small for the glycoproteins to be able to penetrate fully into the bead. Therefore, all of the

SL-glycoprotein interactions are assumed to take place only on the surface despite the bead being fully functionalized throughout the bead. Therefore, having smaller beads would increase the surface area-to-volume ratio of the resin, (0.6 for the 10 μ m beads compared to 0.02 of the 300 μ m beads) providing more room for the glycoproteins to interact with the SLs, and less unused dead space within the bead. Having already used the SL array on 10 μ m PS beads for previous experiments with the flow cytometer, it was the best candidate for the next SPE assays.

Once switching to the 10 μ m beads, other benefits were immediately apparent over the larger beads. The first was that the 10 μ m beads are too small fracture during synthesis and use, which was the main problem when using the 300 μ m polystyrene beads. The second benefit of using the smaller beads is the relative ease of distribution between samples, of which great emphasis was on in Chapter 2.

Since the SLs are used in an array format, the relative amounts of beads and glycoprotein between each sample is of great concern when comparing the SLs binding properties to each other, especially when quantifying the amount of eluted protein. The smaller beads could simply be put into a glycerol-PBS_L suspension as described before, vortexed, and easily distributed with a pipet, whereas the 300 μ m beads would have to be scooped out with a spatula, in a slurry due to the aforementioned electrostatic nature of the dry beads, and distributed into each tube based on visual estimation (As performed by previous members using the fluorescent microscope technique). For all of these reasons, the resin selected to explore the SL array's use as a solid phase extraction material was the 10 μ m polystyrene beads.

USING 10 μ m PS BEADS AS SPE WITH PURIFIED GLYCOPROTEIN

Having selected the resin to be used for these experiments, the total protein binding capacity was revisited using purified proteins and glycoproteins. The incubation condition, or ratio between protein and bead number, used in the previous flow cytometry experiments described in Chapter Two, were kept the same, which included 0.0005mg of protein for every 200,000 beads of the 10 μ m polystyrene beads per sample. In order to examine the glycoprotein binding capacity with the SLs, the amount of protein had to be above a measurable concentration threshold, which for the Bradford Assay is 0.03mg/ml of protein. Therefore, it is was necessary to drastically scale-up this incubation condition 500 fold, so that 0.25mg of protein could be incubated with 51.28mg of unfunctionalized beads, or 10⁸ beads per sample. The purified glycoproteins used in these first assays were ovalbumin (OVA) and bovine submaxillary mucin (BSM), and were chosen because they are two proteins that represent largely different molecular weight and extent of glycosylation, and had been used in previous experiments on this project as well. The samples, after eluting, concentrating, and buffer exchanging with the Amicon 3kDa MWCO spin tubes, had to be extra concentrated by vacuum centrifugation to get the protein concentration well within the detection range of the Bradford Assay.

In the assay, shown in Figure 3.3 below, SL1 and SL2 were used again, and the same relationship was seen in that SL2 consistently eluted more glycoprotein than SL1. As seen previously, both SLs had a greater binding affinity to BSM over OVA, which could be explained by either the larger size or extent of glycosylation of BSM. In total, roughly 6% of the OVA was bound and recovered by the SLs, while around 20% of the BSM was bound and recovered by the SLs. This corresponds to roughly 10-20 μ g of OVA

and 40-50µg of BSM. The extent of glycosylation between OVA and BSM, which 5%mass and 50%mass respectively, appeared to be reflected in the wash step, with 20% of the OVA washing off while only 5% of the BSM washing off. In general though, only about 25% of the total protein was accounted for, suggesting that more than half of the glycoprotein remained bound to the beads even after the three 1hr elutions.

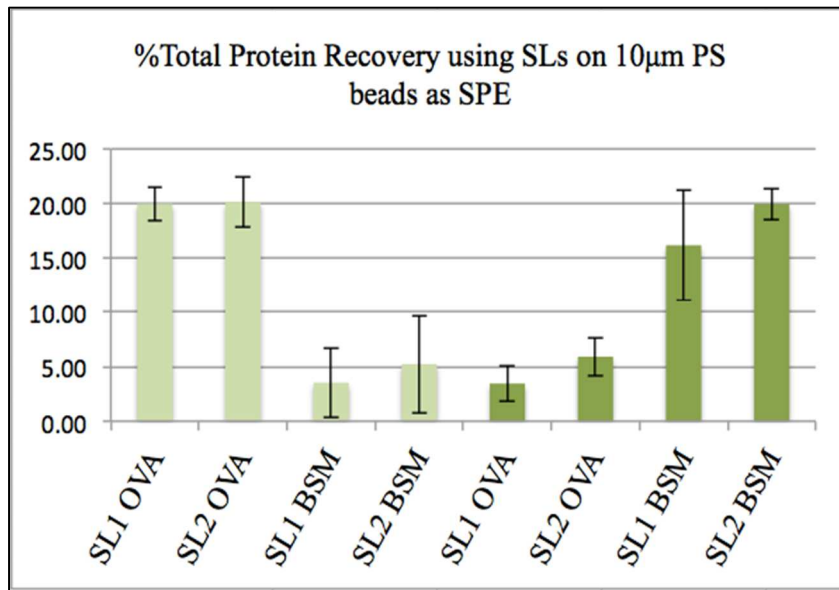


Figure 3.3 Protein % Recovery using SLs on 10µm PS beads as SPE agents with purified glycoproteins.

DETERMINING THE BINDING CAPACITY OF THE SL ARRAY WITH BSA

Before continuing the development of the SLs as a means of SPE, it was important to determine the total amount of BSA able to bind the different SLs, as a baseline for overall binding capacity. This was simply done by incubating 51.28mg of each of the SLs with 500µg of BSA overnight at room temperature on a rocker. The next day, the liquid was removed from the beads and the protein concentration was determined using the BCA assay. The protein concentrations shown below represent amount of

protein detected subtracted from the initial amount of protein of 500 μ g. On average, 180-200 μ g of BSA was able to bind to 51.28mg of the SLs with consistency. Interestingly, SL4 demonstrated a lower binding capacity for the protein than the other SLs. This reaffirms this observation and suggests that the 12mer peptides on SL4 are more polar than the 12mers on the other SLs, and may prefer binding to glycosylated material, rather than with a nonglycosylated protein, such as BSA. Following this logic, it would have been expected that SL1 would have bound the most BSA, however it bound less material than SL2 and SL5, which could mean that SL1 is possibly too hydrophobic for BSA, while the 12mer of SL5 is better suited.

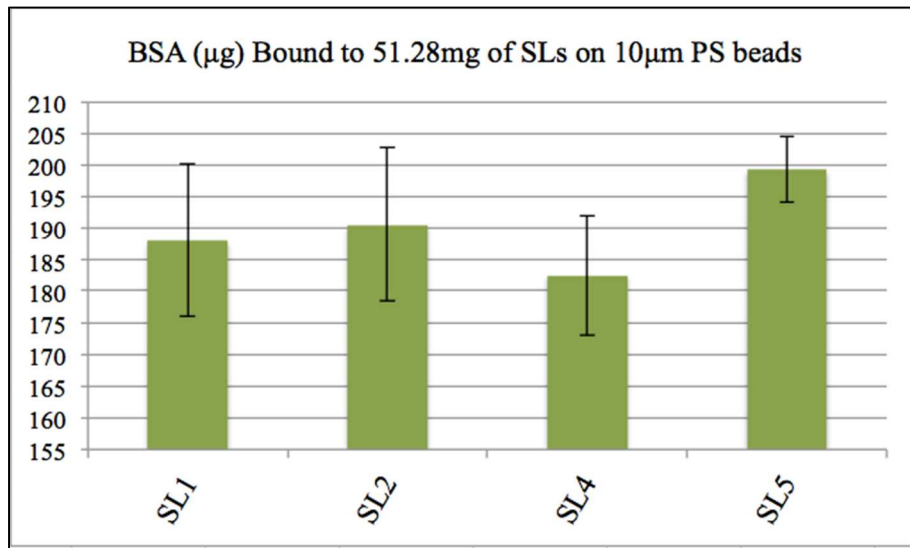


Figure 3.4 Total BSA (μ g) bound to 51.28mg of SLs on 10 μ m PS beads.

BLOCKING THE SL ARRAY WITH 2%BSA, 10%GLYCEROL BEFORE USING THEM AS SPE WITH PURIFIED GLYCOPROTEIN

Once the overall protein binding capacity had been established for SL1 and SL2, the next step was to determine the impact of pre-blocking the beads with BSA and

glycerol on the overall binding capacity of the beads. In all previous work with the SL array, 2%BSA has been used to block any non-specific protein-peptide interactions from occurring, while the 10%glycerol has been used to prevent any weakly binding glycoproteins from binding to the beads. Both of these components in the blocking buffer have allowed the focus to shift more towards the binding of the higher affinity glycan portions of the glycoproteins, rather than the non-covalent interactions that occur between the 12-mers peptides on the beads or the protein portion of the glycoproteins. Preventing the binding of these nonspecific proteins is also important for downstream protocols and subsequent mass spectrometry analysis of the complex glycans in the samples, but that will be described in more detail in Chapter Four.

To analyzing the impact of the blocking buffer, the procedure for the experiment described above was performed again, with 0.25mg of OVA and BSM incubated with 51.28mg of 10 μ m polystyrene SLs. However, prior to the incubation, the beads were blocked with 1mL of the blocking buffer, 2%BSA, 10%glycerol in PBS_L, for 15min. Once the 15min of blocking was finished, the blocking buffer was removed from the beads, and directly replaced with the purified glycoprotein in 1mL of PBS_L, and incubated overnight at room temperature on a rocker.

As shown in Figure 3.5 below, all of the samples looked similar, with about 300 μ g of protein detected in the washes and about 100 μ g of protein detected in the elutions. This total protein recovery was much higher than what was previous seen when not pre-blocking the beads with BSA. The amount of BSA added to each sample for blocking is 1mL of 2%BSA, which calculates to 2g/100mL, or 20mg/1mL, which is 80x more BSA than the 0.25mg of glycoprotein used in these experiments. However, as

previously determined the amount of BSA retained by 51.28mg is around 200 μ g. When calculating the total protein collected from the washes and elutions, around 400 μ g of the total 450 μ g, or 80%, of protein added to the samples was recovered. This alludes to the fact that the target proteins of interest are more easily washed and eluted off of the beads after the beads have been previously blocked with BSA.

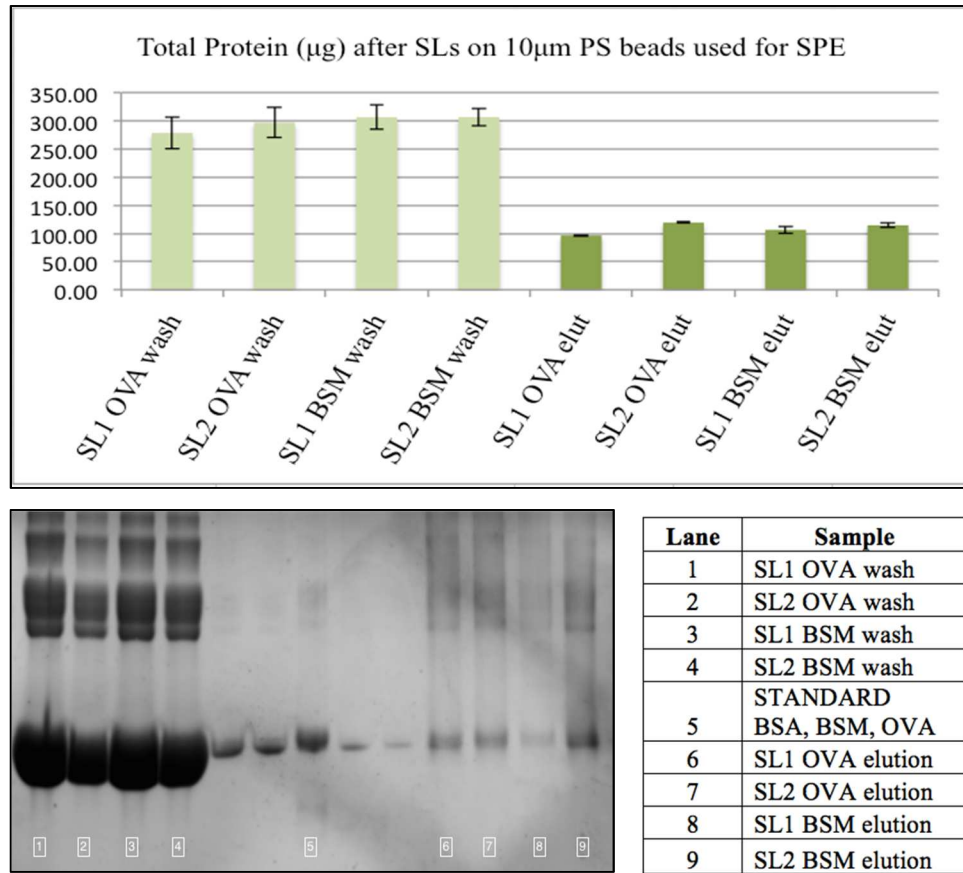


Figure 3.5 Total protein (μ g) recovered from washes and elutions after using SLs on 10 μ m PS beads for SPE (top). Samples run on SDS-PAGE and stained with Coomassie Blue to detect protein bands (bottom left), and corresponding lanes with each sample (bottom right).

To get a visual of the BSA to Glycoprotein ratios in the samples, they were run in an 8% SDS-PAGE gel and stained with Coomassie Blue. The samples proved to run terribly in the gels, the lanes were overloaded and contributed to significant bleed-over

throughout the gel, however the amount of BSA used in the blocking of these samples was determined to be too high, and was eventually adjusted to a 0.1%BSA concentration prior to the glycan mass spectrometry analysis protocol, but that will also be described in detail in chapter 4.

BLOCKING THE SL ARRAY WITH 2%BSA, 10%GLYCEROL BEFORE USING THEM AS SPE WITH MCF10A AND MDA-MB-231 GLYCOPROTEIN EXTRACTS

The next step was to repeat the same experiment above, but with breast cell line glycoprotein extract instead of purified glycoprotein. The incubation conditions included blocking 51.28mg of SL4 and SL5 with 2%BSA, 10%glycerol in PBS_L, and then incubated with 0.10mg of MCF10A and MDA-MB-231 overnight at room temperature on a rocker. The following day, the samples were washed with three 1mL aliquots of PBS_L, and then eluted with three 1mL aliquots of the blocking buffer (1M d-fructose in 20%acetic acid/diH₂O, pH=2.5). The total protein recovered from the washes was very similar to previous experiments, and the total protein recovered from the elutions was roughly half of the amount seen in the previous experiments with purified glycoprotein, which was expected because 100µg of starting material was used instead of 250µg. However, the relative amounts of BSA to the glycoproteins of interest in the cell line extracts was still high, as seen in the 8% SDS-PAGE in Figure 3.6 below.

Interestingly, the 3kDa MWCO Amicon centrifuge tubes for the sample buffer exchange and concentration were re-used from the previous experiments that included OVA and BSM, and reminences of those purified glycoprotein were seen as a contaminant in the 8% SDS-PAGE with the breast cell line extracts. Since, SL1 and SL2

were used in the purified glycoprotein experiments, the protein carryover between experiments had to be from the 3kDa MWCO spin tubes, which also gives at least some explanation as to why the total protein recovery from the earlier experiments were so low, because at least some of the protein stuck to the filters of the centrifuge tubes. Despite the protein contamination from previous experiments, the amount of BSA in the blocking buffer was still too high for the amount of glycoprotein of interest used in the extraction, so the blocking buffer was adjusted once again, to 10% glycerol, 1% BSA in PBS_L for all future experiments.

