Using Synthetic Lectins To Investigate Metastatic Potential In Colon Cancer

Erin E. Gatrone
University of South Carolina

Follow this and additional works at: https://scholarcommons.sc.edu/etd
Part of the Chemistry Commons

Recommended Citation

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.
USING SYNTHETIC LECTINS TO INVESTIGATE METASTATIC POTENTIAL IN COLON CANCER

by

Erin E. Gatrone

Bachelor of Science
Canisius College, 2009

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Chemistry

College of Arts and Sciences

University of South Carolina

2016

Accepted by:

John J. Lavigne, Major Professor

Michael L. Myrick, Chair, Examining Committee

Thomas M. Makris, Committee Member

Edsel A. Pena, Committee Member

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

This dissertation is dedicated to my parents who have served as my first and life-long advisors. It is also dedicated to all educators, both in and out of the classroom, who have shaped the individual that I am today. I will be forever grateful.
ACKNOWLEDGMENTS

I, first and foremost, would like to thank my parents. This gratitude is not just for the past four and a half years, but also for a lifetime of encouragement, guidance, and harsh words of advice, when you both knew that is what I needed. You both began to train me as an independent researcher long before I knew that is the direction I wanted my life to go in, however, you knew this skill would provide a level of critical thinking applicable to any field. At the time, I thought we were doing science experiments and puzzles for fun, and while that may be true you were also instilling a passion for knowledge that would remain with me until today.

To Matthew John, there are not enough words in the dictionary to describe how grateful I am for all the ways you have supported me both throughout this program and for the past eight years. You have been my husband, best friend, personal chef, chauffer, grader, lab assistant, cheerleader, and counselor when I needed any of these things. You have made what would have been impossible for me on my own, to be possible. You also accepted that Anna and I were a package deal for the past four and a half years. Bailey and I will forever be grateful to have you in our lives. I cannot mention Bailey, my dog, without truly thanking her for being my soul mate and the hug I needed at the end of the day.

To Laura, Ralph, Allison, and Mike I could not have asked for better siblings. You were the light I needed at the end of a long day or the texts that gave me a push of encouragement. Laura, who knew that when we were playing school in our basement, or
should I say you were making me play school, that we would both be lifelong learners and educators. I would also like to thank my extended family (Aunt Leslie, Uncle Jake, Shannon, Callan, Jacob, and Gram) as well as my second mom (Deb). Deb, I thank you for raising Matthew to be an incredible person, as well as being my second mom and Bailey’s Grandma.

To my SMEK girls and Hana. You four have been a lifeline for me for the past 15 years. I am grateful for text messages, visits, and phone calls that we have daily. You are my best friends and I can’t imagine one minute without you in my life.

To Kaylee, I would like to thank you for being my first graduate school friend and with that a life-long friend. You and Chris were my piece of home while in South Carolina. While some people did not understand our love of scarves, sweaters, all things fall, baking, apple picking, margaritas, and snow, I am thankful for all of our extremely fun days and nights that were bright spots in the last four and a half years.

To Kathy and Martina you have been incredible friends that I plan on having for the rest of my life. Not only, were you a support system for me, but you also gave Bailey a best friend with Boomer. Kathy, I can’t begin to thank you for your guidance both in and out of lab. You have taught me an incredible amount about lab, writing, and research. You were truly my day-to-day advisor and have had a lot to do with the researcher and person I am today.

To Anna, I do not know where to begin with my gratitude. You and I have been through more ups and downs than most people in my life. I feel like we have fit a lifetime of experiences in four and a half years. I treasure the memories we have had in the front of a classroom together as well as those where we were trying to learn
something from the back of a classroom. I have grown up with you and will continue to grow with you.

To former and current “Team Lavigne” members I am thankful for your guidance and support. I will not forget all of the fun times we had even when doing tasks like inventory, defrosting freezers, and our 3 hour long sub-group meetings. I will look fondly upon the memories I had with our small SL group in the beginning with Tanya, Nadee, Kathy, and Anna.

To my committee members, I truly believe that I won the jackpot when you all agreed to be on my committee. I won’t forget the guidance you have given me not just in my candidacy exams, but also on a day-to-day basis.

To Dr. Lavigne, I am grateful that you took a chance on me when you decided to put an analytical researcher on a biochemistry project in your lab. You provided me with an opportunity to be an independent researcher. I appreciate the freedom you gave me with my project to direct it wherever I wanted it to go, whether you whole-heartedly agreed with this direction or not. You knew when to push and when to stand back and let me figure it out. You also looked upon mistakes as an opportunity to learn, believe me I am very grateful for this. You have taught me how to “see the forest through the trees”, as well as a ridiculous amount of analogies that I use daily in class.
ABSTRACT

Cancer is a disease that affects millions of people each year and it is well established that early diagnosis is a key factor towards improving survival rate, with that in mind there is a great need for better diagnostics. From the earliest onset of cancer, changes can be detected in the type and amount of carbohydrates expressed by the cancer cells, i.e., aberrant glycosylation. This process results in differing glycosylation patterns that can be used to detect cancer. Aberrant glycosylation can also greatly affect the potential of cancer cells to metastasize and can be used to stage the disease as well. In order to study these unique glycosylation patterns, synthetic lectins have been developed that are used in a cross-reactive sensor array. This array has not only been shown to be able to differentiate several cancer cell lines among different types of cancer, but also by metastatic potential.

Specifically, the dissertation will describe 1) investigating the assay’s effectiveness with secreted proteins, rather than membrane proteins, to develop a fluid-based diagnostic, 2) the investigation of the metastatic reversion of an isogenic cell line, 3) the use of the synthetic lectin array to diagnose colon cancer in a longitudinal mouse study, and 4) the application of statistical methods to the sensor array data to draw biochemical information.
TABLE OF CONTENTS

DEDICATION .......................................................................................................................... iii

ACKNOWLEDGMENTS .......................................................................................................... iv

ABSTRACT ........................................................................................................................... vii

LIST OF TABLES .................................................................................................................. xi

LIST OF FIGURES ................................................................................................................. xii

LIST OF ABBREVIATIONS .................................................................................................. xvii

CHAPTER 1 : INTRODUCTION AND BACKGROUND ......................................................... 1

1.0 ABSTRACT ....................................................................................................................... 1

1.1 COLORECTAL CANCER FACTS AND CURRENT DIAGNOSTICS ..................... 1

1.2 ABERRANT GLYCOSYLATION ................................................................................. 4

1.3 CANCER ASSOCIATED GLYANS ........................................................................... 4

1.4 NATURAL LECTINS USED IN CANCER DIAGNOSTICS .................................. 6

1.5 BORONIC ACID-DIOL INTERACTION ..................................................................... 8

1.6 BORONIC ACID SENSORS FOR SACCHARIDES .................................................. 9

1.7 CROSS-REACTIVE SENSOR ARRAYS ..................................................................... 10

1.8 SYNTHETIC LECTINS ............................................................................................... 10

1.9 FINDING SYNTHETIC LECTINS ............................................................................ 12

1.10 SYNTHETIC LECTINS INCUBATED CANCER ASSOCIATED GLYCANS ............ 13

1.11 SYNTHETIC LECTINS WITH CELL LINES ............................................................ 16

1.12 DATA PROCESSING .................................................................................................. 19
CHAPTER 2: GLYCAN DETECTION USING SECRETED PROTEIN

2.0 ABSTRACT

2.1 INTRODUCTION

2.2 NEED FOR FLUID BASED DIAGNOSTIC

2.3 BACKGROUND

2.4 SECRETED PROTEIN PRE-STUDY

2.5 SECRETED PROTEIN STUDY WORKFLOW

2.6 SECRETED PROTEIN STUDY DATA AND DISCUSSION

2.7 CONCLUSION

2.8 FUTURE WORK

2.9 EXPERIMENTAL METHODS

CHAPTER 3: INVESTIGATING METASTATIC POTENTIAL WITH AN ISOGENIC CELL LINE

3.0 ABSTRACT

3.1 INTRODUCTION

3.2 PART 1: INVESTIGATING REVERSION OF METASTATIC POTENTIAL

3.3 PART 1: EXPERIMENTAL WORKFLOW

3.4 DATA ANALYSIS AND DISCUSSION

3.5 CONCLUSIONS AND FUTURE WORK

3.6 PART 2: INVESTIGATING METASTATIC POTENTIAL IN AN IN-VIVO MOUSE STUDY
3.7 PART 2: EXPERIMENTAL WORKFLOW ................................................................. 72
3.8 DATA ANALYSIS AND DISCUSSION .................................................................. 76
3.9 CONCLUSIONS ................................................................................................. 83
3.10 FUTURE WORK ............................................................................................... 83
3.11 EXPERIMENTAL FOR CT26-FL3 REVERSION STUDY
(CHAPTER 3 PART 1) ............................................................................................... 84
3.12 EXPERIMENTAL FOR MOUSE SERUM STUDIES
(CHAPTER 3 PART 2) ............................................................................................... 96

CHAPTER 4 : INVESTIGATING METASTATIC POTENTIAL USING
STATISTICAL METHODS .......................................................................................... 106

4.0 ABSTRACT ........................................................................................................ 106
4.1 INVESTIGATION INTO A HEALTHY CONTROL ............................................. 106
4.2 MICROSCOPE STUDIES WITH FLUORESCENCE REFERENCE BEADS .. 114
4.3 CONCLUSIONS AND FUTURE DIRECTIONS ............................................. 116
REFERENCES ........................................................................................................ 118
LIST OF TABLES

Table 1.1: Overexpression of cancer associated glycans by tissue type. .......................... 6
Table 1.2: Sequences for SLs 1, 3, 4, and 5 ........................................................................ 13
Table 1.3: Cell lines used to evaluate SL array response with cell lines. ................................. 17
Table 1.4: SL 1-9 Sequence and Selectivity Information ............................................................... 28
Table 2.1: Four human colon cell lines used in this experiment with the associated metastatic classification ............................................................................................................ 32
Table 2.2: Table Depicts all cell lines used in this experiment and the associated biological replicate number. It includes the type of protein as well as the number of data points for each cell line. Recall each technical replicate (N=1) is the response of nine synthetic lectins with a particular analyte ................................................................. 37
Table 2.3: Summary of linear discriminant analyses performed and the associated leave-one-out classification accuracies. The number of technical replicates is also listed. Recall, every membrane-bound protein had one biological replicate, while, all secreted protein samples had two biological replicates ...................................................... 45
Table 3.1: Comparison of CT26 and CT26-FL3 cell lines. This table contains metastatic potential information for both cell lines, as well as metastatic potential information after 3-5 passages ................................................................................................. 63
Table 3.2: Sequences for SLs 1-9 ............................................................................................. 65
Table 3.3: Mouse conditions and information for longitudinal mouse study .......................... 72
Table 3.4: Specific blood draws and surgery days for the longitudinal mouse study .......... 73
Table 4.1: Cell lines discriminated with Synthetic Lectin array. The associated metastatic potentials are listed as well as the secondary metastatic site when applicable ..... 107
LIST OF FIGURES