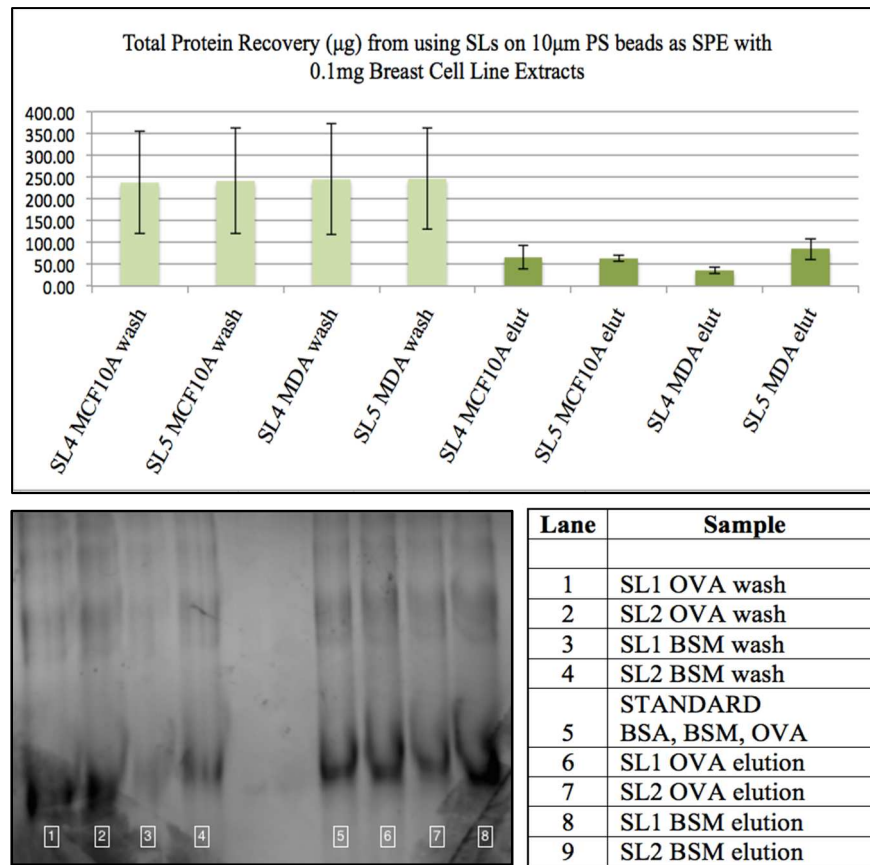


Figure 3.6 Total protein recovery (μg) from 51.28mg of SL4 and SL5 on 10 μm PS beads used as SPE with breast cell line secreted extracts from MCF10A and MDA-MB-231 (top). Samples run on SDS-PAGE and stained with Coomassie blue (bottom left), and corresponding lanes with each sample (bottom right).

BLOCKING THE SL ARRAY WITH 1%BSA, 10%GLYCEROL BEFORE USING
THEM AS SPE WITH MCF10A AND MDA-MB-231 GLYCOPROTEIN EXTRACTS
WITH ELUTION BUFFER ADJUSTED TO PH = 4.5

Once the elution method had proved to extract consistent amounts of protein from both the purified glycoprotein and cell line extracts, the next step was to adjust the elution buffer in order to preserve the carbohydrates, and prevent any hydrolysis of terminal saccharides. The elution buffer (1M d-fructose in 20%acetic acid/diH₂O) that had been used in the previous experiments was pH=2.5, which is reported to be too low for carbohydrates to remain intact. For the next experiments, the pH of the elution buffer was adjusted to 4.5, and tested to see if it could still extract enough material for subsequent N-linked glycan mass spectrometry analysis.

ALIGNING THE SPE PROTOCOL WITH THE MASS SPECTROMETRY
PROCEDURE

Keeping in mind the goal of assessing the glycans present in each cell line extract, the assay that had been chosen for this N-linked glycan analysis was the Waters GlycoWorks Rapifluor-MS N-Glycan kit, which will be described in further detail in the next chapter. However, this protocol requires no more than 15µg of total protein initially, regardless whether it is glycosylated protein or not. This presented two challenges, the first was to be able to extract that much protein using as gentle elution conditions as possible, and the second was to ensure that the majority of the 15µg for the Waters kit is mostly glycoprotein of interest, as apposed to the BSA that was used in the blocking step. The beads, used in Figure 3.7, shown below were blocked with 10%glycerol, 1%BSA in

PBS_L, and the starting material for the samples was 100µg of SKBR3 and T47D incubated with SL1, 2, 4, and 5.

On average, only 5-10µg of glycoprotein extract was recovered from each sample with the newly adjusted elution buffer, which is significantly lower than the total protein eluted with the elution buffer of pH=2.5, however the total protein was also lower than the previous experiment because 1%BSA was used in the blocking buffer instead of 2%BSA. Upon examining the samples in an 8% SDS-PAGE, the samples still seemed to be comprised of mostly BSA, showing a prominent protein band around the 50-60kDa area.

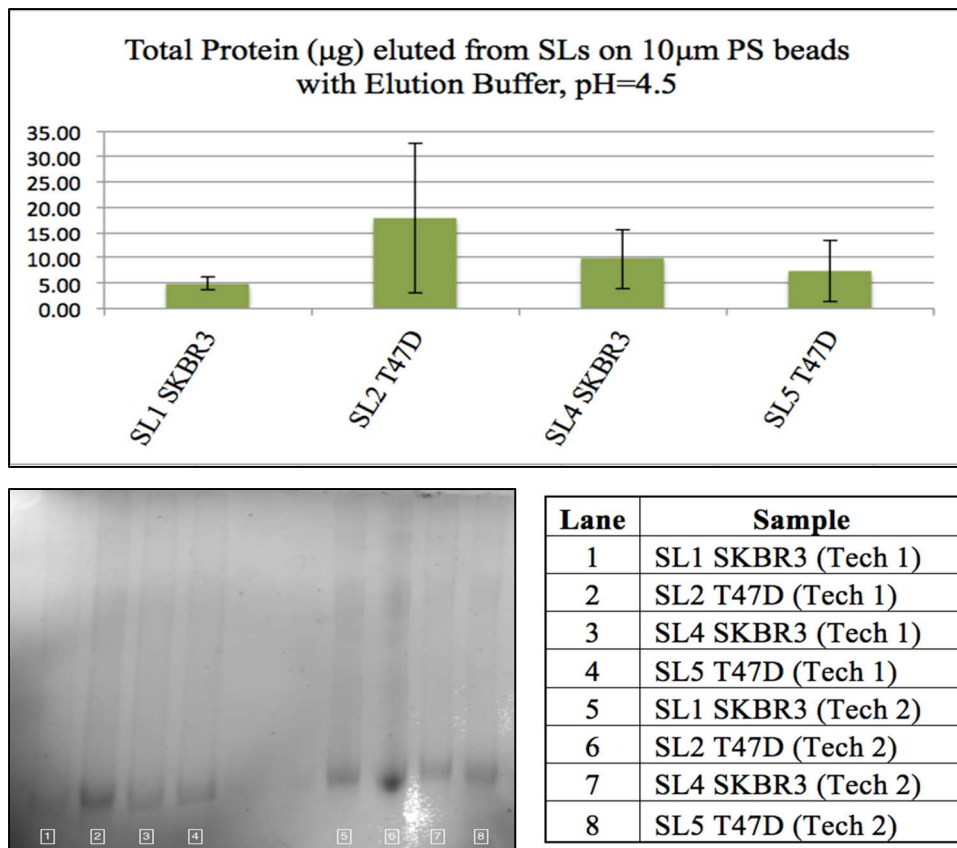


Figure 3.7 Total Protein (µg) of breast cell line secreted extracts eluted from SLs on 10µm PS beads using elution buffer adjusted to pH = 4.5 (top). Samples run on SDS-PAGE and stained with Coomassie blue (bottom left), and corresponding samples and lanes (bottom right).

Keeping the 15 μ g protein maximum in mind for the next procedure, the amount of BSA in the blocking buffer was dropped to 0.1% solution, which compared to the 150 μ g of cell line glycoprotein extract, was still roughly 6.5x more than the glycoprotein of interest. The procedure above was doubled to ensure enough protein material was eluted.

REGENERATING THE SYNTHETIC LECTINS FOR FUTURE USE

The synthesis of the Synthetic Lectins has become much less physically tedious and time consuming since the advent of automated amino acid coupling, however the cost still remains large due to the expenses associated with the amino-functionalized monosized beads and the cost of purified and protected amino acids, in particular the non-naturally occurring diaminobutanoic acid, which is >\$100/gram. For this reason, the SL array as a research tool is much more economic and marketable if the SLs could be regenerated after use, instead of being thrown away. In order to identify new 12mer peptide sequences to be attached to the SLs, a pool of beads with randomly generated sequences, via the split-and-pool method, is incubated with glycoprotein extracts, as previously described in Chapter One. The promising looking beads are individually selected, and the sequences are determined by the Edman's Degradation process. In order for this to work properly, all of the bound material on the beads must be removed. This has been accomplished by alternating 30min acid and base washes, with 70%TFA in diH₂O and 0.1M NaOH in diH₂O. All of the SLs used in the experiments in this chapter had been recycled at least 3-5x each, with the regeneration protocol as six alternating 30min washes of 70%TFA in diH₂O and 0.1M NaOH in diH₂O, and then finally washed twice with PBS_L.

CHAPTER 4

USING THE SL ARRAY FOR SOLID PHASE EXTRACTION TO ANALYZE N-LINKED GLYCANS IN BIOLOGICAL SAMPLES

BACKGROUND

The Synthetic Lectin Array has been shown to differentially bind to purified glycoproteins, and has also demonstrated an ability to discriminate between healthy and cancerous cell line extracts that include a plethora of complex glycoproteins. However, the fluorescent signal from the beads does not necessarily mean that different glycoproteins are binding to different SLs in the array, or that glycans are even playing a role in the signals at all. Again, the original nine SLs were also selected for their differential binding to nonglycosylated proteins, such as BSA, and therefore do not only have a preference for glycoproteins. Since, the FITC used to fluorescently tag the extracts only binds to primary/secondary amines, which are not found on glycans and only on the lysines and N-terminus of a protein, the fluorescent signal only represents the amount of FITC bound to proteins bound to the SL.

Therefore, the next step in developing the SL Array as a cancer diagnostic tool was to prove that the glycoproteins in these extracts were playing a role in the SL Array's ability to discriminate between samples. These next experiments were not only used to demonstrate that ability, but also to serve as a method of feedback analysis, which will be valuable in the making of future SLs to tailor and fortify the array. These experiments also showed that the SL Array could successfully be used as a solid phase extraction

material for different types of glycoproteins, and could serve as a prognostic or research tool as well.

The following procedure was performed to examine the binding relationship of SL1, 2, 4, and 5 with secreted extracts from the same four breast cell lines (MCF10A, SKBR3, MDA-MB-231, and T47D) that were used in the experiments outlined in Chapter 2 and 3. Briefly, the SPE protocol outlined in Chapter 3 was performed with the four SLs and each cell line extracts. The eluted glycoproteins from each sample were then processed, by removing the N-linked glycans from them, which were then isolated, tagged, and examined by LC-ESI-MS.

The efforts outlined in this chapter, including the identification of the N-linked glycans, are all courtesy of the enzyme PNGase F, which quickly and efficiently releases fully intact N-linked glycans from their asparagine anchors. Unfortunately, a corresponding enzyme that releases fully intact O-linked glycans has not yet been discovered, likely because O-linked glycans can be anchored with different types of monosaccharides. Furthermore, the structures of O-linked glycans are more variable than N-linked glycans, due to the differences in biosynthesis, and are therefore much more challenging to properly identify. All N-linked glycans begin from a common template, whereas O-linked glycans are built one monosaccharide at a time de novo, which makes them more variable in shape, size and composition.

As a result of all these aforementioned challenges, this chapter will only discuss the N-linked glycans observed in the four breast cell line secreted extracts. However, in the future, a parallel O-linked glycan analysis should be done to get a more holistic understanding of glycosylation and how it relates to breast cancer, particularly since

mucin-like aberrant glycosylation has been shown in patient tumor biopsies.¹⁷ This subject will be briefly discussed again in Chapter 5, the future works section.

THE GLYCOWORKS RAPIFLUOR-MS N-GLYCAN KIT AND LC-ESI-MS PROTOCOL

In order to accomplish this goal of identifying the N-linked glycans present in the breast cell line extracts, each SL was used as a SPE agent, as described before in Chapter 3. Two 1.5mL Eppendorf tubes were filled with ~162mg of SLs on 10 μ m PS beads each, blocked with 1mL of 0.1%BSA, 1%glycerol in PBS_L for 15min, incubated with 125 μ g of secreted cell line extract, and left overnight on a nutator at 25°C. The following day, the samples were washed 3x with PBS_L, and eluted 3x for 1hr each with the elution buffer, 1M d-fructose in 10%acetic acid/diH₂O, pH 4.4. Once the three aliquots of elution buffer were collected, the glycoprotein material was filtered, buffer exchanged, and desalted before it could be used in the next step. Once these samples were prepared, they were processed with the Waters N-linked glycan preparation kit, separated by liquid chromatography using a Bridged-Ethylene Hybrid (BEH) Amide Column (Waters), and then analyzed by mass spectrometry using an electrospray ionization (ESI) Orbi-Trap in positive ion mode.

The Waters GlycoWorks RapiFluor-MS N-Glycan Kit was used in this project because it contained all of the materials necessary in order to cleave, label, and isolate N-linked glycans from many glycoprotein samples in under 3hrs. Using this kit was argued to be cost effective because of the time it would save over building another protocol to achieve the same goal, but also because of the emphasis it put on the sample cleaning

prior to LC-MS. This last factor is of the utmost importance when it comes to liquid chromatography separation because contaminated samples are both a waste of time and computing power because they are often compromised or suppressed by the background signal, and they also have the potential to clog the column, which is an expensive replacement.

Briefly describing the three modules of this kit, the first step involved the deglycosylation of N-linked glycans from the glycoproteins in the sample. This involved heating the samples to 90°C for 3min in the presence of a surfactant to denature the protein. The samples were then cooled down for 3min, and Rapid PNGase F (New England Biolabs) was added to the samples for 5-6min. As mentioned before, PNGase F is a highly useful enzyme that cleaves N-linked glycans, always between the anchoring N-Acetylglucosamine of the glycan and the asparagine of the protein, preserving the fully intact N-linked glycan and converting the asparagine of the protein, preserving the fully intact N-linked glycan and converting the asparagine amino acid to an aspartic acid (Shown in Figure 4.1 below). This leaves a primary amine on the newly freed reducing end of the N-Acetylglucosamine, but this primary amine is only stable for up to 2 hours, and is therefore labeled soon after it is liberated from the peptide.

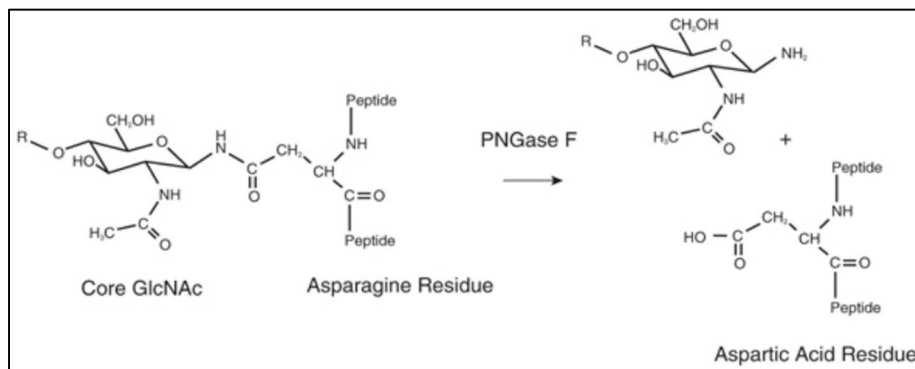


Figure 4.1 The enzyme PNGase F cleaving an N-linked glycoprotein between the core GlcNAc and the asparagine amino acid.

The second step of the kit involved the labeling of the newly liberated N-linked glycans at the reducing end of the N-Acetylglucosamine, with the RapiFluor-MS molecule, which is shown below in a Figure 4.2. This glycan label greatly aids in the ionization of the glycans, which is essential for MS analysis because in order for the molecules to reach the detectors, they must have a charge to fly through the magnetic field. This is often a great concern when analyzing N-linked glycans because, with the exception of N-acetylneuraminic acid (Sialic acid), all of the other monosaccharides found in human N-linked glycans are of neutral charge, including mannose, galactose, N-Acetylglucosamine, and fucose. The RapiFluor-MS molecule is easily ionizable because it contains several secondary and tertiary amine groups, which are easily protonated to ensure that the N-linked glycans have a +1, but more commonly a +2 or +3 overall charge. This +2 or +3 charge is also important for the detection of these glycans because the mass detection range of the ion-trap is 500-2000Da, whereas the masses of most N-linked glycans are in the 2,000-4,000Da range. However, the MS detects molecules by their mass/charge ratio, so having molecules with a +2 or +3 charge allows the detection of molecules up to 4000-6000Da in size.

The third step in the Waters Kit is the cleanup module. This is necessary before running the samples through LC-ESI-MS system because in order to receive a clean signal, any excess materials, salts, or anything other than the N-linked glycans, must be removed to ensure a clean separation of the N-linked glycans by the liquid chromatography column. This main feature of this cleanup module was a hydrophilic interaction liquid chromatography (HILIC) μ Elution plate that was attached to a vacuum manifold. The μ Elution plate wells contained an amino propyl resin that worked as both a

hydrophilic interaction and a size exclusion column, which effectively collected the N-linked glycans as the rest of the materials passed through, which were then washed, eluted, and concentrated prior to analysis by LC-ESI-MS.

The individual samples were loaded into individual wells of the μ Elution plate with 15:85 diH₂O/ACN, and washed with 1:9:90 formic acid/diH₂O/ACN to remove proteins, including PNGase F, and any other hydrophobic or other larger materials. The N-linked glycans were then eluted from the plate wells using 90-100 μ l of 200mM ammonium acetate in 5% ACN, and were further concentrated to \sim 20 μ l using a vacuum centrifuge on a heat setting of 40°C. Once these samples were prepared, they were kept at -20°C until they were injected into the LC-ESI-MS system no more than a week later.