Figure 1.1: Structures of common cancer associated glycans. ................................. 5

Figure 1.3: Interaction of boronic acid with 1,2 and 1,3 diols to form a boronate ester. .... 9

Figure 1.5: This figure depicts the interaction of nine different sensors with a single analyte to produce a unique pattern. .......................... 10

Figure 1.6: Schematic of a generalized synthetic lectin attached to a resin bead. ........ 11

Figure 1.7: Boronic acid and SL interaction with a glycan or glycoprotein. ................. 11

Figure 1.8: a) Split and pool method of creating a library. b) Identifying hit and sequencing of the library. ................................................................. 12

Figure 1.9: SLs 1, 3, 4, and 5 interaction with cancer associated glycans................. 13

Figure 1.10: LDA plot generated with response from SL 1, 3, 4, and 5 with cancer associated glycans. ...................................................................... 15

Figure 1.11: SL array response with cell lines of varying metastatic potential. a) illustrates classification based on individual cell line while b) illustrates classification based on metastatic potential. .......................... 18

Figure 1.12: Experimental workflow used for the data presented in the following chapters. .................................................................................................... 19

Figure 2.1: Data processing and acquisition workflow. ............................................. 28

Figure 2.2: Secreted Protein Initial Study Results. The CT26 +3 cells (black outlined green triangles) were grown with FBS prior to secreted protein isolation from media, while the CT26 +5 cells (black outline pink triangles) were starved (no FBS) prior to secreted protein isolation from media. It was decided that the approach taken would be that or starving the cells prior to secreted protein isolation from media due to the perceived scattering of the CT-26 +3 data points. ........................................ 30
Figure 2.3: Experimental workflow for secreted proteins and membrane-bound proteins and glycoproteins. This workflow was used for all four human colon cell lines. ... 34

Figure 2.4: Data acquisition and processing workflow. This is the same workflow depicted in the secreted protein pre-study. ................................................................. 36

Figure 2.5: Three class LDA plot generated with normalized bright numbers from the SL array generated by incubating fluorescently labeled secreted and membrane proteins. The leave one out classification accuracy was 78% and the N=212. The three classes used were healthy, cancerous non-metastatic, and cancerous metastatic. .... 38

Figure 2.6: Fingerprint patterns for the membrane bound proteins and secreted proteins from the same HCT 116 biological replicate. As mentioned, in text, this was generated using normalized bright numbers. ................................................................. 40

Figure 2.7: Three class LDA plot generated with normalized bright numbers from the SL array generated by incubating fluorescently labeled membrane proteins. The leave one out classification accuracy was 97% and the N=69. The three classes used were healthy, cancerous non-metastatic, and cancerous metastatic. ......................... 42

Figure 2.8: Three-class LDA plot generated with normalized bright numbers from the SL array generated by incubating fluorescently labeled membrane proteins. The leave one out classification accuracy was 88% and the N=143. The three classes used were healthy, cancerous non-metastatic, and cancerous metastatic. ......................... 43

Figure 3.1: A) Picture of the internal organs of a mouse. The cecum is highlighted which is where the cells were injected. B) Pictorial representation of the experimental procedure Dr. Pena's lab used in order to make CT26-FL3 cells. ......................... 62

Figure 3.2: Workflow used for reversion of CT26-FL3 experiments......................... 64

Figure 3.3: Principal Component Analysis (PCA) of secreted proteins from CT26-FL3 passages. ........................................................................................................... 66

Figure 3.4: PCA plot seen in previous figure with circles added to indicate possible groupings. ........................................................................................................... 67
Figure 3.5: LDA plot of CT26-FL3 passages modeled with a healthy cell line (CCD-841 CON) N=116 and the leave one out cross validation was 98%............................68

Figure 3.6: LDA plot of CT26-FL3 passages modeled with a healthy cell line (CCD-841 CON), non-metastatic cell lines (HCT 116 and HT29) and metastatic cell line (LoVo). N=149 and the leave one out cross validation was 97%............................69

Figure 3.7: Invasion assay. a) pictorial representation of an invasion assay being performed. b) Both pictures are the bottom of the trans-membrane well of an invasion assay. The left picture is a photograph of a non-invasive cell line after an invasion assay and the right picture is a highly invasive cell line after an invasion assay.........................................................70

Figure 3.8: Experimental flow chart of both experimental options for the mouse serum samples. A) represents the experimental work-up if the albumin depletion kit is not used. B) represents the experimental work-up if the albumin depletion kit is used.74

Figure 3.9: Fingerprint patterns from SL s 1-9. 3M1 is the mouse used. 3M1-A is experiment A from the flow chart above where no albumin depletion was performed. 3M1-B is experiment B from the flow chart above where albumin depletion was performed..........................................................75

Figure 3.10: PCA plot of day 1 (pre-surgery bleed) and day 8 (sacrificing bleed). ........77

Figure 3.11: PCA plot of day 1 (pre-surgery bleed) and day 8 (sacrificing bleed). The boxes around CT26-FL3-8 and CT26-8 represent that these two groups should be clustering together, while all of the unboxed samples should also be clustering together. This ideally would create two populations. This is created with bright raw numbers..........................................................78

Figure 3.12: PCA plot of day 1 (pre-surgery bleed) and day 8 (sacrificing bleed). The boxes around CT26-FL3-8 and CT26-8 represent that these two groups should be clustering together, while all of the unboxed samples should also be clustering together. This ideally would create two populations. This is created with normalized bright numbers..........................................................79

Figure 3.13: PCA plot of day 1 (pre-surgery bleed-points without outlines), day 4 (near midpoint of the experiment-points with dashed outlines), and day 8 (sacrificing bleed-points with black outline). The boxes around CT26-FL3-8 and CT26-8
represent that these two groups should be clustering together. This is created with normalized bright numbers.

Figure 3.14: PCA plot of day 1 (pre-surgery bleed-points without outlines), day 4 (near midpoint of the experiment-points with dashed outlines), and day 8 (sacrificing bleed-points with black outline). The boxes around CT26-FL3-8 and CT26-8 represent that these two groups should be clustering together. Arrows are drawn to the points that information is not known about where these points should cluster. This is created with normalized bright numbers.

Figure 4.1: LDA array response with cell lines of varying metastatic potential. This analysis was a three-class classification based on metastatic potential (i.e., healthy, cancerous non-metastatic (lowly metastatic), cancerous metastatic (highly metastatic)).

Figure 4.2: LDA plot of N=433 and the leave-one-out classification accuracy was 85%.

Figure 4.3: Training/Test analysis for all cell lines represented in Figure 4.2. The training set was all cell lines except one, while the test set was that individual cell line. The three-class classification accuracies are colored blue, while the two-class classification accuracies are colored green.

Figure 4.4: Class Training/Test analysis for most of the cell lines represented in Figure 4.2. The training set was all cell lines except one, while the test set was that individual cell line. The three-class classification accuracies are colored blue, while the two-class classification accuracies are colored green. 3T3 was left out of this analysis to look at its affect on the data and MCF 10A was removed because there would not be another healthy cell line in the model if this was removed to test.

Figure 4.5: Training test plot of 3T3 with the cell lines. As can be seen, the test data should overlap with the healthy data, however, it overlaps with the cancerous non-metastatic data points (yellow diamonds).

Figure 4.6: Time Study done with reference beads to investigate the change in luminosity with exposure time.
Figure 4.7: Luminosity study over 5 minutes elapsed. The luminosity remains consistent at least over 5 minutes.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Amino Acid</td>
</tr>
<tr>
<td>ATCC</td>
<td>American Type Culture Collection</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>BSM</td>
<td>Bovine Submaxillary Mucin</td>
</tr>
<tr>
<td>CCCR</td>
<td>Center for Colon Cancer Research</td>
</tr>
<tr>
<td>CEA</td>
<td>Carcinoembryonic Antigen</td>
</tr>
<tr>
<td>CRC</td>
<td>Colorectal Cancer</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-Dimethylformamide</td>
</tr>
<tr>
<td>Dab</td>
<td>Diaminobutyric Acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic Acid</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal Bovine Serum</td>
</tr>
<tr>
<td>FIT</td>
<td>Fecal Immunochemical Test</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluorescein Isothiocyanate</td>
</tr>
<tr>
<td>Fmoc</td>
<td>Fluorenylmethoxycarbonyl</td>
</tr>
<tr>
<td>gFOBT</td>
<td>Guaiac Fecal Occult Blood Testing</td>
</tr>
<tr>
<td>HBTU</td>
<td>(2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate)</td>
</tr>
<tr>
<td>H₂O</td>
<td>Water</td>
</tr>
<tr>
<td>iFOBT</td>
<td>Immunochemical Fecal Occult Blood Testing</td>
</tr>
<tr>
<td>ivDde</td>
<td>1-(4,4-Dimethyl-2,6-dioxocyclo-hexylidene)-3-methylbutyl</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
</tbody>
</table>
MALDI .......................................................... Matrix Assisted Laser Desorption/Ionization
MeOH ........................................................................................................... Methanol
MS/MS .......................................................................................... Mass Spectrometry/Mass Spectrometry
OVA ................................................................................................................ Ovalbumin
PBA ........................................................................................................ Phenylboronic Acid
PBS ........................................................................................................ Phosphate Buffered Saline
PCA .................................................................................................. Principal Component Analysis
PSM ................................................................................................ Porcine Submaxillary Mucin
ROI ................................................................................................ Region of Interest
SL ........................................................................................................... Synthetic Lectin
sLe^x ........................................................................................................ Sialyl Lewis x
TACA ................................................................................................ Tumor Associated Carbohydrate Antigens
TFA ........................................................................................................... Trifluoro Acetic Acid
TIS ........................................................................................................ Triisopropylsilane
CHAPTER 1:
INTRODUCTION AND BACKGROUND

1.0 ABSTRACT

Colon cancer is the third leading cause of cancer for both men and women. Early diagnosis is essential for optimal patient outcome. This chapter will discuss current colon cancer diagnostics and the necessity for a new colon cancer diagnostic. It will also discuss the Lavigne group’s approach to investigating aberrant glycosylation as it relates to cancer development and progression as a possible diagnostic target.

1.1 COLORECTAL CANCER FACTS AND CURRENT DIAGNOSTICS

In 2008, 7.6 million deaths worldwide were attributed to cancer, accounting for 1 in every 4 deaths in the United States. For both men and women, colorectal cancer is the 3rd deadliest in the United States.¹ In 2015, it was estimated that approximately 50,000 deaths would be due to colorectal cancer.²

With the number of deaths attributed to colorectal cancer, early detection of colorectal cancer is essential for improved patient outcome. If this disease is diagnosed before metastasis the five-year survival rates are reported to be around 90%, however, if diagnosed after metastasis the five-year survival rates are only around 12%.³ Current diagnostics for colon cancer are invasive and lack accuracy especially in the early stages of cancer.

The most common diagnostic for colon cancer is a colonoscopy, which allows visualization of the colon and removal of tissue (polyps) for diagnosis and staging. There
can be many risk factors associated with this procedure, including but not limited to, bleeding, inflammation of the abdominal cavity, and in some cases, perforation of the intestinal wall. Studies suggest that the efficacy of the colonoscopy depends upon the site of the cancer or neoplasm. One particular study followed approximately 55,000 individuals who had received a colonoscopy and found that receiving a colonoscopy reduced colorectal cancer mortality by 29%. When diagnosing distal colorectal cancer there was a 47% reduction in mortality, however, there was no reduction in mortality when it had to do with proximal colorectal cancer.

There are a couple of newer (within the last decade) diagnostics that screen for colon cancer by looking for hidden blood in the stool. These tests include the fecal immunochemical test (FIT) and the guaiac-based fecal occult blood test (gFOBT). The drawback with both diagnostics is that they can diagnose cancer, but cannot identify neoplasms. Unless a tumor is actively bleeding this test will show a negative result, thus requiring that multiple tests be performed. Studies demonstrated that the FIT, previously known as immunochemical fecal occult blood testing (iFOBT), is a superior test in diagnosing cancer relative to the gFOBT. If either of these diagnostics find a positive result, the patient is then scheduled for a colonoscopy, which has its own disadvantages that were discussed previously. While the FIT has success with identifying advances neoplasms, the main disadvantage with the FIT is that it has higher false-negative results in detecting T1 cancer versus T2-T4.

In comparing both the fecal immunochemical test and a colonoscopy, there a few factors to consider: diagnostic usefulness, accessibility to underserved patients, and favorability of the screening technique. With the above information it can be seen that a
colonoscopy is a better diagnostic tool than the fecal immunochemical test (FIT).\textsuperscript{12} In terms of accessibility to underserved patients, the FIT is a less expensive test than a colonoscopy, which is particularly important to the uninsured. Uninsured people have screening rates of 20.7\% compared with people who have private insurance having screen rates of 65\%.\textsuperscript{13} In 2012, it was reported that in the United States 48 million people do not have health insurance. With data suggesting that the cost for health care for cancer in the United States in 2011 was $88.7 billion dollars, the cost of these screening methods needs to be considered.\textsuperscript{2} Studies also show that patients, whether insured or uninsured, are more inclined to participate in screening when the FIT is suggested rather than a colonoscopy.\textsuperscript{13-14}

There are other less-used diagnostics that utilize biomarkers to diagnose and stage colorectal cancer. Carcinoembryonic antigen (CEA) is a glycosylated cell-surface glycoprotein. In healthy patients, CEA levels can range from 2.5 to 5 ng/mL. Testing of CEA levels in patients with colorectal cancer indicated that 4\% of patients with stage I cancer had elevated levels of CEA, while 25 \% of patients with stage II cancer had elevated levels.\textsuperscript{15} In 2000, the American Society of Clinical Oncology (ASCO) recommended that the CEA test no longer be used as a diagnostic tool, but rather be used to monitor disease progression if no other test was available. The ASCO also recommended that CA 19-9, another biomarker often used to diagnose colon cancer, not be used to diagnose or to monitor the disease.\textsuperscript{16}

With the given information about current colon cancer diagnostics it is clear that a novel colon cancer diagnostic is necessary. Ideally, this would be inexpensive in order to
be a feasible option for those in a low socioeconomic class and for those uninsured. It would be best if this had high accuracy with low risk and limited test preparation.

1.2 ABERRANT GLYCOSYLATION

Glycosylation is a post-translational modification where a carbohydrate chain is attached to a membrane or secretory protein. There are two major types of oligosaccharides found on glycoproteins—N-linked and O-linked. O-linked are oligosaccharides are linked to a serine or threonine residue while N-linked are linked to the amide nitrogen of an asparagine residue.\(^\text{17}\)

Aberrant glycosylation is known to take place at the onset of many diseases including cancer as well as inflammation. The glycans present on the glycoproteins can be under, over, or neo-expressed. It is important to note that aberrant glycosylation occurs on both the membrane-bound proteins as well as the secretory proteins.\(^\text{18}\) This process takes place at the onset of cancer and continues as the disease progresses.\(^\text{19}\) Aberrant glycosylation can lead to an increased amount of N-branching in malignant cells leading to more sialyl lewis x (sLe\(^x\)) residues. There is also a correlation with increased N-branching and increased metastatic potential. More will be discussed surrounding sLe\(^x\) in the Section 1.3.\(^\text{20}\)

1.3 CANCER ASSOCIATED GLYANS

The previous section mentioned the importance of glycosylation in terms of understanding cancer and cancer progression. Cancer associated glycans (CAGs), sometimes referred to as tumor associated carbohydrate antigens (TACAs), are glycans that have been shown to be of particular importance to different areas of cancer, i.e.
disease progression, proliferation, metastasis, etc. The structures of common cancer associated glycans are shown below in Figure 1.1.

<table>
<thead>
<tr>
<th>Glycan</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>fucose</td>
<td><img src="image" alt="Fucose" /></td>
</tr>
<tr>
<td>sialic acid</td>
<td><img src="image" alt="Sialic Acid" /></td>
</tr>
<tr>
<td>Tn</td>
<td><img src="image" alt="Tn" /></td>
</tr>
<tr>
<td>Le&lt;sup&gt;x&lt;/sup&gt;</td>
<td><img src="image" alt="Le-x" /></td>
</tr>
<tr>
<td>Le&lt;sup&gt;y&lt;/sup&gt;</td>
<td><img src="image" alt="Le-y" /></td>
</tr>
<tr>
<td>sLe&lt;sup&gt;x&lt;/sup&gt;</td>
<td><img src="image" alt="sLe-x" /></td>
</tr>
<tr>
<td>sLe&lt;sup&gt;a&lt;/sup&gt;</td>
<td><img src="image" alt="sLe-a" /></td>
</tr>
</tbody>
</table>

Figure 1.1: Structures of common cancer associated glycans.

One cancer-associated glycan, sialyl lewis x, a tetra-saccharide, has been theorized to be involved in metastasis, the cancer spreading from the primary tumor to other areas in the body. It is thought that the sLe<sup>x</sup> present on the surface of the cell is able to bind with a lectin (E/P selectin) in order to begin the process of intravasation, where the cancerous cell makes its way into the vasculature, as well as extravasation, where the cancerous cell makes its way out of the vasculature at a secondary tumor site. Selectins, a class of lectins, are cell adhesion molecules that bind glycans. Selectins play a role in the intravasation of cancer cells in order for these cells to migrate to secondary tumor sites. 21
A large review of the role of glycans in cancer reported that it is not one particular change in a glycan that happens in cancer, but rather a pattern of glycan changes. Table 1.1 indicates that over-expression of cancer-associated glycans for different cancer types.

Table 1.1: Overexpression of cancer associated glycans by tissue type.\textsuperscript{18}

<table>
<thead>
<tr>
<th>Cancer glycan</th>
<th>Blood</th>
<th>Breast</th>
<th>Colon</th>
<th>Brain</th>
<th>Prostate</th>
<th>Skin</th>
<th>Lung</th>
</tr>
</thead>
<tbody>
<tr>
<td>sLe\textsuperscript{a}</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>sLe\textsuperscript{a}</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>sTn</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>TF</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Le\textsuperscript{a}</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>GloboH</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSA</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GD2</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GD3</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fucosyl GM1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>GM2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

It is interesting to note, that each cancer type has its unique pattern, however, some cancer types display similar patterns. For instance, breast cancer and colon cancer have very similar patterns, except breast cancer has the over-expression of polysialic acid (PSA), while colon cancer does not.\textsuperscript{22} An ideal diagnostic would be one that would be able to detect changes in glycosylation.

1.4 NATURAL LECTINS USED IN CANCER DIAGNOSTICS

Lectins are carbohydrate-binding proteins that serve a variety of functions in the body. Lectins can serve as recognition signals on the surface of the cell\textsuperscript{23}, play a role in protein folding and cell-cell interaction\textsuperscript{24}. They can also be involved in cellular functions like endocytosis, cell adhesion, cell migration, and transport. Their interactions with carbohydrates are generally reversible in nature.\textsuperscript{25} A common example of a viral protein
that binds to carbohydrates is influenza hemagglutinin. This protein has sites that bind to sialic acid present on a cell surface to initiate the interaction of the virus and host cell.\textsuperscript{24}

With lectins having high selectivity for a particular target, natural lectins and natural lectin arrays have been of interest in research in a variety of areas, however, we will limit this discussion to cancer-related areas. Researchers are investigating lectins as possible diagnostic and therapeutic tools.\textsuperscript{26-27}

One such study investigated natural lectin microarrays as a means for distinguishing glycosylation patterns in serum from normal patients and patients with pancreatic cancer. This study used five different lectins, however, due to the specificity of lectins only seven different glycan structures could be detected. The researchers were able to distinguish between serum samples; however, the experiments required a large amount of sample preparation. It also only probed a small amount of glycosylation changes that could occur from patient to patient.\textsuperscript{26}

Another study investigated the glycan profile from healthy patients, patients with colon adenomas, and those with colorectal cancer in order to determine a potential colon cancer biomarker. This study also used five different natural lectins, which as discussed previously only probes a limited amount of glycan structures. The study concluded that while using some natural lectins, patients with colon cancer, adenomas, and no cancer/adenomas could be successfully separated. However, there were some natural lectins that did not provide this type of information. They also discussed that the patients with cancer had an increased level of sialylation and fucosylation\textsuperscript{27}, however another study mentioned that sialylation could be increased up to a two-fold difference from patient to patient.\textsuperscript{26}
While natural lectins can prove to be valuable for gaining information about glycosylation patterns, there are a few disadvantages to using them. They are very specific for a particular sugar and or sugar linkage, which can be good in some cases. However, it is known that a glycan profile can change depending on cancer type. If there is no known lectin to probe this change, the efficacy of the lectin in a diagnostic reduces. The natural lectins can be particularly difficult to work with and many of the studies that are utilizing these lectins require a large amount of sample preparation and purification. There is not one particular glycan change that can differentiate diseased tissue from normal tissue. This would indicate that it would be advantageous to look for a way of detecting glycan changes that are not as specific as natural lectins. The above reasons are why the research groups have chosen to create synthetic lectins\textsuperscript{28-30}, which will be talked about in further detail in subsequent sections of this chapter.