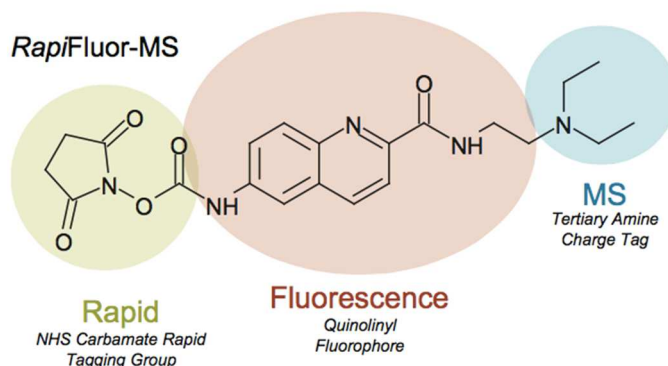


Figure 4.2 RapiFluor-MS molecule that binds to the reducing end of the terminal GlcNAc in 5min.

BRIDGING THE GAP BETWEEN SPE SAMPLES AND LC-ESI-MS PROTOCOL

As mentioned at the end of Chapter 3, the GlycoWorks RapiFluor-MS N-Glycan Kit only processes samples that maximally contain 15 μ g of protein, regardless of whether the samples is comprised of glycosylated protein or not. This is because the amounts of

surfactant, PNGase F, and RapiFluor-MS for each sample are calculated based on a ratio consistent with 15µg of glycoprotein. For example, extraneous BSA in the sample would interfere with the surfactant-to-protein ratio, allowing incomplete denaturation of all the protein in the sample, and therefore potentially incomplete N-linked glycan cleavage by enzyme PNGase F, due to the steric hindrance caused by the proper protein folding. For this reason, special emphasis had to be put on the amount of BSA used in the blocking buffer in Chapter 3, so that it was still a high enough concentration to limit nonspecific binding interactions, but didn't interfere with the downstream processing and analysis.

Prior to introducing the samples to the N-linked Glycan Kit, additional steps had to be taken with the glycoprotein samples eluted from the SLs. These steps involved an initial buffer exchanging and samples concentration into diH₂O using 10kDa MWCO centrifuge tubes (Amicon). The samples were then desalted through 5mL Zeba Desalting Columns (Pierce Fisher Scientific), to remove excess acetic acid, d-fructose, or salt that could interfere with any of the components in the Waters kit.

After this, the amount of BSA in the samples was still a concern, so the samples were subjected to an albumin depletion kit (Pierce), which used albumin-affinity beads and a microcentrifuge to selectively bind and remove albumin from the samples. This last step is a common technique used when dealing with blood samples, because as mentioned before, albumin is the main blood protein, which circulates at concentrations of 2-5% in humans. The samples were then further concentrated to ~25µl by vacuum centrifugation at 40°C. 1µl was then taken from each of these samples, diluted to 10µl in diH₂O, and run in triplicate to determine the protein concentration via Bradford Assay.

In general, the 25µl samples only contained 10-20µg of protein, which was likely a result of all the steps in the sample processing, which involved transferring the samples to multiple different vessels. In any event, all of the samples were used in their entirety for processing by the Water Glycan Kit. Using the full amount of the samples, regardless of the concentration, likely introduced an inconsistency the samples. However, since the sensitivity of this entire method was not known, and the Waters Kit purchased only processes 24 samples, each of which were time consuming and expensive, the risk of not getting a signal far outweighed the subtle differences in the initial protein concentrations of the samples.

HILIC BEH AMIDE COLUMN CHROMATOGRAPHY

The column used for the liquid chromatography was a custom Waters 1 x 150mm Acquity UPLC Glycan BEH (bridged ethylene hybrid) Amide column with 1.7µm particle size with 130Å pockets. This hydrophilic interaction liquid chromatography (HILIC) column separates polar molecules, such as N-linked glycans, using hydrophilic interaction. UPLC grade ACN was used as the stationary phase and 50mM ammonium formate in UPLC grade diH₂O, pH adjusted to 4.4 with UPLC grade formic acid, was used as the mobile phase. The solvents were used over a 60-minute gradient as shown below in Figure 4.3.

Only 5µl of sample was maximally injected in a single run, but on average, only 1-2µl was injected per sample. After each run, the chromatogram of each sample was initially assessed by the sharpness and intensity of several common N-linked glycans peaks. These common glycans included Mannose-5, Mannose-6, Mannose-7, Mannose-8,

FA₂, FA₂S₁, and FA₂S₂. This rough template ensured that good chromatographic separation occurred with the sample, and that the data generated was of comparable quality. If these common glycan peaks were not present, or did not have sharp peaks, meaning that they separated poorly, the samples were rerun until a passable chromatogram was achieved.

Flow rate:	0.4 mL/min				
Gradient:	Time (min)	Flow rate (mL/min)	%A	%B	Curve
	0.0	0.4	25	75	6
	35.0	0.4	46	54	6
	36.5	0.2	100	0	6
	39.5	0.2	100	0	6
	43.1	0.2	25	75	6
	47.6	0.4	25	75	6
	55.0	0.4	25	75	6

Figure 4.3 Gradient used for HILIC protocol. Solvent A was 50mM ammonium formate (pH 4.4), and Solvent B was pure acetonitrile.

THE N-LINKED GLYCAN BIOSYNTHESIS PATHWAY AND CANCER

As alluded to before, the ability to rapidly predict and identify +100 different types of N-linked glycans is a result of the well defined human N-glycan biosynthesis pathway. This biosynthesis begins with a template, as shown in Figure 4.4, in the Endoplasmic Reticulum, where it is attached to an asparagine of a protein, and proceeds through the Golgi apparatus, from the cis- to the medial- to the trans-golgi.²⁹ All along the way, monosaccharides are cleaved off from the terminal ends by enzymes called glycosidases, and other monosaccharides are attached to the terminal ends by enzymes called glycosyltransferases. Over 200 glycosyltransferase genes have been identified in the human genome to date³⁰, but all have strict substrate specificity. For example, Fut8

only transfers a fucose from GDP-β-L-fucose to the innermost N-acetylglucosamine (GlcNAc) in a α1-6 linkage, and is the only enzyme to do so.

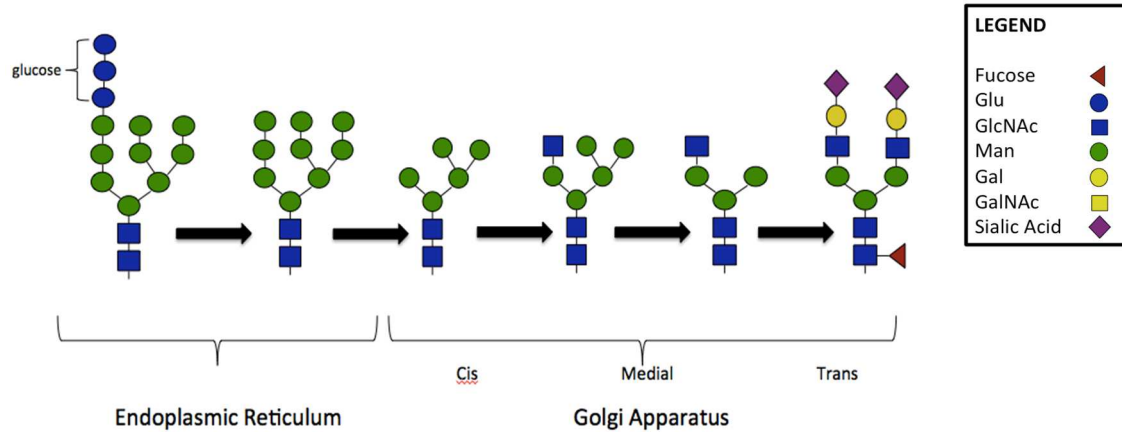


Figure 4.4 The human N-linked glycan biosynthesis pathway. A step-wise process including many specific glycosidases and glycosyltransferases to cleave and add individual monosaccharides to the terminal ends of the glycans.

As a result of this enzyme specificity, the changes that have been observed in cancer-associated aberrant glycosylation are a result of changes in the activity of these glycosidases and glycosyltransferases. This can happen for many reasons, including epigenetic changes to these genes by methylation, the availability of substrates, and the mislocation of these glycan-modifying enzymes.³¹ Therefore, the focus of truly aberrant N-linked glycosylation should be generally limited to the rules already defined by the past decades of research in the field.³² That is to say, there have not been many novel glycosidic bonds or unprecedented structures reported in the literature to date. Although possible, it is highly unlikely that a mutation would occur in an existing glycosyltransferase gene that would allow it to catalyze a less specific or new chemical reaction.

This knowledge was highly beneficial in the predicting and identifying of N-linked glycans because it ruled out any bonds that have not been seen in human glycans before. For example, a galactose has never been reported to bond to another galactose, and a fucose has never been observed to bond to any sugar other than GlcNAc. In fact, all of the possible N-linked glycan bonding combinations (With the exception of polysialic acid) are depicted in the final glycan structure on the bottom right of Figure 4.4.

N-LINKED GLYCAN PEAK IDENTIFICATION

Individual N-linked glycans were identified by their mass to charge ratios, and the presence of a corresponding isotope pattern. To aid in this process, an extensive excel spreadsheet was created, which included permutations of all known N-linked glycan structures that have been observed in human tissue samples. This glycan database included +1000 N-linked glycans of different masses, but was easy to manage because of the bonding patterns of human N-linked glycans, discussed before. Glycans peaks that were not identified using the excel spreadsheet, were pieced together using an isotopic glycan mass calculator on excel, along with the glycan bonding rules previously discussed.

The MS computational software was able to assign the charge of a peak by its corresponding isotope pattern. These isotope patterns are almost always seen in biological molecules above a certain mass because the frequency of carbon-13 in nature is ~1%, while nitrogen-15 is ~0.4%, and oxygen-18 is ~0.2%. Therefore, since carbohydrates are roughly comprised of the general molecular formula $(\text{CH}_2\text{O})_n$, or $40n$, and the smallest N-linked glycan, Man-3, which has a molecular weight of 892.3172Da,

an individual molecule is ~25% likely to have an isotopic atom because a Man-3 has 22 carbon atoms and 22 oxygen atoms. Therefore, if any isotopes versions of the molecule are detected, then there will be a pattern of peaks corresponding to the charge of that peak. If the peak has a +1 charge, then the isotope peaks will be exactly 1Da apart. If the peak has a +2 charge, then the isotope peaks will be exactly 0.5Da apart. If the peak has a +3 charge, then the isotope peaks will be exactly 0.33Da apart.

Any peaks that had a charge of +4 charge or greater were disregarded because they likely to be proteins, or some molecule other than a glycan, generally because the corresponding mass was larger than 5000Da. On the other end, any peaks that had an a charge of +1, were generally disregarded because they either represented some impurity or a glycan of a size smaller than <2,000Da, of which there are only so many and were all generally detected in greater abundance in their +2 and +3 charged states.

Using the BEH Amide column with the solvent gradient provided above in Figure 4.3, the bulk of all the glycan data was collected between the retention times of 20-35min, and in the mass range of 700-1800Da. All other material seen before 15min or after 45min was also likely to be protein crashing out of the column before or after the hydrophilic separation, with the acetonitrile.

ISOTOPIC GLYCANS AND IDENTIFICATION CHALLENGES

Without the added benefit of MS-MS data, or data collected after molecule collisions are induced to provide additional structural clues, there are some glycans structures that could only be narrowed down to several candidates. The more prominent structure of the glycans, or the ones referred to more in the literature, were reported in

bold, while the lesser reported structure were in normal font. In addition to isomeric carbohydrates, another confusing mass relationships hindered the identification of some peaks. This mass relationship was that one sialic acid is 291.0954Da and two fucoses are 292.1158Da. These two masses are only separated by 1Da, and thus made it challenging to know whether a glycan contained one sialic acid, two fucose molecules, especially on larger structures than may be obscured by the presence of isotopes. However, since there is a difference in the relative polarity of two fucoses versus one sialic acid, the relative retention time of the surrounding glycans could be used as a reference, as more polar molecules elute later from the HILIC column. Using this information, the most probably glycan structure could be inferred.

N-LINKED GLYCAN PROFILES FROM MCF10A, SKBR3, MDA-MB-231, AND T47D CELL LINE GLYCOPROTEIN EXTRACTS

Before analyzing the N-linked glycans extracted using the SLs, a background N-linked glycan profile had to be established for each of the four cell lines used. In order to do this, the secreted cell line extracts were concentrated to ~1mg/ml in PBS 1X, and 15 μ l was then taken and processed through the Waters N-linked glycan kit exactly as described before. Once a full profile had been established for each individual cell line, a plot of the N-linked glycans extracted from each SL was devised to better visualize their preferred glycan types.

Not surprisingly, the number of individual glycans observed in these secreted cell line extracts was staggeringly large, and generally very similar, which was to be expected because they are all derived from human breast tissue. It was surprising to see that the

MCF10A, the healthy control cell line, had the most unique N-linked glycans detected in the secreted extracts, with 129 glycans of unique masses identified +20 more unknown peaks. This was all detected from injecting 1 μ L of the cell line extract into the LC-ESI-MS system.

The other cancerous cell line extracts did not present as many unique N-linked glycans as MCF10A. This again, came as a surprise because it was anticipated that the glycosylation structures would become much more erratic and random with tumorigenesis, resulting in more aberrant glycosylation and unique glycan masses. However, this seemed not to be the case. For the cancerous cell line extracts, SKBR3, MDA-MB-231 and T47D, there were 100 identified glycans +25 unknown peaks, 95 identified glycans +20 unknown peaks, and 101 identified glycan +30 unknown peaks, respectively. The structures of all glycans discussed in this chapter, due to limited space are not shown in the data tables, but are provided in the Appendix B section in the tables B.1 through B.5. The hybrid structures were not provided because without MS/MS data, the proposed structures would be largely speculative.

The total number of individual glycan peaks and overall intensity from the MDA-MB-231 and T47D cell line secreted extracts was lower than the intensity and number of individual glycan peaks observed during their respective extractions using SL1. This meant that the MDA-MB-231 and T47D cell line extract samples were suppressed somehow because they should encompass all the glycans observed in the samples eluted from the four SLs. This was likely due to the fact that these were the first two samples injected onto the column, the first time using a HILIC column, and the first glycan mass analysis ever done at The University of South Carolina. Needless to say, there was a

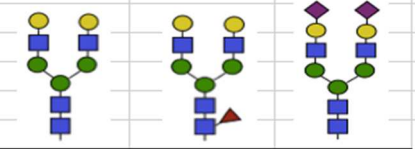
learning curve to overcome, however, unfortunately the Waters N-linked glycan kit only had 24 samples preparations, so the data collected could not be redone and had to be used for further analysis. To overcome this disparity in raw intensities between the cell line extracts, the data from each cell line was normalized to the highest peak intensity within that sample. This put all of the data between cell lines on the same scale, and highlighted the glycan intensities in relation to one another within each sample. In general though, the data presented is qualitative, and could only be considered quantitative after more trials allow for a tighter protocol to be established.

The wide variety of N-linked structures observed from these samples were divided and analyzed by their larger subgroups. Those subgroups of N-linked glycans are high mannose, bi-antennary, tri-antennary, tetra-antennary, and hybrid structures. The most abundant glycan types across the board for all of the samples were the high mannose and the bi-antennary glycans. A few of these glycans peaks were consistently the highest signals within each sample, so these were recorded and used as a reference for the total intensity per sample. This is essentially what the MS computer does with every sample coming off of the LC column, which is scale everything to the “base peak” or scale every peak to the peak with the greatest intensity.

These highly abundant N-linked glycans were FA2G2 and A2G2S2, which were detected at 2098.8231Da and 2534.956Da, respectively. These were also conveniently used as markers to generally show the extent of fucosylation and sialylation on biantennary glycans, such as A2G2. To complete this comparison, A2G2, although not consistently one of the peaks of higher intensity, was also recorded as the unfucosylated and unsialylated control (1952.8Da). The relative abundance of these glycans are shown

in Table 4.1 below. The FA2G2 and A2G2S2 glycans were highlighted in light green if they were 3x higher in intensity than A2G2, highlighted in dark green if 10x higher intensity than A2G2, and highlighted in red if they were lower in intensity than A2G2.

Table 4.1 Glycans FA2G2 and A2G2S2 compared in relative abundance to the glycan A2G2, within each samples. Light green shading represents 3x increase in peak intensity, dark green is a 10x increase, and red represents a decrease in relative peak intensity.









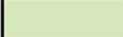

	A2G2	FA2G2	A2G2S2
MCF10A	1.29E+04	1.16E+05	1.54E+05
SKBR3	3.19E+05	6.52E+05	3.60E+05
MDA-MB-231	7.89E+03	4.11E+04	8.85E+03
T47D	2.21E+03	1.44E+04	1.18E+04
MCF10A SL1	1.28E+04	3.31E+04	2.38E+04
SKBR3 SL1	2.77E+03	2.72E+04	3.55E+04
MDA-MB-231 SL1	6.65E+03	4.31E+04	7.16E+04
T47D SL1	1.74E+03	1.71E+04	3.10E+04
MCF10A SL2	1.26E+04	4.28E+04	1.62E+04
SKBR3 SL2	1.08E+03	7.34E+03	4.85E+03
MDA-MB-231 SL2	2.90E+03	1.59E+04	1.32E+04
T47D SL2	5.26E+03	4.79E+04	1.17E+05
MCF10A SL4	1.91E+03	5.22E+03	1.23E+03
SKBR3 SL4	2.26E+02	1.74E+03	2.95E+03
MDA-MB-231 SL4	1.43E+03	7.36E+03	1.43E+04
T47D SL4	1.39E+03	7.02E+03	1.87E+03
MCF10A SL5	3.25E+03	4.13E+03	1.36E+03
SKBR3 SL5	4.41E+02	2.62E+03	3.21E+02
MDA-MB-231 SL5	4.59E+03	2.03E+04	7.58E+03
T47D SL5	3.71E+03	3.98E+04	1.94E+04

Of all the samples, only MCF10A showed lower intensity of A2G2S2 in SL4 and SL5, and SKBR3 showed a lower intensity of A2G2S2 in SL5, compared to A2G2. None of the samples showed lower FA2G2 than A2G2, and T47D appeared to have the most consistently higher signal of FA2G2 for the cell line extract and all of the SLs.

Interestingly, for MCF10A the relative amount of A2G2S2 to A2G2 was only significantly higher in the cell line extract, and was roughly equal or lower in all of the samples eluted from the SLs. The decrease in A2G2S2 when MCF10A was eluted with SL4 and SL5, again demonstrated that the polarity of the SLs plays a major role in the binding interactions with the glycans. This data suggests that MCF10A had less sialylated material bound to SL4, which then led to less sialylated material being eluted from SL4. In contrast, the MDA-MB-231 extract appeared to have a lot of sialylated material by the maintaining the same ratio of A2G2S2:A2G2 when eluted from SL4. The three cancer cell lines had the greatest A2G2S2 signals when eluted from SL1, and slightly lower signals when eluted from SL2. This also demonstrates that it is easier to elute more polar glycans from the more hydrophobic SLs. These numbers will be revisited in the bi-antennary section.

The peak intensities for each glycan in the tables below are represented by one of the colors in Table 4.2 below, which adds up to seven colors plus the blank, reducing the data to a 3-bit scale (2^3). These peak intensities are in arbitrary units, however, they representative the relative abundance of the detected masses within a sample.

Table 4.2 Peak intensity and color scale.

Color	Peak Intensity
	>1.0E5
	>1.0E4
	>7.0E3
	>3.33E3
	>9.0E2
	>5.0E2
	>3.0E2
	<3.0E2

All peak intensities were grouped into one of these levels and were used when normalizing the data within each sample. Any peaks that had registered intensities of below 5.0E2 were noted as on the detection of threshold, and any peaks below intensities of 3.0E2 were disregarded because they were too close to the overall background noise to be considered significant.

Table 4.3 shown below contains some properties of all of the monosaccharides (-H₂O because included in a polymer) observed in human N-linked glycans. All of these sugars are polar, however, only sialic acid has a negative charge, whereas the other four molecules are neutral. In addition, GlcNAc is slightly more polar than mannose and galactose, which are isomers of each other, because it contains an extra nitrogen and a carbonyl functional group. On the other hand, fucose is slightly more hydrophobic than mannose and galactose because it has one less oxygen. This difference in polarity between sialic acid and fucose was useful in determining the structure of certain glycan peaks because, as mentioned before, one sialic acid is 1Da less in mass than two fucoses.

Table 4.3 Properties of all glycans found in human N-linked glycosylation. Polarity rank based on sum of oxygen and nitrogen atoms compared to carbons atoms in monomer.

monosaccharide	mol. formula	mass (Da)	polarity rank
Sialic Acid	C ₁₁ H ₁₇ O ₈ N ₁	291.0954	1
GlcNAc	C ₈ H ₁₃ O ₅ N ₁	203.0794	2
Galactose	C ₆ H ₁₀ O ₅	162.0528	3
Mannose	C ₆ H ₁₀ O ₅	162.0528	3
Fucose	C ₆ H ₁₀ O ₄	146.0579	4

Depending on the relative retention times of the other surrounding peaks, it could be inferred whether a peak contained one sialic acid or two fucose because it would elute off of the HILIC column at different times. If the glycan contained two fucoses, then it would be more hydrophobic and would elute earlier off of the column. But, if the glycan contained a sialic acid, then it would be more polar and elute later off of the column. This comparison between glycan polarity will be discussed in further detail when comparing the relative hydrophobicities of the four SLs to each other, which has a large impact over their glycan preferences.

HIGH MANNANOSE N-LINKED GLYCANS IN CELL LINE EXTRACTS

The first and simplest subgroup of N-linked glycans to be analyzed was the high mannose group, which after the first two GlcNAcs, only contains mannose monomers for the rest of the glycan. In terms of biosynthesis, the original N-linked glycan template contains nine mannoses and three glucoses, which are cleaved off as the template moves to the Golgi apparatus. This results in a Man9 (Or M9) glycan, or more mannoses can be removed using mannosidases to get the other structures of Man4, Man5, Man6, Man7, and Man8. All of these high mannose glycans were highly abundant in all four of the cell line extracts, and seems to be essential, at least among breast tissue. Man5, Man6, and Man7 were the most abundant of these in all cell lines, but no significant difference in the expression of these high mannose glycans was observed between the cell line extracts.

Table 4.4 High Mannose N-linked glycans from four breast cell line extracts. Raw data (left), data normalized to highest internal peak (right).

High Mannose								
	MCF10A	SKBR3	MDA-MB-231	T47D	MCF10A	SKBR3	MDA-MB-231	T47D
M4					0.8	1.0	0.6	0.6
M5					1.0	1.0	0.8	1.0
M6					1.0	1.0	1.0	0.8
M7					1.0	1.0	0.6	0.6
M8					1.0	0.8	0.6	0.6
M9					1.0	0.8	0.6	0.4

As noted before, the MDA-MB-231 and T47D cell line extract samples were relatively lower in overall intensity than MCF10A and SKBR3. In order to better compare the relative abundances of the glycans, the data of each cell line was internally normalized, by dividing each glycan by the highest glycan in that sample. This put all of the data between the cell lines on the same scale, as shown by the red heat diagram on the right of Table 4.4.

HIGH MANNANOSE N-LINKED GLYCANS ELUTED WITH SL1, 2, 4, AND 5

When analyzing the high mannose glycans eluted after the cell line extracts were incubated with the four SLs, the aforementioned relationship of hydrophobicity between the SLs resurfaced again. High mannose glycans, because they don't have any sialic acids or additional GlcNAcs, are relatively more hydrophobic glycans. For the cell line MCF10A, SL1 and SL2 had higher peak intensities than SL4 and SL5. SKBR3, and MDA-MB-231 had similar patterns with the highest peak intensities coming from SL1, and SL2, SL4, and SL5 were all comparably lower in intensity. For these three cell lines,

this correlated with the relative hydrophobicity of the SLs, which suggests that SL1 was the most hydrophobic SL and therefore would have the greatest preference for high mannose N-linked glycans. However, SL4 still had a large relative abundance of all six of the high mannose sugars, suggesting that high mannose glycans were a large component of the N-linked glycan profiles of the cell line secreted extracts.

Interestingly, the cell line T47D had a pattern that was opposite to SKBR3 and MDA-MB-231, where the smallest peak intensities came from SL1, and the other three SLs were comparably higher in peak intensities. This may suggest that the cell line MCF10A has a wide variety and even distribution of N-linked glycans because the polarity of the SL has a large influence on the glycan profile. This also may suggest that cell lines SKBR3 and MDA-MB-231 have more polar secreted extracts because the high mannose, or the most hydrophobic, glycans were only seen in their greatest abundance when eluted from the most hydrophobic SL, SL1. This would also suggest that T47D has a relatively less polar glycan profile because more hydrophobic glycans were bound and eluted from the three more polar SLs. These assertions, however, need more supporting evidence and will be re-examined while analyzing the other N-linked glycan subgroups.

Table 4.5 High mannose glycans from four breast cell lines eluted from SL1, 2, 4, and 5, raw data organized by cell line.

High Mannose	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5
	MCF10A	MCF10A	MCF10A	MCF10A	SKBR3	SKBR3	SKBR3	SKBR3	MDA-MB-231	MDA-MB-231	MDA-MB-231	MDA-MB-231	T47D	T47D	T47D	T47D
M4																
M5																
M6																
M7																
M8																
M9																

BI-ANTENNARY N-LINKED GLYCANS FROM CELL LINE EXTRACTS

The next subgroup of N-linked glycans analyzed was the bi-antennary group. This group contains a common core of two GlcNAcs, followed by three mannose sugars, to which two more GlcNAcs are attached. This is the simplest bi-antennary N-linked glycan, which is referred to as A2. From here, galactoses can be attached to the terminal GlcNAcs, sialic acids or more GlcNAcs can be attached to those galactoses, and fucoses can be attached to either the core GlcNAc and/or the antennary GlcNAcs. The “B” in the table refers to an additional GlcNAc that is attached to the base mannose that is attached to the second core GlcNAc. As referred to before, no MS-MS data was collected, so it was impossible to discern between A2B and A3, two isomers, with the information at hand. Both versions of these glycans have been reported in the literature, but emphasis will be put on A3 and not A2B for organizational purposes.

The bi-antennary glycans were the most abundant subgroup of N-linked glycans in all cell lines extracts and was the only group to have glycan peaks over 1.0E5. All four cell lines appeared to favor fucosylated bi-antennary glycans over their nonfucosylated counterparts, as there were more types of fucosylated glycans with more intense peaks, as shown in the bottom half of Table 4.6 below. However, as shown in the heat diagram to the right, the cell line MCF10A maintained a more even distribution of the bi-antennary glycans detected, as did the cell line SKBR3 to a lesser extent. Whereas, the cell line T47D had a shift in the distribution towards the bottom of the table, where the fucosylated glycans are listed. In addition, more glycans with multiple fucosylation sites (FF and FFF), or outer-arm fucosylation, were observed in SKBR3, MDA-MB-231, and T47D cell line extracts than in the MCF10A cell line extract.

Table 4.6 Bi-Antennary N-linked glycans from four breast cell line extracts. Corresponding fucosylated glycans on the bottom half of data. Raw data (left), and data internally normalized (right).

Bi-Antennary	MCF10A	SKBR3	MDA-MB-231	T47D	MCF10A	SKBR3	MDA-MB-231	T47D
A2					0.9	0.7	0.5	0.5
A2B					0.0	0.4	0.5	0.3
A2BG1					0.9	0.4	0.5	0.3
A2BG1S1					0.4	0.6	0.5	0.5
A2BG2A2					0.0	0.0	0.0	0.3
A2BG2S1					0.4	0.4	0.0	0.3
A2BG2S2					0.0	0.7	0.0	0.0
A2G2A2G2A1					0.0	0.3	0.0	0.0
A2G1					0.9	0.4	0.7	0.5
A2G1S1					0.9	0.6	0.5	0.5
A2G2					0.9	0.6	0.8	0.5
A2G2A2					0.4	0.0	0.0	0.3
A2G2S1					0.9	0.9	0.8	0.8
A2G2S2					1.0	1.0	0.8	0.8
FA2					0.9	0.9	1.0	1.0
FA2B					0.4	0.7	0.5	0.5
FA2BG1					0.9	0.7	0.7	0.7
FA2BG1S1					0.9	0.4	0.5	0.2
FA2BG2					0.9	0.0	0.5	0.0
FA2BG2S1					0.6	0.4	0.5	0.5
FA2BG2S2					0.4	0.3	0.0	0.0
FA2G1					0.9	0.9	1.0	1.0
FA2G1S1					0.7	0.6	0.5	0.3
FA2G2					1.0	1.0	1.0	1.0
FA2G2S1					0.9	0.9	1.0	0.5
FA2G2S2					0.9	0.7	0.5	0.5
FFA2					0.7	0.7	0.5	0.0
FFA2B					0.0	0.0	0.3	0.3
FFA2BG1S1					0.9	0.4	0.5	0.7
FFA2G1S1					0.0	0.6	0.3	0.5
FFFA2BG1S1					0.0	0.0	0.0	0.3

BI-ANTENNARY N-LINKED GLYCANS ELUTED FROM SL1, 2, 4, AND 5

When analyzing the bi-antennary glycans eluted after the cell line extracts were incubated with the four SLs, the patterns were very similar to what was seen when analyzing the high mannose samples. By sorting Table 4.7 by the cell lines instead of the SLs, it became apparent again that the cell line MCF10A eluted significantly more material when incubated with SL1 and SL2, than with SL4 and SL5 as before. The cell line SKBR3 eluted the most material from SL1, less with SL2 and SL5, and hardly any material from SL4. MDA-MB-231 showed a similar pattern to SKBR3 as before with SL1 having the most intense peaks, but had much more intense and consistent signals for the other three SLs. The only inconsistency between the overall patterns in the high

mannose and the bi-antennary tables was with the cell line T47D, which again had SL2 with the highest peak intensities, followed by SL5, but SL1 and SL4 reversed in order of intensity from what they were in the high mannose results. However, this may just be a result of the bi-antennary glycans being much more polar than high-mannose glycans because they can contain GlcNAcs and sialic acids, which allows them to more tightly bind to SL4 than SL1.

Again, all of the cell lines demonstrated a preference for fucosylated bi-antennary glycans over their nonfucosylated versions, however, and was more pronounced with MDA-MB-231 and SKBR3 than with MCF10A, and most pronounced with T47D. This is shown by the darker shading of green color in the bottom right corner of Table 4.7.

Table 4.7 Bi-antennary N-linked glycans of the four breast cell extracts eluted from SL1, 2, 4, and 5, raw data, sorted by cell line.

Bi-Antennary	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5
	MCF10A	MCF10A	MCF10A	MCF10A	SKBR3	SKBR3	SKBR3	SKBR3	MDA-MB-231	MDA-MB-231	MDA-MB-231	MDA-MB-231	T47D	T47D	T47D	T47D
A2																
A2B																
A2BG1																
A2BG1S1																
A2BG2A2																
A2BG2S1																
A2BG2S2																
A2G2A2G2A1																
A2G1																
A2G1S1																
A2G2																
A2G2A2																
A2G2S1																
A2G2S2																
FA2																
FA2B																
FA2BG1																
FA2BG1S1																
FA2BG2																
FA2BG2S1																
FA2BG2S2																
FA2G1																
FA2G1S1																
FA2G2																
FA2G2S1																
FA2G2S2																
FFA2																
FFA2B																
FFA2BG1S1																
FFA2G1S1																
FFFA2BG1S1																

As discussed earlier, the ratio of A2G2S2 compared to A2G2 was greater than 1.0 for each sample except MCF10A SL4, MCF10A SL5, and SKBR3 SL5, and was always greater than 1.0 when comparing the ratio of FA2G2:A2G2. This demonstrates that SL4 and SL5 are more polar SLs, compared to SL1 and SL2, and therefore bind more tightly to glycans containing sialic acids, making them more difficult to elute from the beads. This was again most evident in the cell lines MCF10A, which had A2G2S2:A2G2 ratios less than one for SL4 and SL5, and SKBR3, which has a A2G2S2:A2G2 ratio lower than one for SL5.

TRI-ANTENNARY N-LINKED GLYCANS FROM CELL LINE EXTRACTS

The next subgroup of N-linked glycans analyzed was the tri-antennary group. These glycans also contains the common core of two GlcNAcs and 3 mannoses, but three more GlcNAcs are branched from the mannoses to make the simplest tri-antennary N-linked glycan, called A3. The rest of the binding is the same as the bi-antennary glycans, where galactoses can be attached to the terminal GlcNAcs, sialic acids or more GlcNAcs can be attached to those galactoses, and fucoses can be attached to either the core GlcNAc and/or the antennary GlcNAcs. The overall pattern of tri-antennary N-linked glycans from the four breast cell line extracts, shown below in Table 4.8, conveys a similar relationship as seen when examining the bi-antennary glycans. Namely, all four cell lines have a slight preference for fucosylated tri-antennary glycans over the nonfucosylated, and that MCF10A appeared to have a wider variety of more evenly distributed glycans than the MDA-MB-231 and T47D extracts. SKBR3 had a similar

variety of tri-antennary glycans to MCF10A, but had more variability in the signal intensities across the table.

Table 4.8 Tri-Antennary N-linked glycans from four breast cell line extracts. Raw data (left), data normalized internally (right).