1.5 BORONIC ACID-DIOL INTERACTION

To investigate the changes in glycosylation and design a sensor to monitor these changes, a part of the sensor has to interact with the glycans. In looking at some of the structures in Figure 1.1, it can be seen that all of the structures contain diols. Boronic acids have the ability to bind 1, 2 and 1, 3-diols covalently, with high affinity, yet reversibly to form 5- or 6-membered boronate esters.\textsuperscript{31-34} This interaction is generally stronger with cis-diols than trans-diols.\textsuperscript{35} Figure 1.2 depicts the formation of this boronate ester from a boronic acid binding covalently with a 1,2 and 1,3 diol.
The reversibility of this interaction is what makes boronic acids useful in a saccharide sensor. The boronic acid is capable of binding strongly to a carbohydrate, however, if a different carbohydrate has a stronger interaction the first will unbind and the second will bind. This property would ensure that the sensor would bind with whichever saccharides had the strongest interaction.

1.6 BORONIC ACID SENSORS FOR SACCHARIDES

With the strong, yet reversible, interaction boronic acids are capable of making with 1,2 and 1,3 diols, boronic acids have been used as part of a sensor for saccharides.\textsuperscript{36-39} This research has been applicable to cancer research as well as other diseases, such as, diabetes.

Sialyl lewis X has been associated with metastasis as well as part of the glycan pattern seen in different cancer types. Some research has been used to find a fluorescent di-boronic acid sensor that can correctly identify those cells containing sLe\textsuperscript{x}, while also having a negative interaction with those cells that do not have sLe\textsuperscript{x}.\textsuperscript{38-39}

While the above examples used boronic acid sensors for applications in cancer, they have also been studied for use as a non-invasive glucose sensor for patients with diabetes.\textsuperscript{34-35} Boronic acid sensors have also been used as a way to study non-covalent drug delivery.\textsuperscript{40}
1.7 CROSS-REACTIVE SENSOR ARRAYS

The previous applications with natural lectins involve the lectins binding with high affinity with the glycans specific for lectin. This interaction is similar to an antigen/antibody interaction. In looking at these applications as well as the biomarkers discussed in the colon cancer diagnostics section, looking for one biomarker that has a high affinity may not be the best approach. Rather than thinking about one sensor for one analyte, one can think about a set of cross-reactive sensors forming a sensor array. This type of interaction is depicted in Figure 1.3.\textsuperscript{41}

![Figure 1.3](image)

Figure 1.3: This figure depicts the interaction of nine different sensors with a single analyte to produce a unique pattern.\textsuperscript{41}

With this approach, one analyte (the red oval) has an interaction with a series of sensors. This interaction can then produce a unique pattern depending on the interaction (or lack of) and the strength of the interaction. Therefore, the response is the entire pattern. This is similar to the mammalian sense of taste or smell. There is not a receptor for each type of food, but rather the response of 4 taste buds (sensors) that send a signal to your brain to recognize what is being eaten.

1.8 SYNTHETIC LECTINS

In order to mimic the interactions between natural lectins and glycans while avoiding the disadvantages of using natural lectins, the Lavigne lab developed synthetic
lectins. These synthetic lectins were short peptide chains functionalized with boronic acids attached to a resin. The advantages to using synthetic lectins are that they are easy to synthesize, biocompatible, and have low toxicity.

![Figure 1.4: Schematic of a generalized synthetic lectin attached to a resin bead.](image)

Figure 1.4 depicts a general synthetic lectin on a resin. It depicts one synthetic lectin on a bead, however it reality there are many more than one. The M, R, and B represent the one-letter amino acid codes with B standing for β-alanine. The X represents a random amino acid chosen. The boronic acids are attached to a non-natural amino acid, diaminobutyric acid, represented as Dab in Figure 1.4.

![Figure 1.5: Boronic acid and SL interaction with a glycan or glycoprotein.](image)

The attached boronic acids would interact with 1,2 and 1,3-diols present on sugar residues forming a boronate ester. This interaction is depicted in Figure 1.5.⁴²
1.9 FINDING SYNTHETIC LECTINS

In order to find the synthetic lectins to use, a library was made using a split and pool method depicted in Figure 1.6a. A portion of the resin was split into individual tubes and a single amino acid was attached, a different amino acid for each tube. The resin was pooled into a tube and re-split so that another amino acid could be added. After all of the random amino acids were added and then end of the peptide was acylated, boronic acids were attached. Theoretically the synthetic lectin, and more specifically the boronic acids would interact with 1,2 and 1,3-diols.

Figure 1.6: a) Split and pool method of creating a library. b) Identifying hit and sequencing of the library.

In order to find SLs that would interact with carbohydrate structures, the library was incubated with fluorescently labeled purified glycoproteins, Ovalbumin (OVA), porcine stomach mucin (PSM), and bovine sub-maxillary mucin (BSM). Figure 1.6b depicts this process. The purified glycoproteins were of varying carbohydrate composition. After incubation, the excess FITC was washed away and hits were identified using fluorescence microscopy. A “hit” was determined to be a bead that fluoresced more than the others beads. The fluorescing of the bead would indicate an interaction between the synthetic lectin and the glycoprotein. The hits were manually
removed from the other beads and the bead sequenced by MS/MS. The sequences for SL 1, 3, 4, and 5 are shown in Table 1.2.

Table 1.2: Sequences for SLs 1, 3, 4, and 5

<table>
<thead>
<tr>
<th>SL ID</th>
<th>Sequence</th>
<th>Selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL 1</td>
<td>Ac-RGD<em>VTFD</em>RBBRM</td>
<td>OVA</td>
</tr>
<tr>
<td>SL 3</td>
<td>Ac-RSD<em>VTTD</em>RBBRM</td>
<td>PSM</td>
</tr>
<tr>
<td>SL 4</td>
<td>Ac-RRD<em>TQTD</em>QBBRM</td>
<td>OVA and PSM</td>
</tr>
<tr>
<td>SL 5</td>
<td>Ac-RAD<em>TRVD</em>VBBRM</td>
<td>PSM</td>
</tr>
</tbody>
</table>

The research involved in Chapter 2 and 3 discuss the use of SLs 1-9, however, the following research presented in Chapter 1 will only discuss the use of SLs 1, 3, 4, and 5. It was not until after this research was published, that an expanded SL array was investigated. The sequences for the remaining SLs are listed in Table 2.1.

1.10 SYNTHETIC LECTINS INCUBATED CANCER ASSOCIATED GLYCANS

Following the hits being sequenced, the SLs were re-synthesized. There were a variety of experiments done in order to optimize the use of the SLs, however, only ones pertinent to this dissertation will be discussed. The re-synthesized SLs were incubated with fluorescently labeled cancer associated glycans.

Figure 1.7: SLs 1, 3, 4, and 5 interaction with cancer associated glycans
Figure 1.7 depicts the interaction that each SL had with each cancer associated glycan. It is important to note that each SL was incubated with each CAG separately. The y-axis is the normalized average luminosity. These values were found by manually circling each bead in Adobe Photoshop®. Essentially, the stronger boronic acid-diol interaction is the larger the normalized average luminosity value.

The important part about the interactions shown above is that each cancer-associated glycan has a unique fingerprint pattern that the four SLs create. For instance, if one looks at the response of SL1 with Le\(^x\) and sLe\(^x\) it may be difficult, taking into account the error bars, to distinguish one from another. However, if this response is expanded to looking at the SL 1 and SL 3 response, it starts to become more clear that these are different.

It is challenging to look at a bar graph to decipher which responses are different, which is why a statistical method called linear discriminant analysis (LDA) is used. This method attempts to reduce the dimensionality of the data into two dimensions. When reducing the dimensionality, the statistical model attempts to minimize the separation within the same group, while maximizing the differences between other groups. The important thing to note about LDA is that it is a guided statistical method, meaning that the computer program is given information about which group each data point should belong. Figure 1.8 is the LDA plot generated from the data in Figure 1.7.
Figure 1.8: LDA plot generated with response from SL 1, 3, 4, and 5 with cancer associated glycans.

Each shape in Figure 1.8 represents the array’s response from four different synthetic lectins. Recall that the program is attempting to group all of the sLe$^x$ points (black triangles) as close together as possible, while attempting to group these points away from the Le$^x$ data points (green triangles). The circles surrounding the data points represent the 95% confidence intervals. LDA is capable of giving cross-validation values meaning that it will take a point out, for example a black triangle, and the program will attempt to put that black triangle back in and see where that point will classify itself. This particular data set has a classification accuracy of 97%. The interesting thing to note is that while there is some overlapping above, for instance between Le$^a$ (red squares) and sLe$^a$ (pink triangles), these sugars are structurally similar. The only difference between these two polysaccharides is the addition of a sialic acid group in the sLe$^a$ polysaccharide.$^{42}$
1.11 SYNTHETIC LECTINS WITH CELL LINES

Being that the SL array response was able to discriminate between structurally similar polysaccharides, the next step was to determine if the SLs were able to distinguish differing cell lines of varying metastatic potential. Metastatic potential will be used throughout the duration of this dissertation and refers to whether the cell line is considered healthy/normal, cancerous non-metastatic/lowly-metastatic, and cancerous metastatic. Table 1.3 lists the cell lines used in the experiment described. For each cell line the table lists the metastatic potential, cancer type (if applicable), and the origin of the cancer cell line if it was metastatic, and the species this cancer is derived from. For instance, LoVo is a human colon cancer metastatic cell line. While it is a colon cancer cell line, it was obtained from its secondary site, a lymph node. CT26-F1 and CT26-FL3 have an asterisk next to them denoting that these are isogenic cell lines meaning that they are derived from the same parent cell line, CT26. For clarity, CT26 is considered non-metastatic/lowly metastatic. No cell line can be completely non-metastatic because given the right conditions a cell can migrate, however, if a cell line has a very low metastatic potential rate it is usually considered non-metastatic. CT26 has a very low metastatic potential (0-5%) and thus is considered non-metastatic. CT26-F1 is cancerous metastatic with a metastatic rate of 50%, and CT26-FL3 is highly metastatic with a metastatic rate of 95%.
The cell lines were grown according to ATCC guidelines and the membrane proteins/glycoproteins were extracted using the Qiagen® membrane extraction kit. Proteins were labeled with FITC and incubated with the four SLs. The beads were analyzed using fluorescence microscopy and imaged. The bright values were obtained using Adobe Photoshop® as described earlier in this chapter. The fingerprint pattern was obtained and LDA was used to better distinguish the similarities and differences in the array response, similarly to how LDA was used while analyzing the SL array with cancer associated glycans described previously. Figure 1.9 illustrates the LDA plot obtained from the SL array’s response with each cell line. Each point is the individual response from each of the four SLs with that particular cell line.
In Figure 1.9a, the data was grouped/classified based on cell line. For example, the program knew that each NIH/3T3 point (black diamond) was different than each HT29 point (orange circle). The program had a cross-validation of 92% when classified by cell line. However when thinking about a clinically relevant diagnostic, at first pass it is more important to answer the following questions: Do I have cancer? and Has the cancer metastasized than to know if one has HCT 116 vs. HT29. Figure 1.9b takes these questions into account. In this LDA plot, the data is grouped by metastatic potential (healthy, cancerous non-metastatic, and cancerous metastatic). When classified by metastatic potential, the cross-validation accuracy is 97%. The false-negative and false-positive rates were found by manually going through the output and looking at whether a cancerous data point was misclassified as healthy or a healthy point was misclassified as cancerous respectively.42
1.12 DATA PROCESSING

The data analyzed in this chapter was analyzed differently than what will be discussed in the rest of the dissertation. The previous luminosity data for the interaction between the SL and carbohydrates was found by incubating the SL with fluorescently labeled glycoprotein and imaging with a fluorescence microscope. After imaging, in order to obtain fluorescence intensity information, Adobe Photoshop® was used. Within this program, a bead was isolated and a histogram generated to give a “bright value”. The work presented in Chapters 2-4 will use a slightly varied experimental set-up. Figure 1.10 depicts the workflow used.