Tri-Antennary	MCF10A	SKBR3	MDA-MB-231	T47D	MCF10A	SKBR3	MDA-MB-231	T47D
A3					0.0	0.5	0.8	0.4
A3G1					1.0	0.5	0.8	0.4
A3G2A1					0.2	0.0	0.0	0.4
A3G2A2					0.0	0.0	0.0	0.4
A3G2S1					0.5	0.5	0.0	0.4
A3G1A1					0.5	0.0	0.5	0.0
A3G1S1					0.8	0.7	0.8	0.6
A3G2S2					0.0	0.8	0.0	0.0
A3G3					0.8	0.0	0.5	0.0
A3G3A1					0.0	0.3	0.5	0.4
A3G3A1S1					0.0	0.3	0.0	0.0
A3G3A2					0.8	0.5	0.5	0.2
A3G3A3					0.5	0.0	0.0	0.0
A3G3A3G2					0.3	0.0	0.0	0.0
A3G3S1					0.7	0.5	0.0	0.0
A3G3S2					0.7	0.5	0.0	0.4
A3G3S3					0.8	0.7	0.5	0.6
FA3G1					1.0	0.8	1.0	0.8
FA3G1S1					1.0	0.5	0.8	0.2
FA3G1A1					0.0	0.0	0.5	0.0
FA3G2					1.0	0.8	0.8	0.6
FA3G2S1					0.7	0.5	0.8	0.6
FA3G2S2					0.5	0.3	0.0	0.0
FA3					1.0	0.8	0.8	0.6
FA3G3					0.5	0.5	0.8	0.4
FA3G3A1G1					1.0	1.0	1.0	1.0
FA3G3A1G1S2					0.5	0.3	0.5	0.0
FA3G3A2G2S1					0.5	0.3	0.0	0.0
FA3G3A3G1					0.0	0.3	0.0	0.0
FA3G3A3G2					0.0	0.0	0.0	0.0
FA3G3S1					1.0	0.5	0.8	0.0
FA3G3S2					0.5	0.3	0.8	0.0
FA3G3S3					1.0	0.5	0.0	0.0
FA3G2A1					0.7	0.5	0.8	0.6
FA3G2A2					0.7	0.5	0.0	0.0
FA3G3A1					0.0	0.0	0.0	0.0
FA3G3A2					0.5	0.3	0.0	0.4
FA3G3A3					0.0	0.0	0.0	0.0
FA3G3A2S1					0.5	0.0	0.0	0.0
FA3G3A1S1					0.5	0.5	0.0	0.0
FFA3					0.0	0.0	0.5	0.4
FFA3G1					0.7	0.7	0.0	0.0
FFA3G1S1					1.0	0.5	0.8	0.8
FFFA3G1S1					0.0	0.0	0.0	0.4

TRI-ANTENNARY N-LINKED GLYCANS ELUTED FROM SL1, 2, 4, AND 5

Analyzing the tri-antennary glycans eluted after incubating the cell line extracts with the four SLs, some of the patterns previously seen in the other tables became more pronounced, and new patterns emerged as well. Again, by sorting Table 4.9 below by the cell lines instead of the SLs, some of these disparities were better highlighted. The most apparent of these, again, was that MCF10A eluted much more material when incubated with SL1 and SL2, than with SL4 and SL5, and SKBR3 eluted significantly less material from SL2, SL4, and SL5 than it did from SL1. The SKBR3 extract, when eluted from the SLs, showed the least amount of tri-antennary glycans with the lowest peak intensities. All cell lines, again, preferred fucosylated tri-antennary glycans over nonfucosylated tri-antennary glycans. However, this enrichment of fucosylated tri-antennary glycans was particularly apparent for the MCF10A extract when eluted from SL1 and SL2, when comparing these data sets to amount of nonfucosylated tri-antennary glycans present in the MCF10A cell line extract.

In addition to these previously noted observations, MDA-MB-231 eluted with SL1 and T47D eluted with SL2 both had the largest variety of tri-antennary glycans, particularly with nonfucosylated tri-antennary glycans. Again, MDA-MB-231 had very comparable results when eluted from SL2, SL4, and SL5, but the amount of tri-antennary glycans, when eluted from SL4, far surpassed any of the other cell line extracts eluted with SL4. When comparing the MDA-MB-231 eluted with SL1 data set to the MDA-MB-231 eluted with SL4 data set, many of the nonfucosylated and nonsialylated tri-antennary glycans disappear in the SL4 data set, but most of the fucosylated and sialylated tri-antennary glycans remained. This again suggests that MDA-MB-231 has a

more polar secreted extract than the other three breast cell line extracts, and that SL4, the most polar SL, was a good candidate in enriching these tri-antennary sialylated glycans. SL4 had the opposite effect with the other three cell line extracts, where only a few sialylated glycans were detected in these three samples.

The T47D extract was similar to the MCF10A extract when it came to SL1 and SL2, but differed in SL5, which had significantly more tri-antennary glycans with higher peak intensities. However unlike MDA-MB-231, the T47D extract eluted with SL5 did not also enrich sialylated tri-antennary glycans. In fact, the most sialylated glycans from T47D were eluted from SL2, and then closely followed by SL1. This suggests that the T47D extract, much like the MCF10A and SKBR3 extracts, has a relatively smaller amount of sialylated glycans that bind tighter to the more polar SLs, SL4 and SL5, and are not easily to elute from the beads.

This is again thought to be the binding relationship with SL4 because this SL has been experimentally the brightest SL in the array, as shown in Chapter 2, and these sialylated glycans have been detected in the cell line extracts and when incubated with other SLs. Conversely, the overall intensity of the tri-antennary glycans for all four cell lines was lower than what was observed when analyzing the bi-antennary glycans. This could suggest that the SLs prefer to bind to smaller glycans in general, but could also mean that the breast tissues prefers to express bi-antennary glycans over tri-antennary glycans. Or, on the other hand, more branched glycans could allow for more surface interaction between the SLs, and could lead to tighter binding than smaller glycans.

Table 4.9 Tri-antennary N-linked glycans from four breast cell lines eluted from SLs 1, 2, 4, and 5, organized by cell line, fucosylated glycans (bottom).

Tri-Antennary	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5
	MCF10A	MCF10A	MCF10A	MCF10A	SKBR3	SKBR3	SKBR3	SKBR3	MDA-MB-231	MDA-MB-231	MDA-MB-231	MDA-MB-231	T47D	T47D	T47D	T47D
A3																
A3G1																
A3G2A1																
A3G2A2																
A3G2S1																
A3G1A1																
A3G1S1																
A3G2S2																
A3G3																
A3G3A1																
A3G3A1S1																
A3G3A2																
A3G3A3																
A3G3A3G2																
A3G3S1																
A3G3S2																
A3G3S3																
FA3																
FA3G1																
FA3G1S1																
FA3G1A1																
FA3G2																
FA3G2S1																
FA3G2S2																
FA3G3																
FA3G3A1G1																
FA3G3A1G1S2																
FA3G3A2G2S1																
FA3G3A3G1																
FA3G3A3G2																
FA3G3S1																
FA3G3S2																
FA3G3S3																
FA3G2A1																
FA3G2A2																
FA3G3A1																
FA3G3A2																
FA3G3A3																
FA3G3A2S1																
FA3G3A1S1																
FFA3																
FFA3G1																
FFA3G1S1																
FFFA3G1S1																

TETRA-ANTENNARY N-LINKED GLYCANS FROM CELL LINE EXTRACTS

The second to last subgroup of N-linked glycans analyzed was the tetra-antennary group. These glycans, similar to the bi- and tri-antennary, has the common core of two GlcNAcs and 3 mannoses, but then has four more GlcNAcs, two branched from each of the two mannoses extending from the base mannose. This makes the simplest tetra-antennary N-linked glycan, called A4. From here, all the same rules for the bi- and tri-

antennary glycans are the same for the tetra-antennary glycans. More permutations of these glycans were detected between the four cell lines. Therefore, Tables 4.10 and 4.11 look more sparsely populated than the other tables previously analyzed.

Similar to the cell line extracts tables examined before, Tables 4.10 and 4.11 below show that the MCF10A extract had a larger variety of tetra-antennary glycans than the other three cell line extracts. All cell line extracts still presented a slight shift towards the fucosylated glycans, but was as apparent as the tri-antennary glycan table analyzed in the previous section. In, fact many of the tetra-antennary glycans detected in each cell line extracts were detected in both their fucosylated and nonfucosylated forms. In addition, the MCF10A extract displayed several multi-fucosylated glycans that were not seen in the other cell line extracts, shown at the bottom of Table 4.11.

Table 4.10 Nonfucosylated tetra-antennary glycans from four breast cell line secreted extracts, raw data (left), internally normalized data (right).

Tetra-Antennary	MCF10A	SKBR3	MDA-MB-231	T47D	MCF10A	SKBR3	MDA-MB-231	T47D
A4G4S2					0.3	0.0	0.0	0.0
A4G3A1					0.8	0.0	0.0	0.0
A4G3A1G1					0.0	0.0	0.0	0.0
A4G3A1GIS1					0.5	0.0	0.0	0.0
A4G3A2					0.5	0.0	0.0	0.0
A4G3A2G2					0.3	0.0	0.0	0.0
A4					1.0	0.0	0.7	0.6
A4G1S1					0.7	0.3	0.0	0.4
A4G1A1					0.0	0.0	0.0	0.0
A4G4A1					0.0	0.3	0.0	0.0
A4G1					0.0	0.0	0.3	0.0
A4G2					0.2	0.0	0.0	0.4
A4G3					0.0	0.3	0.3	0.4
A4G3S1					0.0	0.3	0.0	0.0
A4G3A1					0.0	0.5	0.3	0.2
A4G2S1					0.5	0.0	0.3	0.0
A4G2S2					0.5	0.3	0.0	0.0
A4G2A1					0.0	0.0	0.0	0.4
A4G2A2					0.0	0.3	0.0	0.4
A4G2A2G2					0.0	0.0	0.0	0.0
A4G3A1GIS2					0.5	0.0	0.0	0.0
A4G3A1GIS3					0.0	0.0	0.0	0.0
A4G3A1S2					0.3	0.0	0.0	0.0
A4G4A3G2S1					0.0	0.0	0.0	0.0
A4G3A2G1S1					0.3	0.3	0.0	0.0
A4G3A2					0.0	0.0	0.0	0.0
A4G3A2G2					0.0	0.0	0.0	0.0
A4G3A3					0.5	0.5	0.0	0.2
A4G3A3G2					0.3	0.0	0.0	0.0
A4G3A3G3					0.2	0.0	0.0	0.0
A4G4A3G1S1					0.0	0.0	0.0	0.0
A4G4A3G3					0.0	0.0	0.0	0.0
A4G4A4					0.5	0.3	0.0	0.0
A4G4A4G1					0.0	0.0	0.0	0.0
A4G4S1					0.5	0.5	0.0	0.0
A4G4S3					0.3	0.0	0.0	0.0
A4G4S4					0.3	0.0	0.0	0.0

Table 4.11 Fucosylated tetra-antennary N-linked glycans from four breast cell line extracts, raw data (left), and internally normalized data (right).

Tetra-Antennary	Raw Data				Internally Normalized Data			
	MCF10A	SKBR3	MDA-MB-231	T47D	MCF10A	SKBR3	MDA-MB-231	T47D
FA4G4					1.0	1.0	0.7	1.0
FA4G4S2					0.5	0.3	0.3	0.0
FA4G4A1G1S1					0.5	0.3	0.0	0.0
FA4					1.0	0.5	0.8	0.0
FA4G1A1					1.0	1.0	1.0	0.6
FA4G1S1					0.7	0.5	0.0	0.4
FA4G1					0.0	0.0	0.3	0.0
FA4G2					0.7	0.5	0.5	0.6
FA4G3					0.0	0.0	0.0	0.0
FA4G2A1					0.7	0.5	0.0	0.2
FA4G2A2					0.0	0.0	0.0	0.0
FA4G2A2G2					0.0	0.3	0.0	0.0
FA4G2S1					0.5	0.5	0.5	0.6
FA4G2S2					0.5	0.2	0.0	0.0
FA4G2A1G1S3					0.0	0.3	0.0	0.0
FA4G3A1					0.5	0.3	0.0	0.4
FA4G3A1G1S1					0.0	0.0	0.0	0.0
FA4G3A1G1S2					0.0	0.0	0.0	0.0
FA4G3A1G1S3					0.0	0.0	0.0	0.0
FA4G3A1S1					0.5	0.0	0.0	0.0
FA4G3A1S2					0.3	0.0	0.0	0.0
FA4G3A2G2					0.0	0.0	0.0	0.0
FA4G3A3G1					0.0	0.0	0.0	0.0
FA4G3A3G2					0.0	0.0	0.0	0.0
FA4G3A3G3					0.0	0.0	0.0	0.0
FA4G3S1					0.5	0.5	0.0	0.0
FA4G3S3					0.0	0.3	0.0	0.0
FA4G3A3					0.0	0.0	0.0	0.0
FA4G4A2G2S1					0.3	0.0	0.0	0.0
FA4G4A2G2S2					0.5	0.0	0.0	0.0
FA4G4A3G1S1					0.0	0.0	0.0	0.0
FA4G4A3G2S1					0.0	0.0	0.0	0.0
FA4G4A3G3					0.3	0.0	0.0	0.0
FA4G4A3G3S1					0.5	0.0	0.0	0.0
FA4G4A4G3					0.5	0.0	0.0	0.0
FA4G4S1					0.7	0.5	0.5	0.6
FA4G4S2					0.0	0.0	0.0	0.0
FA4G4S3					0.3	0.5	0.0	0.0
FA4G4A1S3					0.0	0.0	0.0	0.0
FA4G4S4					0.7	0.5	0.0	0.0
FA4G4A4					0.0	0.0	0.0	0.0
FFA4					0.7	0.7	0.5	0.0
FFA4G1S1					0.7	0.3	0.0	0.0
FFA4G1A1					0.0	0.0	0.0	0.0
FFA4G2S2					0.5	0.0	0.0	0.0
FFA4G4A3G1S1					0.0	0.0	0.0	0.0
FFA4G4A3G3S1					0.5	0.0	0.0	0.0
FFFA4					1.0	0.5	0.5	0.6
FFFA4G1					0.5	0.0	0.0	0.0

TETRA-ANTENNARY N-LINKED GLYCANS ELUTED FROM SL1, 2, 4, AND 5

When analyzing the tetra-antennary glycans, in Figure 4.12 below, eluted from the cell line extracts after incubating them with the four SLs, the greatest variability between the four cell lines was seen compared to the other N-linked glycan subgroups. Although the disparities were more pronounced, the overall patterns were very similar to what was seen when analyzing the tri-antennary glycans eluted from the four SLs, but

were almost the opposite of what was seen when analyzing the tetra-antennary glycans in cell line extracts in Tables 4.10 and 4.11 above. MCF10A, once again, showed the most types of tetra-antennary glycans in the cell line extract, however, once the MCF10A extract was eluted from the four SLs, it detected the least amount of tetra-antennary glycans with the lowest peak intensities. For MCF10A eluted with SL4 and SL5, only two and three tetra-antennary glycans were detected, respectively. This was surprising after seeing a plethora of tetra-antennary glycans in the MCF10A cell line extract. This was particularly true with the abundance of nonfucosylated tetra-antennary glycans detected in all of the cancerous cell line extracts compared to MCF10A, which was also seen during the tri-antennary glycan analysis.

The SL1 elutions with SKBR3 and MDA-MB-231, and the SL2 elution with T47D all showed a larger variety of tetra-antennary glycans than any of the MCF10A samples. For SKBR3, this only true when eluted from SL1, whereas when eluted with SL2, SL4, and SL5, hardly eluted any tetra-antennary glycans. In fact, SL2 only eluted one sialylated tetra-antennary glycan, SL4 eluted two, and SL5 eluted none.

The MDA-MB-231 extract eluted with SL1 and the T47D extract eluted with SL2 again rivaled for the most amount of tetra-antennary glycans with the greatest peak intensities, as seen in the tri-antennary glycan table. The MDA-MB-231 extract eluted with SL4, again distinguished itself from the other three cell line extracts by containing the most of tetra-antennary glycans, the majority of which were sialylated and fucosylated.

The T47D extract also showed the same pattern as before, with the SL2 elution containing the most amount of glycans, followed by SL2 and SL5, and then a sharp

decline in glycans when eluted from SL4. The T47D SL1 elution contained a few more sialylated tetra-antennary glycans than the SL5 elution, which hardly contained any. This demonstrates again that the T47D extract only elutes polar glycans from SL1 and SL2, which again suggests that sialylated glycans may not be in large enough abundance to be eluted from SL4 and SL5.

Table 4.12 Tetra-Antennary N-linked glycans from four breast cell lines eluted from SLs 1, 2, 4, and 5, nonfucosylated glycans (top), and fucosylated glycans (bottom).