![Diagram](image)

**Figure 1.10:** Experimental workflow used for the data presented in the following chapters.

The synthetic lectins on beads were incubated with fluorescently labeled glycoproteins and these beads were imaged in the same way as discussed previously in this chapter. A typical image is shown in the figure above. The image was run through a program in MATLAB® rather than using Adobe Photoshop.
The program used was one written by Dr. Andrew Greytak, University of South Carolina Chemistry Department. As can be seen from the figure, there is a white box in the middle of the photo. This is called the Region of Interest (ROI) and was used in the MATLAB® program. Only beads inside this region (not touching the border) are analyzed. This was used because it was noted that beads directly under the light source fluoresce brighter than those on the outside of the ROI. The beads may not have necessarily been different intensities, but the difference may have been due to how much of the light source was over the bead.

The program gives a variety of information about the bead. It gives information about bead size, circularity, area, and luminosity information. The luminosity information is given in three channels (red, green, and blue) as well as a “bright” number, which is a weighted average of the first three with the green number being a much larger factor than the red or blue. The green number would be the largest value for these experiments because the bead is incubated with FITC labeled glycoproteins and imaged under a GFP filter. In order for a bead to be counted as “valid” it needs to have an appropriate area to ensure that the program is not counting two beads as one. The bead must be circular to avoid the program looking at bead fragments.

Due to the bit size of the camera, the maximum value for bead intensity is 255. Therefore, in order to obtain information for each bead, the intensity value must be between 0 and 255. To optimize the exposure time and be sure that no synthetic lectin is too bright or dim, each synthetic lectin used (SL 1-9) with a particular analyte is placed under the camera. The brightest SL is found and the time is optimized in order to make
sure that with the exposure time picked the bright number is not above 255. Each SL is then imaged at the optimized exposure time.

Following imaging and data extraction from the MATLAB® program, the data is processed using Excel®. The sensor array data will be referred to as either normalized or raw. If the data is raw, the bright or green numbers used will be the numbers given directly from the MATLAB® program, therefore all of the data was on a scale from 0-255, the scale of the camera. Being that each analyte will have a different amount of FITC present, it was thought that normalizing the data may also be beneficial. Therefore, within each individual analyte (usually a cell line) the brightest SL was found. For instance, SL 1 was the brightest. If this were the case all other SLs data points would be divided by SL1 to put the data on a scale from 0 to 1. Throughout the paper, it will be discussed which type of data analysis was used for the sets of data.

1.13 CONCLUSIONS

Colorectal cancer is a disease that directly affects thousands of men and women, however with an early diagnosis the mortality rate decreases dramatically. Current colon cancer diagnostics are expensive, unreliable, and have the possibility of serious complications that make a novel diagnostic necessary.

Aberrant glycosylation-the under, over, and neo expression of glycans is a process that occurs at the onset of cancer and progress throughout the disease both on membrane bound proteins and secreted proteins. This process provides a target for a novel cancer diagnostic, however, the diagnostic would require something that can probe carbohydrates. Synthetic lectins functionalized with boronic acids can mimic natural
lectins without the disadvantages to natural lectins. The boronic acids can form covalent, yet reversible, bonds with 1, 2 and 1, 3 diols present on the surface of sugar moieties.

A synthetic lectin library was created in order to find synthetic lectins, both selective and cross-reactive, to be used in a sensor array. This sensor array was used to investigate its utility in diagnosing several different cell lines of varying metastatic potential. The array was successful in classifying these cell lines by metastatic potential. The subsequent chapters in this dissertation will discuss the applications of this array including the possibility of the array being used as a fluid-based diagnostic. It will also discuss using the array to investigate metastatic potential in an isogenic cell line as well as in a longitudinal mouse study.
CHAPTER 2:

GLYCAN DETECTION USING SECRETED PROTEIN

2.0 ABSTRACT

Aberrant glycosylation begins at the onset of cancer and continues throughout the progression of the disease. A synthetic lectin array was developed to investigate glycosylation changes. The sensor array was effective in discriminating cell lines based on metastatic potential (healthy, cancerous non-metastatic (lowly metastatic), and cancerous metastatic (highly metastatic)). However, the proteins and glycoproteins used were membrane-bound and as a clinical diagnostic would require a biopsy for which there are a variety of associated disadvantages. Ideally, a fluid-based diagnostic would be used.

This chapter will discuss the need for a fluid-based diagnostic, a pre-study to optimize experimental conditions, and the use of secreted proteins from four human colon cell lines as a proof-of-concept experiment to investigate the synthetic lectin array to be used as a fluid-based assay. The analysis of the synthetic lectin array is able to discriminate membrane bound proteins of four different human colon cell lines with a classification accuracy of 97% and could discriminate secreted proteins from four human colon cell lines with a classification accuracy of 88%. This indicates that the synthetic lectin array was able to classify secreted proteins based on metastatic potential. The chapter will also discuss the future directions for this project.
2.1 INTRODUCTION

Colon cancer is the third most common type of cancer in both men and women worldwide with at least 600,000 colon cancer-related deaths annually.\textsuperscript{1-2} Current colon cancer diagnostics, e.g. colonoscopy, are invasive and often inaccurate because they rely on visual inspection, which can be subjective. Monitoring cancer biomarkers using a fluid-based diagnostic, often referred to as a liquid biopsy, is an optimal strategy though existing methods suffer greatly from low effectiveness associated with high false negative and false positive rates. The work in this chapter describes progress towards developing a blood-based colon cancer diagnostic.\textsuperscript{15-16}

As previously mentioned, the approach used focuses on monitoring aberrant glycosylation, which impacts integral cell membrane and secreted glycoproteins from the onset and throughout progression of the disease.\textsuperscript{18-19} Here, boronic acid functionalized peptides, termed Synthetic Lectins (SLs), are used in a cross-reactive sensor array to monitor changes in glycosylation patterns displayed on cancer cells as well as proteins secreted into media/blood.

2.2 NEED FOR FLUID BASED DIAGNOSTIC

Current colon cancer diagnostics have serious disadvantages including reliability, cost, and time for results.\textsuperscript{4-13} With these factors in consideration a fluid-based diagnostic is ideal. The fluids that can be considered for this type of diagnostic would include blood, semen, urine, etc. A fluid-based diagnostic would reduce the cost, which is advantageous to uninsured populations. It would also allow for easier sample collection.
2.3 BACKGROUND

The synthetic lectin array had been investigated to determine its discriminatory power in differentiating between healthy, cancerous non-metastatic (lowly metastatic), and cancerous metastatic (highly metastatic) cell lines. These cell lines were of different species with some being from mice and some from human. The synthetic lectin array was effective at discriminating seven colon cell lines with a leave one out classification accuracy of 97%. The analysis was run with a three-class classification which means the seven cell lines were classified into three groups (healthy/normal, cancerous non-metastatic, and cancerous metastatic).

In terms of a clinical diagnostic, obtaining cells includes many of the same risk factors and disadvantages as current diagnostics, discussed in Chapter 1. The following chapter will describe a pre-study used to optimize the conditions to grow cells and obtain secreted proteins as well as a proof-of-concept experiment with four human colon cell lines or varying metastatic potential.

2.4 SECRETED PROTEIN PRE-STUDY

A fluid-based diagnostic requires secreted proteins. The SL array, up until this point, had investigated the array’s response with membrane bound proteins and glycoproteins that were extracted from cultured cells. In order to investigate whether the SL array would work with secreted proteins, a proof-of-concept experiment was created in which the proteins secreted into media from cultured cells would be used. The primary questions to be answered were: Can secreted proteins be used? and What growth conditions would be best for this experiment?
Cell culture conditions usually include some sort of base media, fetal bovine serum (FBS), antibiotics/antimycotics, and any additives it requires. For this first experiment the cell line used was CT-26 (ATCC® CRL-2638™). This is a cancerous, lowly, metastatic mouse cell line. It was chosen because it is easily grown and the membrane glycoproteins were used in the initial study.

Two different growing conditions were used for this pre-study. The +3 and +5 seen in the paragraph below refer to the passage numbers from the cell stocks (i.e. how many times the cells grew confluent and were split). Both CT26 +3 and CT26 +5 were grown with FBS (serum) present in the media prior to the experiment. When the cells reached 70% confluence, CT26 +3 had its media (that contained FBS) replaced, while CT26 +5 had media without FBS (serum) added to the cells. The lack of FBS added means that CT26 +5 were starved for 48 hours prior to media removal. Both media conditions were added the cells 48 hours prior to media removal. Antibiotics/antimycotics were present in both cell culture scenarios.

After 48 hours the media was removed from the cells and centrifuged to remove cell debris. The media was pipetted off and put in a new tube. The media was centrifuged in Amicon® 3K centrifuge tubes (EMD Millipore) and the proteins were precipitated with cold acetone. The acetone was then removed and the protein was put into carbonate buffer for fluorescent labeling. The amount of carbonate buffer added changes depending on the amount of protein present. Generally 1-2 mL was used to make a homogenous protein solution. In order to quantify the protein concentration, a BCA assay was done. A small amount of protein solution (200-300 µL) was taken to label. It is advantageous to label a portion of the total protein if further experiments need
to be done. The protein was labeled with fluorescein isothiocyanate (FITC) and the solvent exchanged to PBS using Amicon® 10K centrifuge tubes (EMD Millipore). A BCA assay was repeated to quantify the amount of labeled protein.

Previously, the SL array has been used with membrane proteins/glycoproteins from the seven different cell lines of varying metastatic potential as well as purified proteins (BSA, BSM, PSM, and OVA). For the pre-study, the experimental workflow that was followed was the same workflow for the incubation of purified proteins, which requires a larger protein concentration (0.1mg/mL) and a longer incubation time (20 hours) with the SLs. This protocol was used due to the homogeneous nature of the secreted proteins. Prior to incubation the beads containing the SLs were washed with PBS twice. The beads were incubated with PBS containing 1% BSA for 15 minutes. This solution was removed and 1 mL of the protein solution at the desired concentration was added to nine tubes, each independently containing approximately 2 mg of a different SL (SL 1-9). After incubation, the beads were washed three times with PBS. Each tube of beads was imaged using a fluorescence microscope, Leica® MZ 16F, using a GFP filter set.

Previous work used SLs 1,3,4, and 5 for the array, however, work done concurrently in the lab at the time of this experiment demonstrated an expanded SL array using SLs 1-9 had higher classification accuracies. For this reason, the expanded array was used for all of the experiments presented in this chapter. The sequences for SLs 1-9 are shown below in Table 2.1. The glycoprotein that each SL was found to be selective for is also shown. Synthetic lectins 6-8 were found to be cross-reactive.
Table 2.1: SL 1-9 Sequence and Selectivity Information

<table>
<thead>
<tr>
<th>SL ID</th>
<th>Sequence</th>
<th>Selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL 1</td>
<td>Ac-RGD<em>VTFD</em>RBBRM</td>
<td>OVA</td>
</tr>
<tr>
<td>SL 2</td>
<td>Ac-RTD<em>RFLD</em>VBBRM</td>
<td>OVA</td>
</tr>
<tr>
<td>SL 3</td>
<td>Ac-RSD<em>VTTD</em>RBBRM</td>
<td>PSM</td>
</tr>
<tr>
<td>SL 4</td>
<td>Ac-RRD<em>TQTD</em>QBBRM</td>
<td>OVA and PSM</td>
</tr>
<tr>
<td>SL 5</td>
<td>Ac-RAD<em>TRVD</em>VBBRM</td>
<td>PSM</td>
</tr>
<tr>
<td>SL 6</td>
<td>Ac-RTD<em>NRND</em>FBBRM</td>
<td>-</td>
</tr>
<tr>
<td>SL 7</td>
<td>Ac-RSD<em>YFTD</em>QBBRM</td>
<td>-</td>
</tr>
<tr>
<td>SL 8</td>
<td>Ac-RTD<em>YGND</em>NBBRM</td>
<td>-</td>
</tr>
<tr>
<td>SL 9</td>
<td>Ac-RTD<em>YQVD</em>ABBRM</td>
<td>OVA</td>
</tr>
</tbody>
</table>

Figure 2.1 represents the data acquisition and processing after incubation with the SLs. Each tube of beads was imaged using a fluorescence microscope, Leica® MZ 16F, using a GFP filter set. The images are analyzed with a MATLAB® program and process with Microsoft Excel®. Following data processing, is data analysis using various statistical methods (i.e. LDA or PCA).

Figure 2.1: Data processing and acquisition workflow.
Each synthetic lectin incubated with each analyte was imaged using a camera and fluorescent microscope. Each synthetic lectin incubated with each analyte will give a fluorescent response depending upon the level of interaction. The pictures imaged with the camera are run through a program in MATLAB® that was written by Dr. Andrew Greytak. This allows a numerical value to be assigned for fluorescent intensity. As discussed in Chapter 1, each analyte will have a unique fingerprint pattern created by all nine synthetic lectins. While looking at the fingerprint patterns can be valuable it becomes confusing when attempting to find similarities and differences between analytes. For this reason, statistical methods are implemented. For this analysis, linear discriminant analysis (LDA) was used. LDA is a guided statistical method meaning that the statistical software knows which group each analyte (cell line) should belong. Because the cell lines used have well-established metastatic potentials this was the method chosen.

Figure 2.2 is the linear discriminant analysis plot generated with Systat® 7 and plotted in Microsoft Excel®. This plot was generated as a three-class classification, which means that each analyte (cell line) classifies as one of three groups. The three main groups of importance are the healthy (red squares), non-metastatic (purple circles), and metastatic (blue diamonds) cell lines. CT26 is a non-metastatic (lowly) cell line, however, for ease of data interpretation all CT26 points are plotted as a green triangle. It is important to note that all of the data points except for CT26 +3 and CT26 +5 were generated from membrane bound proteins and glycoproteins. The data points used except for CT26 +3 and CT26 +5 represent the data collected by Kevin Bicker and Jing Sun.
Figure 2.2: Secreted Protein Initial Study Results. The CT26 +3 cells (black outlined green triangles) were grown with FBS prior to secreted protein isolation from media, while the CT26 +5 cells (black outline pink triangles) were starved (no FBS) prior to secreted protein isolation from media. It was decided that the approach taken would be that or starving the cells prior to secreted protein isolation from media due to the perceived scattering of the CT-26 +3 data points.

The first question was whether secreted proteins worked with the SL array. As seen in Figure 2.2, it is seen that a response is generated from SLs 1-9 with both CT-26 +3 and CT-26 +5. Both of these analytes classify well with the existing cell membrane data. This was a promising sign because the secreted protein contained in the media is a more heterogeneous mixture of proteins than the cell membrane proteins.

The second question had to do with what growth conditions were optimal during cell culture. All of the colon cell lines use Fetal Bovine Serum (FBS) as a growth substitute. The main issue with this is that albumin is a large component of FBS. Being that the proteins secreted are heterogeneous, it was a concern that the albumin would mask proteins/glycoproteins of interest in the sample. It was assumed that the albumin would be in all of the sample in a large quantity. By removing this factor the glycoproteins with aberrant glycosylation could be detected. As mentioned before, two passages of CT-26 were grown labeled as CT-26 +3 and CT-26 +5 in Figure 2.2. Both
passages were grown to 70% confluence and new media was added to the cells for 48 hours. In the case of CT-26 +3 complete media was added to the flasks of cells, meaning that FBS was included, however for CT-26 +5 media without FBS was used. In analyzing Figure 2.2, it was expected that both CT26 passages would overlap with the cell membrane array response and sit over the CT-26 data points (the green triangles) or at least over the rest of the non-metastatic data points (the purple circles). It can be seen that while both overlap reasonably well, the cells that were starved (CT-26 +5) seem to overlap better with the previous data including less scattering.

It was expected that the cells that were starved would have a consistent response with the array, as seen from the CT26 membrane bound glycoproteins. This was advantageous for a few reasons. The first of which was sample workflow which will be talked about in Section 2.5. Using albumin-free media significantly cut down on time in the sample centrifugation part of the workflow. The second reason this was advantageous was because studies in blood-based proteomic techniques generally discuss removing the albumin as an optimal strategy for unmasking proteins of interest. Albumin can tend to overwhelm the signal not allowing the SLs to detect glycosylation changes on the secreted proteins. The third reason this was advantageous was the clinical relevance of removing the albumin. Many diagnostics that are used clinically remove albumin after the blood draw as albumin is the most abundant protein in serum samples. Using serum-free media would allow the SL array to more closely mimic protocols that are clinically relevant.
2.5 SECRETED PROTEIN STUDY WORKFLOW

With the results of the pre-study in mind, a larger secreted protein study was developed. In order to provide a more clinically relevant study, only human colon cell lines were chosen. Four human colon cell lines of varying metastatic potential were used for these experiments. CCD-841-CON (ATCC® CRL-1790™)\textsuperscript{45} was the normal or healthy colon cell line used. HCT-116 (ATCC® CCL-247™)\textsuperscript{46} and HT29 (HTB-38™)\textsuperscript{47} were the cancerous non-metastatic (lowly metastatic) cell lines used. LoVo (ATCC® CCL-229™)\textsuperscript{48} was the highly metastatic cell line used derived from the left supraclavicular region (secondary site) of a male. All cell lines were grown according to ATCC® guidelines. Table 2.2 lists the cell lines used in this experiment with the associated metastatic classification. If the cell line is metastatic, it also lists the origin of the cell line (i.e. the secondary site).

Table 2.2: Four human colon cell lines used in this experiment with the associated metastatic classification.

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Classification</th>
<th>Cancer Type</th>
<th>Origin</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCD-841 CON</td>
<td>Healthy</td>
<td>N/A</td>
<td>Human</td>
<td></td>
</tr>
<tr>
<td>HCT116</td>
<td>Cancerous, Non-Metastatic</td>
<td>Colon</td>
<td>Human</td>
<td></td>
</tr>
<tr>
<td>HT29</td>
<td>Cancerous, Non-Metastatic</td>
<td>Colon</td>
<td>Human</td>
<td></td>
</tr>
<tr>
<td>LoVo</td>
<td>Metastatic</td>
<td>Colon</td>
<td>Lymph Node</td>
<td>Human</td>
</tr>
</tbody>
</table>

The pre-study used protein concentrations of 0.1 mg/mL, however, to keep this study consistent with the cell membrane data already collected a 50-fold dilution of a 0.01 mg/mL solution was used for a final concentration of 0.0002 mg/mL. The pre-study indicated that a 0.1 mg/mL solution would correlated with membrane bound protein, however, the concentration was chosen in order to stay consistent with published data.
The experiment designed used both membrane-bound proteins as well as secreted proteins. In order to keep cellular transformation between passage numbers to a minimum, the membrane bound proteins would be extracted from the cells and the secreted proteins would be taken from the media all from the same culture. While the experimental section will go into further detail, Figure 2.3 depicts the workflow used for both the secreted proteins as well as the membrane-bound proteins.

The membrane-bound proteins were included in the experiment for a few reasons. The first of which was to include the analyte that was previously used for studies. All of the previously published data, referenced in Chapter 1, used only proteins extracted from the plasma membrane. The second reason was that we would be able to see if the array would classify both the secreted and membrane-bound proteins and if the array results would correlate well to one another. If this correlation existed, the array could start being used with secreted proteins only rather than membrane-bound if using one versus the other was not losing information. Being that there may be cellular changes from passage to passage with cell lines, it was important that the secreted proteins taken from the media was from the same flask that the membrane-bound proteins were taken. This
would remove this possible variable that was present in the protein pre-study.

![Experimental workflow for secreted proteins and membrane-bound proteins and glycoproteins. This workflow was used for all four human colon cell lines.](image)

The cells were cultured to approximately 70% confluence. At this time, the media was removed and new media was added to begin the starvation process. Recall that starvation indicates that FBS was not present in the media. Antibiotics/mycotics were also removed. After 48 hours of starvation the media was pipetted out of the flask and put in a separate tube (right branch of Figure 2.3), while the cells were scraped using PBS and put into a separate tube (left branch of Figure 2.3). Trypsin was not used to lift the cells from the culture flask, as this would cleave the glycans on the cell surface.

Continuing down the left branch of the above figure, the proteins were extracted from the plasma membrane using an extraction kit and then were fluorescently labeled with FITC. Prior to incubation the beads containing the SLs were washed with PBS twice. The beads were incubated with PBS containing 1% BSA for 15 minutes. This solution was
removed and 1 mL of the protein solution at the desired concentration was added to nine tubes, each independently containing approximately 2 mg of a different SL (SL 1-9). After incubation, the beads were washed three times with PBS. Each tube of beads was imaged using a fluorescence microscope, Leica® MZ 16F, using a GFP filter set. The images are analyzed with a MATLAB® program.