Tetra-Antennary	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5
	MCF10A	MCF10A	MCF10A	MCF10A	SKBR3	SKBR3	SKBR3	SKBR3	MDA-MB-231	MDA-MB-231	MDA-MB-231	MDA-MB-231	T47D	T47D	T47D	T47D
AG4S2																
AG3A1																
AG3A1G1																
AG3A1G1S1																
AG3A2																
AG3A2G2																
A4																
AG1S1																
AG1A1																
AG4A1																
AG1																
AG2																
AG3																
AG3S1																
AG3A1																
AG2S1																
AG2S2																
AG2A1																
AG2A2																
AG2A2G2																
AG3A1G1S2																
AG3A1G1S3																
AG3A1S2																
AG4A3G2S1																
AG3A2G1S1																
AG3A2																
AG3A2G2																
AG3A3																
AG3A3G2																
AG3A3G3																
AG4A3G1S1																
AG4A3G3																
AG4A4																
AG4A4G1																
AG4S1																
AG4S3																
AG4S4																
FA4G4																
FA4G4S2																
FA4G4A1G1S1																
FA4																
FA4G1A1																
FA4G1S1																
FA4G1																
FA4G2																
FA4G3																
FA4G2A1																
FA4G2A2																
FA4G2A2G2																
FA4G2S1																
FA4G2S2																
FA4G2A1G1S3																
FA4G3A1																
FA4G3A1G1S1																
FA4G3A1G1S2																
FA4G3A1G1S3																
FA4G3A1S1																
FA4G3A1S2																
FA4G3A2G2																
FA4G3A3G1																
FA4G3A3G2																
FA4G3A3G3																
FA4G3S1																
FA4G3S3																
FA4G3A3																
FA4G4A2G2S1																
FA4G4A2G2S2																
FA4G4A3G1S1																
FA4G4A3G2S1																
FA4G4A3G3																
FA4G4A3G3S1																
FA4G4A4G3																
FA4G4S1																
FA4G4S2																
FA4G4S3																
FA4G4A1S3																
FA4G4S4																
FA4G4A4																
FFA4																
FFA4G1S1																
FFA4G1A1																
FFA4G2S2																
FFA4G4A3G1S1																
FFA4G4A3G3S1																
FFA4																
FFFA4G1																

HYBRID, POLYLACNAC, AND MULTIFUCOSYLATED N-LINKED GLYCANS FROM CELL LINE EXTRACTS

The last subgroup of N-linked glycans that was analyzed was the hybrid group, which was also combined with another group referred to as the PolyLacNAc group. The hybrid group is a combination of any high mannose glycan with any of the antennary glycan moieties seen before in the bi- tri- and tetra- subgroups. These glycans still contain the common core of two GlcNAcs and 3 mannoses, but after the core additional mannoses, GlcNAcs, galactoses, fucoses and sialic acids can be attached in many different ways. This group of glycans was the most challenging to identify from the mass peaks because mannose and galactose are isomers of each other.

The second type of glycan in this group is the polyLacNAc group, which contains repeat units of the disaccharide LacNAc (N-Acetylglucosamine bound to galactose) more than once in a branch. This second group, which was partially included in the tetra-antennary group, was also challenging to identify glycan peaks because, without any MS-MS collision data, it was impossible to determine whether certain glycans had terminal galactoses, or extra mannoses in the hybrid structure. Therefore, the structures were not speculated on, and only the masses and likely saccharide content were reported. As a general note, the names of the glycans in Table 4.13 below are not abbreviated as they were before, like A2, A3 and A4. For clarification purposes, these glycans are named exactly for what they contain. In example, M5A3 contains exactly five mannoses and three GlcNAcs. This glycan presumably contains the standard core of two GlcNAcs and three mannoses, and therefore has antennary branches containing two more mannoses and one more GlcNAc.

Table 4.13 also contains multi-fucosylated glycans, which were denoted in their fully fucosylated form (Ex. FFM3A2, or FFFFM3A2) instead of considering whether two of those fucoses actually represented one sialic acid, which was mentioned before as another challenge in glycan identification. For this reason, all glycan structures were assumed based on their most logical and precedented configuration, but were also chosen to preserve symmetry within the glycan structure stemming from the amount of GlcNAcs calculated to be present in discerning the mass.

This latter consideration may be erroneous, but was a useful guideline in matching glycan structures to mass peaks. In one last example, if the mass peak was equivalent to FFFFM5A4, it would be assumed that the actual structure would be (A2M3)A2G2S2. (The (A2M3) represents the core N-linked structure conserved among all N-linked glycans) This is because four fucoses were initially calculated to be included in the glycan, however only four GlcNAcs were on the glycan, and having never seen any literary precedence of the second core GlcNAc being fucosylated, it seems more likely that at least two of the fucoses are actually one sialic acid, and if there is a sialic acid, then there must be at least one galactose to connect it to the structure, as sialic acid does not bind to any other monosaccharides. It could also be plausible that the other two fucoses are not another sialic acid, and represent one core and one antennary GlcNAc fucosylation, such as FF(A2M3)M1G1S1. But, as mentioned earlier, asymmetrical structures were avoided unless they were outweighed by some other logic.

When analyzing this last subgroup of glycans in the cell line extracts, a large variety of these structures were seen in both the MCF10A and SKBR3 extracts, alongside the same relatively weak signals seen in the MDA-MB-231 and T47D extracts. Many of

the structures identified, were not only fucosylated once, but were also seen fucosylated multiple times, as shown at the bottom of Table 4.13. This seemed to be more prevalent with the SKBR3 extract than the MCF10A extract. These multifucosylated structure are indicative of sialyl Lewis A (sLeA) and sialyl Lewis X (sLeX) structures, which are commonly observed cancer-associated glycans.

Table 4.23 Hybrid N-linked glycans from four breast cell line secreted extracts. Glycans names are not abbreviated, and are named for exactly what they

hybrids	MCF10A	SKBR3	MDA-MB-231	T47D	MCF10A	SKBR3	MDA-MB-231	T47D
M3A3					0.00	0.00	0.33	0.00
M3A4G3S2					1.00	0.50	0.00	0.25
M3A5G5S2					0.50	0.00	0.00	0.00
M3A6G5					1.00	1.00	0.17	0.50
M3A7G6					1.00	0.50	0.00	0.25
M4A3					0.67	0.50	0.33	0.50
M5A3					0.83	0.67	0.50	0.75
M5A3G1S1					0.00	0.17	1.00	1.00
M5A4G2S1					1.00	0.67	0.17	0.25
M5A4G2S2					0.50	0.17	0.00	0.00
M5A5G3					0.67	0.50	0.00	0.00
M6A3					1.00	0.83	0.50	0.75
M6A2G2S2					0.67	0.50	0.50	0.00
M6A3G1S1					0.50	0.33	0.50	0.25
M9A3G1S1					0.50	0.50	0.00	0.25

hybrids	MCF10A	SKBR3	MDA-MB-231	T47D	MCF10A	SKBR3	MDA-MB-231	T47D
FM3A2					0.50	0.00	0.17	0.25
FM3A3					0.83	0.83	0.00	0.75
FM3A4G4					0.00	0.00	0.50	0.75
FM3A4G4S1					1.00	0.50	0.00	0.00
FM3A5G4					0.67	0.00	0.00	0.00
FM3A6G5					0.50	0.17	0.00	0.00
FM3A7G6					0.50	0.17	0.00	0.00
FM3A8G7					0.50	0.00	0.00	0.00
FM3A10G3					0.50	0.00	0.00	0.00
FM4A3					1.00	0.83	0.50	0.75
FM5A3					0.67	0.50	0.33	0.00
FM6A3					0.67	0.50	0.00	0.00
FM6A5G1S1					0.50	0.17	0.00	0.00
FM7A3G1S1					1.00	0.67	0.00	0.00
FFM3A3					0.00	0.00	0.00	0.75
FFM3A4G3					1.00	1.00	0.50	0.75
FFM3A7G4S1					0.00	0.00	0.00	0.00
FFM4A3					1.00	0.83	0.50	0.75
FFM5A3					1.00	0.83	0.50	0.75
FFM5A5G3S2					0.00	0.00	0.00	0.00
FFM6A3					1.00	0.83	0.33	0.75
FFM7A2					0.00	0.50	0.50	0.00
FFM7A3G1S1					1.00	0.50	0.00	0.00
FFM8A3					1.00	0.00	0.17	0.00
FFM9A3					1.00	0.50	0.00	0.00
FFF3A4G3					1.00	0.67	0.17	0.50
FFF3A5G1					1.00	0.50	0.50	0.25
FFF3A5G2					0.67	0.50	0.00	0.00
FFF3A7G5					0.83	0.50	0.00	0.00
FFF4A3					0.00	0.83	0.00	0.00
FFF5A3					0.00	0.67	0.00	0.00
FFF5A5G3					0.00	0.00	0.00	0.00
FFF6A3					0.00	0.50	0.00	0.00
FFF7A3G1S1					1.00	0.50	0.00	0.00
FFFF3A4G2S2					1.00	0.67	0.50	0.50
FFFF5A3					0.00	0.67	0.00	0.00
FFFF6A3					0.00	0.50	0.00	0.00

HYBRID, POLYLACNAC, AND MULTIFUCOSYLATED N-LINKED GLYCANS ELUTED FROM SL1, 2, 4, AND 5

When analyzing this final subgroup of glycans eluted from the cell line extracts after incubating them with the four SLs, the data table generated most resembles what was seen in the tri-antennary glycans, which stands to reason because most of the hybrid structures are either di- or tri- antennary glycans. However, an enrichment in fucosylated glycans was not seen in this subgroup as it was with the other N-linked glycan subgroups. As noted in the cell line extract table above, many glycans were detected in their nonfucosylated and fucosylated forms.

The MCF10A extract again displayed a stark contrast in the amount of glycans and peak intensities detected from SL1 and SL2, which were almost identical, versus SL4 and SL5, which were very similar as well. Many of the multifucosylated glycans at the bottom of Table 4.14 disappeared when eluted with SL4 and SL5, which suggests that many of these structures listed as containing two fucoses actually contain one sialic acid.

As per usual, SKBR3 also had the greatest signals when eluted with SL1, a much reduced signal when eluted with SL2 and SL4, and almost no signal at all when eluted from SL5. SKBR3 eluted with SL5 only detected one glycan in this subgroup. The MDA-MB-231 and T47D extracts also demonstrated their consistency with this glycan subgroup. MDA-MB-231 again had the most intense signals when eluted with SL1, and equally comparable results from SL2, SL4 and SL5. MDA-MB-231 being the only cell line extract to elute a significant amount of material with SL4. The T47D extract again had the most intense signals from SL2, SL1 and SL5 to a slightly lesser degree, and a weak signal when eluted from SL4.

Table 3.14 Hybrid N-linked glycans eluted from four breast cell line extracts using SL1, 2, 4, and 5, nonfucosylated (top) and fucosylated (bottom).

hybrids	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5
	MCF10A	MCF10A	MCF10A	MCF10A	SKBR3	SKBR3	SKBR3	SKBR3	MDA-MB-231	MDA-MB-231	MDA-MB-231	MDA-MB-231	T47D	T47D	T47D	T47D
M3A3																
M3A4G3S2																
M3A5G5S2																
M3A6G5																
M3A7G6																
M4A3																
M5A3																
M5A3G1S1																
M5A4G2S1																
M5A4G2S2																
M5A5G3																
M6A3																
M6A2G2S2																
M6A3G1S1																
M9A3G1S1																
hybrids	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5	SL1	SL2	SL4	SL5
	MCF10A	MCF10A	MCF10A	MCF10A	SKBR3	SKBR3	SKBR3	SKBR3	MDA-MB-231	MDA-MB-231	MDA-MB-231	MDA-MB-231	T47D	T47D	T47D	T47D
FM3A2																
FM3A3																
FM3A4G4																
FM3A4G4S1																
FM3A5G4																
FM3A6G5																
FM3A7G6																
FM3A8G7																
FM3A10G3																
FM4A3																
FM5A3																
FM6A3																
FM6A5G1S1																
FM7A3G1S1																
FFM3A3																
FFM3A4G3																
FFM3A7G4S1																
FFM4A3																
FFM5A3																
FFM5A5G3S2																
FFM6A3																
FFM7A2																
FFM7A3G1S1																
FFM8A3																
FFM9A3																
FFF3A4G3																
FFF3A5G1																
FFF3A5G2																
FFF3A7G5																
FFF4A3																
FFF5A3																
FFF5A5G3																
FFF6A3																
FFF7A3G1S1																
FFFF3A4G2S2																
FFFF5A3																
FFFF6A3																

N-LINKED GLYCANS RELEASED FROM NATURAL LECTINS

In addition to the cell line extracts, the N-linked glycans from the natural lectins (NLs) used at the end of Chapter 2 were released and processed through the Waters N-linked glycan kit. The three NLs, GSLII, Jacalin, and SNA, are again all derived from

plants, which are known to use different carbohydrates than human cells do, such as the monosaccharide xylose. For this reason, several of the glycan peaks observed during these runs remained unidentified. However, the majority of the glycans present in GSLII and SNA were identified as high mannose glycans, with a few bi-antennary glycans observed as well. The common glycans between these two NLs included: M4, M5, M6, M7, M8, M9, and A2G1S1.

The sample from Jacalin contained some of these glycans as well, but did not produce as clean of a chromatogram as did the other two NLs. This was apparent by the wide glycan peaks that eluted off of the column over a relatively long time. This may possibly be an indication as to why Jacalin was very distinguished from the other two NLs in the experiments at the end of Chapter 2, but the explanation remains unclear.

Analyzing the N-linked glycans present on the NLs gave a good explanation as to why the SLs had a strong affinity to them, because the SLs also demonstrated a strong affinity for high mannose glycans with the cell line extracts as well. Additionally, since the vast majority of the observed N-linked glycans were of one relatively nonpolar glycan subgroup, it further stands to reason that the NLs were able to discriminate, at least partially, the SLs by their relative polarity.

CONCLUSIONS OF THE BREAST CELL LINE N-LINKED GLYCAN STUDIES

After analyzing the N-linked glycans from these four cell lines extracts, and the N-linked glycans eluted after incubating the extracts with the four SLs, it became clear that the hydrophobicity of the SLs plays a major role in the array's ability to discriminate between the cell lines. The overall average peak intensities of the SLs, regardless of the

cell line extracts, were more similar between SL1 and SL2, whereas SL4 and SL5 were more similar to each other. The former two SLs had greater intensities than the latter two. The similarities seen between SL1 and SL2, and between SL4 and SL5, were also observed during the SL-NL flow cytometry experiment discussed at the end of Chapter 2. These two pairs of SLs contributed to the most amount of misclassification, when being discriminated by the three Natural Lectins (NLs), when then overlapped more significantly after the NLs were de-N-linked-glycosylated with PNGase F.

When closely examining the difference in the amino acid sequences between the SLs, the general pattern of glycan intensities was $SL1 > SL2 > SL5 > SL4$. This was the case for every cell line extract, with the exception of T47D, which showed slightly greater intensities when eluted with SL2 over SL1.

The healthy control, MCF10A, appeared to have the most uniform and even distribution of N-linked glycans across all of the subtypes discussed, however it only elute a significant amount of glycans with SL1 and SL2. This trend became more apparent as the branching increased and the glycans became larger. This suggests that the amount of sialylated material in the MCF10A extract is relatively low, and binds tightly to SL4 and SL5, but is more easily eluted from SL1 and SL2. The argument for this case is made when examining the earlier flow cytometry experiments where SL4 and SL5 generally had greater fluorescent intensities than SL1 and SL2.

When looking back at the earlier flow cytometry and NL-SL sandwich assay experiments, T47D was the cell line extract that was most misclassified with the MCF10A extract. When analyzing the T47D extract in relation to the four SLs, the patterns were very similar to MCF10A, in that they both eluted significant amounts of

material from SL1 and SL2, and much less from SL4. The two cell line extracts only differed in SL5, which T47D also eluted a significant amount of material from.

In comparing the two lowly metastatic breast cell line extracts, SKBR3 and T47D, to each other, they both had very dissimilar relationships with the four SLs. The most obvious of which is their difference between them when eluted with SL2 and SL5. SKBR3 was the worst performing sample, in terms of number of peaks and signal intensities, for every SL except for SL1. This made the SKBR3 extract resemble the MDA-MB-231 extract because, although MDA-MB-231 generally had the most intense glycan signals for all SLs, they both clearly eluted the most material from SL1, and had lower relatively consistent signals with SL2, SL4, and SL5. When looking back at earlier experiments, the SKBR3 and MDA-MB-231 extracts had consistent overlap between them. SKBR3 also had overlap with MCF10A, which may be attributed to their common preference for SL1 and SL2 with smaller more hydrophobic glycans. In contrast, the MCF10A and MDA-MB-231 extracts had the lowest amount of overlap between them of all the cell lines.

The cell line elutions from SL4, the most polar of the SLs, proved to be useful in characterizing the polarity of the cell line extracts. SL4 clearly demonstrated that the MDA-MB-231 extract has a large relative amount of sialylated material, which is consistent with previously published literature.³³ For more hydrophobic cell line extracts, SL4 would not be useful in extracting sialylated or more polar glycans, and therefore can be used as a marker of the relative abundance of those types of glycans.