The right branch of Figure 2.3 represents the experimental workflow for the secreted proteins. After removal from the flask, the media is centrifuged in order to remove cell debris. After the centrifugation, the media is concentrated to isolate the secreted proteins while minimizing the liquid volume. The secreted proteins are fluorescently labeled with FITC. Prior to incubation the beads containing the SLs were washed with PBS twice. The beads were incubated with PBS containing 1% BSA for 15 minutes. This solution was removed and 1 mL of the protein solution at the desired concentration was added to nine tubes, each independently containing approximately 2 mg of a different SL (SL 1-9). The beads were washed three times with PBS after incubation and before imaging.

Figure 2.4 represents the data acquisition and processing after incubation with the SLs. This is the same workflow in Figure 2.1. Each tube of beads was imaged using a fluorescence microscope, Leica® MZ 16F, using a GFP filter set. The images are analyzed with a MATLAB® program and process with Microsoft Excel®. Following data processing, is data analysis using various statistical methods (i.e. LDA or PCA).
2.6 SECRETED PROTEIN STUDY DATA AND DISCUSSION

The importance of this study was to investigate the usefulness of the synthetic lectin sensory array in a more clinically relevant way. This would look at the ability of the sensor array to discriminate secreted proteins isolated from media in order to mimic the secreted proteins in blood or urine. The sensor array has already been shown to discriminate cell lines, seen in Chapter 1. Similar cell lines would be used with the addition of a human colon cell line for a healthy/normal cell line. The discriminatory power of the sensor array would be investigated for cell lines again as well as the secreted proteins. Table 2.3 depicts the cell lines used as well as the rep number and total number of data points for each sample.
Table 2.3: Table Depicts all cell lines used in this experiment and the associated biological replicate number. It includes the type of protein as well as the number of data points for each cell line. Recall each technical replicate (N=1) is the response of nine synthetic lectins with a particular analyte.

<table>
<thead>
<tr>
<th>Total</th>
<th>Cell Line</th>
<th>Rep</th>
<th>Type of Protein</th>
<th>N=</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCD-841 CON</td>
<td>1</td>
<td>Secreted</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>CCD-841 CON</td>
<td>2</td>
<td>Secreted</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>HCT 116</td>
<td>1</td>
<td>Secreted</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>HCT 116</td>
<td>2</td>
<td>Secreted</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>HT29</td>
<td>1</td>
<td>Secreted</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>HT29</td>
<td>2</td>
<td>Secreted</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>LoVo</td>
<td>1</td>
<td>Secreted</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>LoVo</td>
<td>2</td>
<td>Secreted</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>Secreted</td>
<td></td>
<td></td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>CCD-841 CON</td>
<td>2</td>
<td>Membrane</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>HCT 116</td>
<td>2</td>
<td>Membrane</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>HT29</td>
<td>2</td>
<td>Membrane</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>LoVo</td>
<td>2</td>
<td>Membrane</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>Membrane</td>
<td></td>
<td></td>
<td>69</td>
</tr>
</tbody>
</table>

There were two biological replicates done for each cell line for the secreted protein type, while there was one replicate done for each of the cell lines for the membrane bound proteins. A biological replicate was one defined as at least two passages (cells split) and allowed to grow before being starved and scraped. It is ideal to thaw two different cell stocks if possible. Each cell line for the membrane bound proteins is listed as replicate 2 because the membrane bound proteins were taken from the cells that the media for replicate 2 was taken. In other words CCD-841 CON rep 2 secreted proteins was the media taken from the flask of CCD-841 CON rep 2 cells. Initially the
experiment was designed to have two biological replicates for the membrane bound proteins as well, however, the entire membrane extraction kit was ineffective in producing any protein in the end. After this experiment, it was decided not to re-grow cell lines as this is expensive and the utility of the sensor array with membrane proteins had already been established, as referenced in Chapter 1.

All protein samples both secreted and membrane bound were dealt with as depicted in Figure 2.3 fluorescently labeled and incubated individually with SLs 1-9. The beads were then imaged using fluorescent microscopy in order to obtain bright numbers. All of the data seen within this chapter is using normalized bright numbers. Figure 2.5 is the Linear Discriminant Analysis (LDA) plot generated from all of the data from secreted and membrane proteins listed in Table 2.3.

![Three class LDA plot generated with normalized bright numbers from the SL array generated by incubating fluorescently labeled secreted and membrane proteins. The leave one out classification accuracy was 78% and the N=212. The three classes used were healthy, cancerous non-metastatic, and cancerous metastatic.](image)

Figure 2.5: Three class LDA plot generated with normalized bright numbers from the SL array generated by incubating fluorescently labeled secreted and membrane proteins. The leave one out classification accuracy was 78% and the N=212. The three classes used were healthy, cancerous non-metastatic, and cancerous metastatic.
Each data point represented in this plot is the sensor array’s response (SL1-9) with that particular analyte. The discriminant analysis was run in Systat®, while the graph was generated using Statistica®. LDA was used because the cell lines have well defined metastatic potential; therefore using a guided statistical method seemed appropriate. The analysis used a three-class classification: healthy/normal, cancerous non-metastatic (lowly), and cancerous metastatic. The green dots represent all of the healthy samples (CCD-841-CON), with the filled in denoting secreted protein and the not filled in representing the membrane-bound samples. The orange triangles represent all of the cancerous non/lowly metastatic samples (HT29 and HCT116), with the filled in denoting secreted protein and the not filled in representing the membrane-bound samples. The red diamonds represent all of the cancerous metastatic samples (LoVo), with the filled in denoting secreted protein and the not filled in representing the membrane-bound samples. The N for this particular analysis was N=212, representing the technical replicates within the biological replicates. The ellipses surrounding the data points represent a 95% confidence interval for that particular data set. The leave-one-out classification accuracy for this analysis was 78%.

Ideally, it was hoped that the leave-one-out classification accuracy would be higher as was referenced in Chapter 1 in dealing with the membrane-bound proteins. Furthermore, it was initially hypothesized that in looking at the LDA plot that the secreted and membrane would overlap nicely. While there is some overlap between the respective secreted and membrane protein of a particular metastatic potential, there is also some overlap of differing metastatic potentials leading to lower classification
accuracy. However, thinking through the experiment and reviewing fluorescent intensities can explain this result.

There are a few reasons why the LDA plots do not overlap above. LDA attempts to group things that are alike together, but this is based up the SL array response, that is the response of SL 1-9 with one particular analyte. Figure 2.6 depicts the fingerprint pattern from SLs 1-9 for HCT 116 membrane-bound proteins and secreted proteins. The membrane bound proteins and secreted proteins were taken from the same biological replicate. This plot is generated from normalized bright numbers.

![Figure 2.6: Fingerprint patterns for the membrane bound proteins and secreted proteins from the same HCT 116 biological replicate. As mentioned, in text, this was generated using normalized bright numbers.](image)

As can be seen from the figure, the patterns are slightly different. This is just an example from one of the cell lines, but is a trend seen among most of the cell lines. Experimentally, the secreted proteins and membrane proteins require slightly different parameters. Typically, when imaging the beads incubated with fluorescently labeled secreted proteins the exposure times used are longer than when imaging the beads incubated with fluorescently labeled membrane bound proteins. The concentration of all
of the fluorescently labeled proteins incubated with the synthetic lectins were the same according to a BCA assay, however, there is a difference in the heterogeneity of the samples. The membrane extraction kit used binds to plasma membrane proteins so that much of the sample is only plasma membrane proteins. The secreted proteins are any proteins that are secreted off of the cells while in culture. The proteins kept after the concentration of the media is above a 3kDa cutoff (the cutoff of the tubes used). The latter approach provides a sample that is significantly more heterogeneous. While all of the protein is labeled and quantified using a BCA assay perhaps only a portion of this protein would interact with the synthetic lectin. Previous and current research in the Lavigne lab has demonstrated that glycoproteins interact through the boronic acid-diol interaction and some proteins have a protein/protein interaction with the synthetic lectin backbone. With this in mind, it was advantageous to look at the LDA plots from both the membrane bound proteins (Figure 2.7) and secreted proteins (Figure 2.8) with the synthetic lectin sensor array separately. Figure 2.7 depicts the LDA plot of the sensor array’s response with only membrane proteins from the four human colon cell lines. Recall there are two non-metastatic cell lines (HCT 116 and HT29) included in the non-metastatic points (orange triangles).
Figure 2.7: Three class LDA plot generated with normalized bright numbers from the SL array generated by incubating fluorescently labeled membrane proteins. The leave one out classification accuracy was 97% and the N=69. The three classes used were healthy, cancerous non-metastatic, and cancerous metastatic.

Again, the discriminant analysis was run in Systat®, while the graph was generated using Statistica®. This analysis also used a three-class classification: healthy/normal, cancerous non-metastatic (lowly), and cancerous metastatic. The green dots represent all of the healthy samples (CCD-841-CON). The orange triangles represent all of the cancerous non/lowly metastatic samples (HT29 and HCT116). The red diamonds represent all of the cancerous metastatic samples (LoVo). The N for this particular analysis was N=69, representing the technical replicates. There is only one biological replicate in each of the cell lines. The ellipses surrounding the data points represent a 95% confidence interval for that particular data set. The leave-one-out classification accuracy for this analysis was 97%.

The leave-one-out classification accuracy is significantly higher than the 78% that resulted from running the LDA analysis with both the secreted and membrane-bound
proteins. It can be argued that there is increased classification accuracy because there is only one biological replicate, however it has been shown that the SL sensor array can discriminate membrane bound proteins/glycoproteins. There are two non-metastatic cell lines which cluster together (orange triangles) which indicates that while two different cell lines they SL sensor array data will group these by metastatic potential. Both of these results were expected from the data mentioned in Chapter 1.

Linear discriminant analysis was then run on only the sensor array’s response to the secreted protein from the four different cell lines. The data set contains both biological replicates from each of the four human colon cell lines.

![Secreted Proteins for 4 Cell Lines (3 class)](image)

Figure 2.8: Three-class LDA plot generated with normalized bright numbers from the SL array generated by incubating fluorescently labeled membrane proteins. The leave one out classification accuracy was 88% and the N=143. The three classes used were healthy, cancerous non-metastatic, and cancerous metastatic.

The discriminant analysis was run in Systat®, while the graph was generated using Statistica®. This analysis again used a three-class classification: healthy/normal,
cancerous non-metastatic (lowly), and cancerous metastatic. The green dots represent all of the healthy samples (CCD-841-CON). The orange triangles represent all of the cancerous non/lowly metastatic samples (HT29 and HCT116). The red diamonds represent all of the cancerous metastatic samples (LoVo). The N for this particular analysis was N=143, representing the technical replicates. There are two biological replicates in each of the cell lines. The ellipses surrounding the data points represent a 95% confidence interval for that particular data set. The leave-one-out classification accuracy for this analysis was 88%. The leave-one-out classification accuracy was very good for the SL array response with the secreted proteins. This was a promising result for the clinical significance of the sensor array.