When assessing SL4 from earlier experiments, it was generally the least useful, or had the lowest F-to-remove score, in discriminating the cell line extracts using LDA. In this context, it seems that SL4 bound the most amount of material from each cell line,

because it was always the SL with the brightest fluorescence. In thinking of SL4 as an SPE agent, it would only be useful if the cell line extract contained relatively large amount of polar glycans, so that at least a portion of them could be eluted from the beads. Although difficult to say with the knowledge at hand, it may be that a stronger elution buffer would be needed to elute the more polar glycans off of a more polar SL. This consideration will be revisited again in the future works section when discussing the use of TentaGel beads as SPE agents.

This analysis of SL4 was almost the complete opposite case for SL1, the most hydrophobic of the SLs. SL1 eluted the most amount of material for every cell line, except for T47D, which again consistently eluted the most material with SL2. This relative ease of eluting significant amounts of material from this SL reaffirms the claim made above about the strength needed in the elution buffer. The elutions from SL1 generally had a wide variety of glycans across the board, and did not appear to enrich any subtype in particular. This amount of hydrophobicity may allow SL1 to be overly cross-reactive with all N-linked glycans. However, all of the SLs did seem to have a subtle enrichment of fucosylated glycans over their nonfucosylated counterparts, because more variety of fucosylated glycans was generally observed in the samples. This effect was subtle though because all the cell line extracts by themselves appeared to have a slight preference for fucosylated glycans.

When analyzing the tri- and tetra-antennary glycans, the three cancerous cell line extracts eluted with the four SLs showed a significant increasing in these branched structures over MCF10A. This was particularly true with nonfucosylated tri- and tetra-antennary glycans, which were shown in greater abundance in the three cancer extracts

compared to the MCF10A extract. This is consistent with what has been seen in the literature regarding aberrant glycosylation and cancer, which has reaffirmed in many cancers an upregulation of the branching enzymes, GnT-V in particular, during tumorigenesis.²¹

MDA-MB-231 was shown in these experiments to have a larger relative amount of sialylated N-linked glycans, apparent by the amount of sialylated glycans eluted from SL4 relative to the other cell line extracts. Since the glycans being eluted from the beads are the same glycans, then they should elute the same way, and only their abundance and the presence of other more tightly binding glycans should have an impact on them. MDA-MB-231 is a TNBC, which is characterized by its lack of upregulated hormone receptors, but also by its aggressive behavior and high metastatic potential. This finding is also consistent with what has been reported in the literature, which is a correlation between an increase in sialylation and a decrease in integrin-mediated cellular adhesion³⁰.

This notion follows through when analyzing SKBR3 and T47D in comparison to MCF10A. SKBR3 and T47D are HER2+ and Luminal A breast cancer subtypes, respectively. The HER2+ enriched subtype is known as another fairly aggressive type of breast cancer, while the Luminal A subtype is known to be a slower growing and generally has a much more favorable prognosis. This information fits well with the patterns observed in the N-linked glycan analysis, which is that SKBR3 most closely resembles MDA-MB-231 and T47D most resembles the healthy MCF10A cell line. When comparing this to previously published findings, Lee, et al.³⁴ described the SKBR3 cell line fit in between the MDA-MB-231 cell line and the HMEC (Healthy control-Human Mammary Epithelial Cells) cell line, resembling both on either side.

CONCLUDING REMARKS

The information gathered from the experiments throughout Chapter 2, 3, and 4 were all for the primary purpose of validating the SL array as a valuable sensor for aberrant glycosylation with respect to breast cancer. Once the flow cytometry platform was established and certain variables were in control, the SL array became reproducible and consistent with other findings in this important and growing field of cancer diagnostics. Furthermore, the ability of the individual SLs to enrich for specific glycans provides a strong foundation for the use of these SLs, not only as a valuable research tool, but also as a prognostic tool to characterize cancers into subtypes based on their glycan fingerprints.

The results discussed in this paper were also externally validated because they coalesced with other findings regarding the wealth of existing information on these cell lines and breast cancer associated aberrant glycosylation. This is valuable to the project because it allows for a better way to predict and create new and useful SLs rather than try and identify them based on random screening and Edman degradation.

In summary, the current SL array that was used in these experiments is fairly limited in its variation between SLs. Despite all of the different interactions that were observed between the SLs and the different cell line extracts, the SLs are all at least 57% similar in amino acid sequence and all have a +3 overall charge. More variations of SLs should be explored to exploit different properties for better discrimination ability and for more specific glycan enrichment as well. Examples of these changes will be further discussed in Chapter 5, the future works section.

CHAPTER 5

FUTURE WORK AND EXPERIMENTS

Throughout the past experiments described in Chapter 2, it was observed that SL3 and SL9 consistently had relatively low signals, for any cell line, which were barely above the blank SL negative controls. Furthermore, SL3 and SL9 rarely contributed to the discrimination between the cell lines (Low F-to-remove factor), and therefore added little or no benefit to the array. Although these SLs were not analyzed for their use as SPE agents, replacing these SLs should be considered, to add more diversified to the SL array. However, these SLs may be much more useful if they were synthesized on a different bead resin. Instead of just using PS or TentaGel beads in the array, how can they be combined to make a more diverse and powerful sensor array?

The original nine SLs, used in previous experiments with the fluorescent microscope and flow cytometer, were selected based on their cross-reactivity to different purified glycoproteins, two of which were bovine mucin and porcine mucin. Mucins are glycoproteins secreted by the submaxillary glands, and are classified by their extensive O-linked glycosylation, which aid in their role as a gastrointestinal lubricant. For this reason, it is quite plausible that the existing SL array used in all prior experiments, is more catered to detect glycoproteins with more extensive O-linked glycosylation than N-linked glycosylation. Classically, O-linked glycosylation is generally restricted to signaling within the cell, whereas N-linked glycosylation is generally restricted to the cell

surface and secretome. However, O-linked glycoproteins, including the MUC-1 glycoprotein, have been detected circulating in the blood stream of breast cancer patients. This is more evidence that tumorigenesis results in random cellular behavior, and therefore any diagnostic tool should not be limited to typical homeostatic behavior. In addition to O-linked glycosylation, the SL array could also be used to analyze glycolipids, if a glycolipid isolation protocol could be developed.

In addition to these ameliorations of the SL array, the SPE procedure from chapter 3 should be repeated with the SLs on TentaGel beads, to see if they could potentially be used as a better SPE material. Some potential drawbacks for this would be that the TentaGel beads have a 4x lower binding capacity than the PS beads and are much more hydrophilic, which may make eluting the glycan material off of the beads more challenging. This is again a concern because the elution buffer must be gentle enough so it does not hydrolyze any of the terminal sialic acids.

In speaking of the elution buffer, the efficiency of the elution buffer at stripping the N-linked glycans off of the beads was not thoroughly examined, and the amount of incidental sialic acid hydrolysis was unknown during the Chapter 4 experiment as well. A further investigation into the elution buffer should take place to be able to safely recover the greatest amount of glycoproteins from the SLs.

Finally, the experiments in Chapter 4 could be repeated with cell line extracts that were previously fluorescently labeled with FITC, to see if the FITC is playing a role in the binding interactions with the SLs.

With all these considerations in mind, the SL array has the great potential to be useful in our collective fight against cancer.

BIBLIOGRAPHY

1. Stringhini S. The Shift From Heart Disease to Cancer as the Leading Cause of Death in High-Income Countries: A Social Epidemiology Perspective. *Ann Intern Med.* 2018;169(12):877-878.
2. Bradley CJ, Given CW, Roberts C. Race, socioeconomic status, and breast cancer treatment and survival. *J Natl Cancer Inst.* 2002;94(7):490-6.
3. Plenderleith IH. Treating the treatment: toxicity of cancer chemotherapy. *Can Fam Physician.* 1990;36:1827-30.
4. National Cancer Institute. <http://www.cancer.gov/types/breast/mammograms-fact-sheet#q4>
5. Celià-Terrassa T, Kang Y. Distinctive properties of metastasis-initiating cells. *Genes Dev.* 2016;30(8):892-908.
6. Hong B, Zu Y. Detecting circulating tumor cells: current challenges and new trends. *Theranostics.* 2013;3(6):377-94.
7. Sathwara JA, et al. Sociodemographic Factors and Late-stage Diagnosis of Breast Cancer in India: A Hospital-based Study. *Indian J Med Paediatr Oncol.* 2017;38(3):277-281.
8. Malmberg KJ, et al. Natural killer cell-mediated immunosurveillance of human cancer. *Semin Immunol.* 2017;31:20-29.
9. Sihto H, Lundin J, Lundin M, et al. Breast cancer biological subtypes and protein expression predict for the preferential distant metastasis sites: a nationwide cohort study. *Breast Cancer Res.* 2011;13(5):R87.
10. Nwabo Kamdje AH, et al. Signaling pathways in breast cancer: therapeutic targeting of the microenvironment. *Cell Signal.* 2014;26(12):2843-56.
11. Reddy SM, Barcenas CH, Sinha AK, et al. Long-term survival outcomes of triple-receptor negative breast cancer survivors who are disease free at 5 years and relationship with low hormone receptor positivity. *Br J Cancer.* 2017;118(1):17-23.

12. Varki A. Biological roles of glycans. *Glycobiology*. 2016;27(1):3-49.
13. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer*. 2015;15(9):540-55.
14. Joenvaara S, et al. Quantitative N-glycoproteomics reveals altered glycosylation levels of various plasma proteins in bloodstream infected patients. *PLoS One*. 2018;13(3):e0195006.
15. Krištić J, et al. Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci*. 2014;69(7):779-89.
16. Alan Kirwan, et al. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *Biomed Res Int*. 2015;2015:490531.
17. Jing X, et al. Overexpression of MUC1 predicts poor prognosis in patients with breast cancer. *Oncol Rep*. 2018;41(2):801-810.
18. James TD, et al. Boronic acids in saccharide recognition. Royal Society of Chemistry; Cambridge, UK: 2006.
19. Bicker KL, et al. Boronic acid functionalized peptidyl synthetic lectins: combinatorial library design, peptide sequencing, and selective glycoprotein recognition. *ACS Comb Sci*. 2011;13(3):232-43.
20. Bicker KL, et al. Synthetic lectin arrays for the detection and discrimination of cancer associated glycans and cell lines. *Chem Sci*. 2012;3:1147.
21. Potapenko IO, et al. Glycan-related gene expression signatures in breast cancer subtypes; relation to survival. *Molecular oncology*. 2015;9:861-876.
22. Ortmann O, et al. Weak estrogenic activity of phenol red in the pituitary gonadotroph: re-evaluation of estrogen and antiestrogen effects. *J Steroid Biochem*. 1990;35(1):17.
23. <https://www.thermofisher.com/order/catalog/product/10010023>
24. Crowell AM, et al. Maximizing recovery of water-soluble proteins through acetone precipitation. *Anal Chim Acta*. 2013;796:48-54.
25. Houben W, et al. Synthesis of Peptides and Peptidomimetics. Stuttgart Literature. 2002;22:676.

26. Upton G, Cook I. Understanding Statistics. Oxford University Press. 1996;55.
27. McLachlan GJ. Discriminant Analysis and Statistical Pattern Recognition. Wiley Interscience. 2004.
28. Ni Y, Tizard I. Lectin-carbohydrate interaction in the immune system. *Vet Immunol Immunopathol.* 1996;55(1-3):205-23.
29. Bard F and Chia J. Cracking the Glycome Encoder: Signaling, Trafficking, and Glycosylation. *Trends in cell biology.* 2016; 26:379-88. □
30. Marth JD and Grewal PK. Mammalian glycosylation in immunity. *Nature reviews Immunology.* 2008; 8:874-887. □
31. Vojta A, et al. Glyco-genes change expression in cancer through aberrant methylation. *Biochim Biophys Acta.* 2016;1860:1776-85. □
32. Stowell SR, Ju T and Cummings RD. Protein glycosylation in cancer. *Annual review of pathology.* 2015; 10:473-510.
33. Xiaolu Ma, et al. Functional roles of sialylation in breast cancer progression through miR-26a/26b targeting ST8SIA4. *Cell Death Dis.* 2016;12.
34. Lee YL, et al. Comprehensive N-Glycome Profiling of Cultured Human Epithelial Breast Cells Identifies Unique Secretome N-Glycosylation Signatures Enabling Tumorigenic Subtype Classification. *J. Proteome Res.* 2014;13:4783–4795.

APPENDIX A

MATLAB ALGORITHM FOR FLOW CYTOMETRY PROTOCOL FOR PC

Figure A.1 Matlab algorithm to unbin 10-bit data from FC500, remove outliers from the data, and arrange the data in an array format.

```
%Input the Excel Data File
XLS_Data_Sheet = input('Please Enter Full File Name:\n','s')
Data_Sheet = XLS_Data_Sheet;
%Input each Cell Line
XLS_Sheet = input('Please Enter Cell Line:\n','s')
Cell_Line = XLS_Sheet;
[SL_array,~,~] = xlsread(Data_Sheet,Cell_Line);

%Each column (vector) represents each SL in binned data
Bin = SL_array(:,1);
SL1_binned = SL_array(:,2);
SL2_binned = SL_array(:,3);
SL3_binned = SL_array(:,4);
SL4_binned = SL_array(:,5);
SL5_binned = SL_array(:,6);
SL6_binned = SL_array(:,7);
SL7_binned = SL_array(:,8);
SL8_binned = SL_array(:,9);
SL9_binned = SL_array(:,10);
%-----

are                                     %Unbin the data by extracting elements that
to                                     %non-zero and greater than count, then adding
to                                     %count and repeating.

SL1_data = [];
count = 0;
while count < max(SL1_binned);
    SL1_unbinned = find(SL1_binned>count);
    count = count+1;
    disp(SL1_unbinned);
    SL1_data = [SL1_data; SL1_unbinned];
end

%Removing the Outliers from SL_unbinned data using IQD
%IQD = input('Please enter IQD value')

% results = find(SL>Q1 & SL<Q3)                                     %Interquartile range = Q3-Q1
SL1_IQR = iqr(SL1_data);                                           %Lower Limit = Q1-1.5*SL1_IQR
```

```

%Upper Limit = Q3+1.5*SL1_IQR
SL1_Q1 = quantile(SL1_data, 0.25);
SL1_Q3 = quantile(SL1_data, 0.75);

IQD_value = input('Please enter Interquartile distance value\n');

Lower_Cut_Off = (SL1_Q1-(IQD_value*SL1_IQR));
Upper_Cut_Off = (SL1_Q3+(IQD_value*SL1_IQR));
%Remove Outliers w/ prompted IQD
SL1_clean = find(SL1_data > Lower_Cut_Off & SL1_data < Upper_Cut_Off);
SL1 = SL1_data(SL1_clean);
length(SL1); %Number elements in SL array
removed %Number of data points
Number_Outliers_Removed = length(SL1_data) - length(SL1);

%Randomize Data for LDA
SL1_rand = SL1(randperm(length(SL1)));

%-----
-
%Now to process SL2-SL9 in the same way repeated

SL2_data = [];
count = 0;
while count < max(SL2_binned);
    SL2_unbinned = find(SL2_binned>count);
    count = count+1;
    disp(SL2_unbinned);
    SL2_data = [SL2_data; SL2_unbinned];
end

SL2_IQR = iqr(SL2_data);
SL2_Q1 = quantile(SL2_data, 0.25);
SL2_Q3 = quantile(SL2_data, 0.75);

Lower_Cut_Off = (SL2_Q1-(IQD_value*SL2_IQR));
Upper_Cut_Off = (SL2_Q3+(IQD_value*SL2_IQR));

SL2_clean = find(SL2_data > Lower_Cut_Off & SL2_data < Upper_Cut_Off);
SL2 = SL2_data(SL2_clean);

length(SL2);
Number_Outliers_Removed = length(SL2_data) - length(SL2);

SL2_rand = SL2(randperm(length(SL2)));

%-----

SL3_data = [];
count = 0;
while count < max(SL3_binned);
    SL3_unbinned = find(SL3_binned>count);
    count = count+1;
    disp(SL3_unbinned);
    SL3_data = [SL3_data; SL3_unbinned];
end

SL3_IQR = iqr(SL3_data);
SL3_Q1 = quantile(SL3_data, 0.25);
SL3_Q3 = quantile(SL3_data, 0.75);

```



```

Lower_Cut_Off = (SL3_Q1-(IQD_value*SL3_IQR));
Upper_Cut_Off = (SL3_Q3+(IQD_value*SL3_IQR));

SL3_clean = find(SL3_data > Lower_Cut_Off & SL3_data < Upper_Cut_Off);
SL3 = SL3_data(SL3_clean);

length(SL3);
Number_Outliers_Removed = length(SL3_data) - length(SL3);

SL3_rand = SL3(randperm(length(SL3)));

%-----

SL4_data = [];
count = 0;
while count < max(SL4_binned);
    SL4_unbinned = find(SL4_binned>count);
    count = count+1;
    disp(SL4_unbinned);
    SL4_data = [SL4_data; SL4_unbinned];
end

SL4_IQR = iqr(SL4_data);
SL4_Q1 = quantile(SL4_data, 0.25);
SL4_Q3 = quantile(SL4_data, 0.75);

Lower_Cut_Off = (SL4_Q1-(IQD_value*SL4_IQR));
Upper_Cut_Off = (SL4_Q3+(IQD_value*SL4_IQR));

SL4_clean = find(SL4_data > Lower_Cut_Off & SL4_data < Upper_Cut_Off);
SL4 = SL4_data(SL4_clean);

length(SL4);
Number_Outliers_Removed = length(SL4_data) - length(SL4);

SL4_rand = SL4(randperm(length(SL4)));

%-----

SL5_data = [];
count = 0;
while count < max(SL5_binned);
    SL5_unbinned = find(SL5_binned>count);
    count = count+1;
    disp(SL5_unbinned);
    SL5_data = [SL5_data; SL5_unbinned];
end

SL5_IQR = iqr(SL5_data);
SL5_Q1 = quantile(SL5_data, 0.25);
SL5_Q3 = quantile(SL5_data, 0.75);

Lower_Cut_Off = (SL5_Q1-(IQD_value*SL5_IQR));
Upper_Cut_Off = (SL5_Q3+(IQD_value*SL5_IQR));

SL5_clean = find(SL5_data > Lower_Cut_Off & SL5_data < Upper_Cut_Off);
SL5 = SL5_data(SL5_clean);

length(SL5);
Number_Outliers_Removed = length(SL5_data) - length(SL5);

SL5_rand = SL5(randperm(length(SL5)));

```

```

%-----
SL6_data = [];
count = 0;
while count < max(SL6_binned);
    SL6_unbinned = find(SL6_binned>count);
    count = count+1;
    disp(SL6_unbinned);
    SL6_data = [SL6_data; SL6_unbinned];
end

SL6_IQR = iqr(SL6_data);
SL6_Q1 = quantile(SL6_data, 0.25);
SL6_Q3 = quantile(SL6_data, 0.75);

Lower_Cut_Off = (SL6_Q1-(IQD_value*SL6_IQR));
Upper_Cut_Off = (SL6_Q3+(IQD_value*SL6_IQR));

SL6_clean = find(SL6_data > Lower_Cut_Off & SL6_data < Upper_Cut_Off);
SL6 = SL6_data(SL6_clean);

length(SL6);
Number_Outliers_Removed = length(SL6_data) - length(SL6);

SL6_rand = SL6(randperm(length(SL6)));

%-----
SL7_data = [];
count = 0;
while count < max(SL7_binned);
    SL7_unbinned = find(SL7_binned>count);
    count = count+1;
    disp(SL7_unbinned);
    SL7_data = [SL7_data; SL7_unbinned];
end

SL7_IQR = iqr(SL7_data);
SL7_Q1 = quantile(SL7_data, 0.25);
SL7_Q3 = quantile(SL7_data, 0.75);

Lower_Cut_Off = (SL7_Q1-(IQD_value*SL7_IQR));
Upper_Cut_Off = (SL7_Q3+(IQD_value*SL7_IQR));

SL7_clean = find(SL7_data > Lower_Cut_Off & SL7_data < Upper_Cut_Off);
SL7 = SL7_data(SL7_clean);

length(SL7);
Number_Outliers_Removed = length(SL7_data) - length(SL7);

SL7_rand = SL7(randperm(length(SL7)));

%-----
SL8_data = [];
count = 0;
while count < max(SL8_binned);
    SL8_unbinned = find(SL8_binned>count);
    count = count+1;
    disp(SL8_unbinned);
    SL8_data = [SL8_data; SL8_unbinned];
end

```

```

SL8_IQR = iqr(SL8_data);
SL8_Q1 = quantile(SL8_data, 0.25);
SL8_Q3 = quantile(SL8_data, 0.75);

Lower_Cut_Off = (SL8_Q1-(IQD_value*SL8_IQR));
Upper_Cut_Off = (SL8_Q3+(IQD_value*SL8_IQR));

SL8_clean = find(SL8_data > Lower_Cut_Off & SL8_data < Upper_Cut_Off);
SL8 = SL8_data(SL8_clean);

length(SL8);
Number_Outliers_Removed = length(SL8_data) - length(SL8);

SL8_rand = SL8(randperm(length(SL8)));

%-----
SL9_data = [];
count = 0;
while count < max(SL9_binned);
    SL9_unbinned = find(SL9_binned>count);
    count = count+1;
    disp(SL9_unbinned);
    SL9_data = [SL9_data; SL9_unbinned];
end

SL9_IQR = iqr(SL9_data);
SL9_Q1 = quantile(SL9_data, 0.25);
SL9_Q3 = quantile(SL9_data, 0.75);

Lower_Cut_Off = (SL9_Q1-(IQD_value*SL9_IQR));
Upper_Cut_Off = (SL9_Q3+(IQD_value*SL9_IQR));

SL9_clean = find(SL9_data > Lower_Cut_Off & SL9_data < Upper_Cut_Off);
SL9 = SL9_data(SL9_clean);

length(SL9);
Number_Outliers_Removed = length(SL9_data) - length(SL9);

SL9_rand = SL9(randperm(length(SL9)));

%-----
%Find lengths of all SLs in Array
SL1_length = length(SL1_rand);
SL2_length = length(SL2_rand);
SL3_length = length(SL3_rand);
SL4_length = length(SL4_rand);
SL5_length = length(SL5_rand);
SL6_length = length(SL6_rand);
SL7_length = length(SL7_rand);
SL8_length = length(SL8_rand);
SL9_length = length(SL9_rand);

%Export all SL vectors independently to the Excel Data File
xlswrite(Data_Sheet,SL1_rand, Cell_Line, 'AS3');
xlswrite(Data_Sheet,SL2_rand, Cell_Line, 'AT3');
xlswrite(Data_Sheet,SL3_rand, Cell_Line, 'AU3');
xlswrite(Data_Sheet,SL4_rand, Cell_Line, 'AV3');
xlswrite(Data_Sheet,SL5_rand, Cell_Line, 'AW3');
xlswrite(Data_Sheet,SL6_rand, Cell_Line, 'AX3');
xlswrite(Data_Sheet,SL7_rand, Cell_Line, 'AY3');
xlswrite(Data_Sheet,SL8_rand, Cell_Line, 'AZ3');

```

```

xlswrite(Data_Sheet,SL9_rand, Cell_Line, 'BA3');

%Find SL with the least number of elements
SL_lcd = [SL1_length; SL2_length; SL3_length;
SL4_length;SL5_length;SL6_length;SL7_length;SL8_length;SL9_length]

%Use smallest SL data set to truncate and align other SL vectors
LDA_minimum = min(SL_lcd)

SL1_LDA = SL1_rand(1:LDA_minimum);
SL2_LDA = SL2_rand(1:LDA_minimum);
SL3_LDA = SL3_rand(1:LDA_minimum);
SL4_LDA = SL4_rand(1:LDA_minimum);
SL5_LDA = SL5_rand(1:LDA_minimum);
SL6_LDA = SL6_rand(1:LDA_minimum);
SL7_LDA = SL7_rand(1:LDA_minimum);
SL8_LDA = SL8_rand(1:LDA_minimum);
SL9_LDA = SL9_rand(1:LDA_minimum);

ALL_DATA = horzcat(SL1_LDA, SL2_LDA, SL3_LDA, SL4_LDA, SL5_LDA, SL6_LDA,
SL7_LDA, SL8_LDA, SL9_LDA)

%csvwrite('Output_File.csv',ALL_DATA,1,1)

xlswrite(Data_Sheet,ALL_DATA, Cell_Line, 'BD3')

beep
beep
beep

```

APPENDIX B

HUMAN N-LINKED GLYCANS

Table B.1 High mannose N-linked glycan structures and monoisotopic masses $M(i)$ and monoisotopic masses + Rapifluor-MS glycan tag $M(i)+RFMS$








Glycan Structure	Name	$M(i)$	$M(i)+RFMS$
	M3	892.3172	1221.5024
	M4	1054.3700	1383.5552
	M5	1216.4228	1545.6080
	M6	1378.4756	1707.6608
	M7	1540.5284	1869.7136
	M8	1702.5812	2031.7664
	M9	1864.6340	2193.8192

Table B.2 Bi-antennary N-linked glycan structures, monoisotopic masses M(i), and monoisotopic masses + Rapifluor-MS glycan tag M(i)+RFMS.

















































Glycan Structure	Name	M(i)	M(i)+RFMS	Glycan Structure	Name	M(i)	M(i)+RFMS
	A2	1298.4760	1627.6612		FA2	1444.5339	1773.7191
	A2B	1501.5554	1830.7406		FA2B	1647.6133	1976.7985
	A2BG1	1663.6082	1992.7934		FA2BG1	1809.6661	2138.8513
	A2BG1S1	1954.7036	2283.8888		FA2BG1S1	2100.7615	2429.9467
	A2BG2A2	2231.8189	2561.0050		FA2BG2	1971.7189	2300.9041
	A2BG2S1	2116.7564	2445.9416		FA2BG2S1	2262.8143	2591.9995
	A2BG2S2	2407.8518	2737.0370		FA2BG2S2	2553.9097	2883.0949
	A2G2A2G2A1	2555.9254	2885.1106		FA2G1	1606.5867	1935.7719
	A2G1	1460.5288	1789.7140		FA2G1S1	1897.6821	2226.8673
	A2G1S1	1751.6242	2080.8094		FA2G2	1768.6395	2097.8247
	A2G2	1622.5816	1951.7668		FA2G2S1	2059.7349	2388.9201
	A2G2A2	2028.7404	2357.9256		FA2G2S2	2350.8303	2680.0155
	A2G2S1	1913.6770	2242.8622		FFA2	1590.5918	1919.7770
	A2G2S2	2204.7724	2533.9576		FFA2B	1793.6712	2122.8564
					FFA2BG1S1	2246.8194	2576.0046
					FFA2G1S1	2043.7400	2372.9252
					FFFA2BG1S1	2392.8773	2722.0625

Table B.3 Tri-antennary N-linked glycan structures, monoisotopic masses M(i), and monoisotopic masses + Rapifluor-MS glycan tag M(i)+RFMS.

Glycan Structure	Name	M(i)	M(i)+RFMS
	A3	1501.5554	1830.7406
	A3G1	1663.6082	1992.7934
	A3G2A1	2028.7404	2357.9256
	A3G2A2	2231.8198	2561.0050
	A3G2S1	2116.7564	2445.9416
	A3G1A1	1866.6876	2195.8728
	A3G1S1	1954.7036	2283.8888
	A3G2S2	2407.8518	2737.0370
	A3G3	1987.7138	2316.8990
	A3G3A1	2190.7932	2519.9784
	A3G3A1S1	2481.8886	2811.0738
	A3G3A2	2393.8726	2723.0578
	A3G3A3	2596.9520	2926.1372
	A3G3A3G2	2921.0576	3250.2428
	A3G3S1	2278.8092	2607.9944
	A3G3S2	2569.9046	2899.0898
	A3G3S3	2861.0000	3190.1852







































































Glycan Structure	Name	M(i)	M(i)+RFMS	Glycan Structure	Name	M(i)	M(i)+RFMS
	FA3	1647.6133	1976.7985		FA3G2A1	2174.7983	2503.9835
	FA3G1	1809.6661	2138.8513		FA3G2A2	2377.8777	2707.0629
	FA3G1S1	2100.7615	2429.9467		FA3G3A1	2336.8511	2666.0363
	FA3G1A1	2012.7455	2341.9307		FA3G3A2	2539.9305	2869.1157
	FA3G2	1971.7189	2300.9041		FA3G3A3	2743.0099	3072.1951
	FA3G2S1	2262.8143	2591.9995		FA3G3A2S1	2668.9731	2998.1583
	FA3G2S2	2553.9097	2883.0949		FA3G3A1S1	2506.9203	2836.1055
	FA3G3	2133.7717	2462.9569		FFA3	1793.6712	2122.8564
	FA3G3S1	2424.8671	2754.0523		FFA3G1	1955.7240	2284.9092
	FA3G3S2	2715.9625	3045.1477		FFA3G1S1	2246.8194	2576.0046
	FA3G3S3	3007.0579	3336.2431		FFFA3G1S1	2392.8773	2722.0625

Table B.4 Tetra-antennary N-linked glycan structures, monoisotopic masses M(i), and monoisotopic masses + Rapifluor-MS glycan tag M(i)+RFMS.

Glycan Structure	Name	M(i)	M(i)+RFMS	Glycan Structure	Name	M(i)	M(i)+RFMS
	A4G4S2	2935.0368	3264.2220		A4G3A1G1S2	3138.1162	3467.3014
	A4G3A1	2393.8726	2723.0578		A4G3A1G1S3	3429.2116	3758.3968
	A4G3A1G1	2555.9254	2885.1106		A4G3A1S2	2976.0634	3305.2486
	A4G3A1G1S1	2847.0208	3176.2060		A4G4A3G2S1	3577.2852	3906.4704
	A4G3A2	2596.9520	2926.1372		A4G3A2G1S1	3050.1002	3379.2854
	A4G3A2G2	2921.0576	3250.2428		A4G3A2	2596.9520	2926.1372
	A4	1704.6348	2033.8200		A4G3A2G2	2921.0576	3250.2428
	A4G1S1	2157.7830	2486.9682		A4G3A3	2800.0314	3129.2166
	A4G1A1	2069.7670	2398.9522		A4G3A3G2	3124.1370	3453.3222
	A4G1	1866.6876	2195.8728		A4G3A3G3	3286.1898	3615.3750
	A4G2	2028.7404	2357.9256		A4G4A3G1S1	3415.2324	3744.4176
	A4G3	2190.7932	2519.9784		A4G4A3G3	3448.2426	3777.4278
	A4G3S1	2481.8886	2811.0738		A4G4A4	3165.1636	3494.3488
	A4G3A1	2393.8726	2723.0578		A4G4A4G1	3368.2430	3697.4282
	A4G2S1	2319.8358	2649.0210		A4G4S1	2643.9414	2973.1266
	A4G2S2	2610.9312	2940.1164		A4G4S2	2935.0368	3264.2220
	A4G2A1	2231.8198	2561.0050		A4G4S3	3226.1322	3555.3174
	A4G2A2	2434.8992	2764.0844		A4G4S4	3517.2276	3846.4128
	A4G2A2G2	2759.0048	3088.1900				

Table B.5 Fucosylated tetra-antennary N-linked glycan structures, monoisotopic masses M(i), and monoisotopic masses + Rapifluor-MS glycan tag M(i)+RFMS.

Glycan Structure	Name	M(i)	M(i)+RFMS	Glycan Structure	Name	M(i)	M(i)+RFMS
	FA4G4	2498.9039	2828.0891		FA4G3S1	2627.9465	2957.1317
	FA4G4S2	3081.0947	3410.2799		FA4G3S3	3210.1373	3539.3225
	FA4G4A1G1S1	2993.0787	2994.71		FA4G3A3	2946.0893	3275.2745
	FA4	1850.6927	2179.8779		FA4G4A2G2S1	3520.2637	3849.4489
	FA4G1A1	2215.8249	2545.0101		FA4G4A2G2S2	3811.3591	4140.5443
	FA4G1S1	2303.8409	2633.0261		FA4G4A3G1S1	3561.2903	3890.4755
	FA4G1	2012.7455	2341.9307		FA4G4A3G2S1	3723.3431	4052.5283
	FA4G2	2174.7983	2503.9835		FA4G4A3G3	3594.3005	3923.4857
	FA4G3	2336.8511	2666.0363		FA4G4A3G3S1	3885.3959	4214.5811
	FA4G2A1	2377.8777	2707.0629		FA4G4A4G3	3797.3799	4126.5651
	FA4G2A2	2580.9571	2910.1423		FA4G4S1	2789.9993	3119.1845
	FA4G2A2G2	2905.0627	3234.2479		FA4G4S2	3081.0947	3410.2799
	FA4G2S1	2465.8937	2795.0789		FA4G4S3	3372.1901	3701.3753
	FA4G2S2	3081.0947	3410.2799		FA4G4A1S3	3575.2695	3904.4547
	FA4G3A1G1S3	3575.2695	3904.4547		FA4G4S4	3663.2855	3992.4707
	FA4G3A1G1S1	2993.0787	3322.2639		FA4G4A4	3311.2215	3604.4067
	FA4G3A1G1S2	3284.1741	3613.3593		FFA4	1996.7506	2325.9358
	FA4G4A1S3	3575.2695	3904.4547		FFA4G1S1	2449.8988	2779.0840
	FA4G3A1S1	2831.0259	3160.2111		FFA4G1A1	2361.8828	2691.0680
	FA4G3A1S2	3122.1213	3451.3065		FFA4G2S2	2903.0470	3232.2322
	FA4G3A2G2	3067.1155	3396.3007		FFA4G4A3G1S1	3707.3482	4036.5334
	FA4G3A3G1	3108.1421	3437.3273		FFA4G4A3G3S1	4031.4538	4360.6390
	FA4G3A3G2	3270.1949	3599.3801		FFFA4	2142.8085	2471.9937
	FA4G3A3G3	3432.2477	3761.4329		FFFA4G1	2304.8613	2634.0465