All of the analyses were run with metastatic potential in mind, i.e., healthy/normal, cancerous non-metastatic (lowly), and cancerous metastatic. However, in use as a diagnostic the most important question to answer is “Does the patient have cancer?” If the answer to this is yes then there are a variety of imaging options that can be done in a follow-up setting, however that initial question would be answered. It can also be argued that for a diagnostic it would not be idea to model secreted and membrane bound proteins in the same model. In thinking about this question, the linear discriminant analysis was run on the secreted and membrane protein’s SL array response (Figure 2.5), however, this time using a two class method: healthy and cancerous.

Looking at the three-class in Figure 2.5, it can be seen that the major source of overlap is between the cancerous non-metastatic and cancerous metastatic groups (both membrane-bound and secreted protein). This overlap includes all of the orange data points (cancerous non-metastatic) and all of the red data points (cancerous metastatic).
Recall the leave-one-out classification accuracy for the three-class analysis (healthy, cancerous non-metastatic, and cancerous metastatic) was 78%. If the analysis is re-run is a two-class analysis (healthy and cancerous) the classification accuracy improves to 90%. A graph cannot be drawn when performing a two-class analysis. Table 2.4 will summarize the linear discriminant analyses that were performed.

Table 2.4: Summary of linear discriminant analyses performed and the associated leave-one-out classification accuracies. The number of technical replicates is also listed. Recall, every membrane-bound protein had one biological replicate, while, all secreted protein samples had two biological replicates.

<table>
<thead>
<tr>
<th>Secreted Protein and Membrane Protein</th>
<th>Class Discriminant Analysis</th>
<th>Class Identification</th>
<th>N=</th>
<th>Leave-one-out classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three class</td>
<td>Healthy Cancerous Non-met. Cancerous Met.</td>
<td>212</td>
<td>78%</td>
<td></td>
</tr>
<tr>
<td>Secreted and Membrane Protein</td>
<td>Two class</td>
<td>Healthy Cancerous</td>
<td>212</td>
<td>90%</td>
</tr>
<tr>
<td>Three class</td>
<td>Healthy Cancerous Non-met. Cancerous Met.</td>
<td>143</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td>Secreted Protein</td>
<td>Three class</td>
<td>Healthy Cancerous Non-met. Cancerous Met.</td>
<td>69</td>
<td>97%</td>
</tr>
</tbody>
</table>

While it is interesting to look at the classification accuracies of membrane protein and secreted protein, an ideal diagnostic would be fluid-based to avoid an invasive procedure to obtain cells.

2.7 CONCLUSION

The initial sections of this chapter described the need for a fluid-based colon cancer diagnostic. The synthetic lectin sensor array, up until this point, had only been used with membrane-bound proteins and glycoproteins. This would not serve well in a clinical setting as obtaining cells from a tumor or polyp is as invasive as a colonoscopy.
and would be associated with similar disadvantages. With this in mind, optimizing the synthetic lectin sensor array for secreted proteins was essential.

A proof-of-concept experiment was necessary to evaluate the whether the sensor array could discriminate secreted proteins, which would be used in a fluid-based diagnostic. The experiment was optimized to use cell culture with four established colon cell lines of differing metastatic potentials varying from healthy to cancerous metastatic. The media was taken and concentrated for secreted proteins, while the cells were isolated to obtain the membrane-bound proteins and glycoproteins. The nine synthetic lectins were incubated with each fluorescently labeled analyte and imaged using fluorescence microscopy and camera.

Linear discriminant analysis was used to analyze the data. The secreted proteins and membrane bound proteins modeled together had a leave one out classification accuracy of 78%. This was not completely unexpected as the secreted proteins are more heterogeneous and experimentally the SL array patterns are slightly different. The membrane bound proteins had a leave-one-out classification accuracy of 97% which was expected as former members of the Lavigne lab had determined the array was successful in discriminating cell lines of varying metastatic potential using membrane-bound proteins/glycoproteins. The secreted proteins, when analyzed with linear discriminant analysis, had a leave-one-out classification accuracy of 88%. This indicates that the synthetic lectin array was able to classify secreted proteins based on metastatic potential. The next section will describe the future experiments that could be done with the synthetic lectin array and clinical serum samples.
2.8 FUTURE WORK

This proof-of-concept experiment shows great promise with the discrimination that was possible of the secreted proteins as well as the membrane-bound proteins. To further the project, it is important to use the SL array with clinically relevant samples. The next step for this project would be to get clinical blood samples from colon cancer patients.

The ideal route for the next step was to get serum samples from patients of varying stages of colon cancer. Whole blood samples or serum samples could be used. Using serum samples would simplify the experimental procedure, however, if other experiments would require an analyte not present in serum then whole blood should be used. There are a few companies that can give a serum sample, tissue sample from a tumor, and matched healthy portion of cells. This would allow the investigation of whether serum samples or tumor samples provide a more accurate assessment of staging. Calculations should be done to determine a minimum number of samples needed to correlate the SL array to metastatic potential. The lab already has healthy blood samples from volunteers in the lab as well as the Chemistry department at the University of South Carolina. To mimic a clinically relevant diagnostic, albumin would be removed from the human serum samples as well. Chapter 3 includes work done with mouse serum samples. It can be argued that a very similar protocol would be used for the human serum samples.

2.9 EXPERIMENTAL METHODS

Cell Culture

Four human colon cell lines of varying metastatic potential were used for these experiments. CCD-841-CON (ATCC® CRL-1790™) was the normal colon cell line used.
HCT-116 (ATCC® CCL-247™) and HT29 (ATCC® HTB-38™) were the cancerous lowly metastatic cell lines used. LoVo (ATCC® CCL-229™) was the highly metastatic cell line used. This cell line was derived from the left supraclavicular region of a male. All cell lines were grown according to ATCC® guidelines.

**Media Used**

HCT 116 and LoVo used a media of RPMI-1640 (purchased from HyClone SH30605.01, 10% FBS (50mL) purchased from VWR (45000-734) and 1% (5mL) antibiotics/antimycotics purchased from VWR (45000-616). HT29 used a media made up of Dulbecco’s Modified Eagle Medium (DMEM) purchased from VWR (45000-734), 10% FBS (50mL) purchased from VWR (45000-734) and 1% (5mL) antibiotics/antimycotics purchased from VWR (45000-616). CCD-841-CON cells used Minimum Essential Media (MEM), 10% FBS (50mL) purchased from VWR (45000-734) and 1% (5mL) antibiotics/antimycotics purchased from VWR (45000-616). The stockroom stocks some of this media.

Remove 55mL of media from the bottle and place in a centrifuge tube. This can be used this later for starving cells if needed. Add the FBS and antibiotics/antimycotics in with a pipette rather than pouring the solution into the bottle. After everything has been added to the media use a larger pipette and pipette up and down in order to thoroughly mix the media. The previous steps can be performed with the media cold if it is not being used for anything after.

**Thawing Stocks**

Typically, the initial thaw of a stock will want to be done in a T25 flask. Once you know that cell viability is high for a cell stock, you can thaw into a T75 flask. If you
notice the T25 is confluent fast, the next time you can thaw into a T75. Pre-warm your full media (all additives added) to 37°C. Make sure you have everything you need to thaw the cells in your hood ready to go, as DMSO is toxic to cells so they need to be thawed and thawed in a larger volume of media.

Place about 5mL of warm media into a 15mL Falcon tube. Pipette small amounts of warm media into the cryogenic vial. You will need to pipette a small amount in and then pipette up and down. As soon as some of the cell stock thaws place it into the falcon tube with the other warm media to dilute the DMSO. Keep adding the warm media and removing the DMSO/cell solution to media until all of the stock has thawed. The following day remove the media and add fresh media. Be careful to not do this too soon after. Dead cells and live cells look different yet they both float (dead cells will look like broken dark pieces where live unattached cells will look like nice healthy spheres). I would usually split my cells 2-3 times before running any experiments to make sure these were healthy and viable.

*Feeding the Cells*

Feeding the cells requires the pipetting of old media off before putting new media on. I would suggest changing the media every couple of days when not splitting the cells. The pink media will begin to turn orange when the cells are either too confluent or they media needs to be changed (you should not let it get to this point). The media should be warm (37°C) for this step.

*Splitting Cells*

The trypsin/EDTA media should be warmed in incubator to 37°C before beginning the process of splitting the cells. The cells should be at 70-75% confluence.
prior to splitting. Take the existing media off of the cells and put into the waste container. The same pipette may be used as long as it does not touch the outside of the flask or any part of the hood. Wash the back of the flask with 1-2 mL of the trypsin/EDTA solution then pipette this off and put into waste. This will remove the dead cells and excess media that may not have come off with the first media removal. Then (with a fresh pipette) add trypsin/EDTA to the flask. The amount needed will depend on the flask size, but you want enough of the solution to cover the bottom of the flask when horizontal. This will ensure that all cells will have the solution over them. Place the flask back into the incubator (37°C).

The amount of time necessary to trypsinize the cells will depend on the cell line. In the beginning, check the cells every 1-2 minutes. You can either hold the flask up to the light or look under the microscope to see if the cells are attached. If you know that your cell line is hearty you can rub the back of the flask to speed up this process or in extreme cases hit the flask against your hand. The cells should be floating when the have completely detached. Be mindful not to leave too many cells behind, as this will deplete your cell number.

Wash the cells that have detached to the bottom of the flask with fresh warm media. This amount will depend on the size flask used. The flask may need to be rinsed a couple of times to ensure all of the cells have been removed. Add this solution to a 15 mL centrifuge tube and centrifuge for 4-5 minutes at 500 x g. Be sure to not spin at too high of a rate as this will damage the cells. Remove the supernatant and put in the waste container. Add 1-2 mL of fresh media to the cell pellet. Use a small pipette (p1000) and pipette this solution up and down. This will ensure there are not clumps of cells when
adding to a new flask. Add the appropriate amount of microliters to each flask depending on the split you want and how many flasks you need. If there are extra cells these can be placed in the waste container. Add the media to the flask to make it to the total volume for each flask type (5 mL for T25, 10 mL for T75, and 20 mL for T125. The following day check to make sure the cells have attached and there are not too many floating cells. You can change the media if you would like using warmed media.

*Making Stocks*

Only make stocks from cells that are actively growing and no more than 75-80% confluent. If you are unsure, split the cells again and makes stocks from the next passage. One T75 flask should result in 3-5 stocks, however, adjust this number depending on how many cells you want in your stock. I tend to put more in so that it is a better chance of recovery. Cell stocks are frozen in cryovials that can be found in the stockroom. Label each vial with you initials, cell line, and cell passage.

Make sure everything is ready in the hood prior to these steps. DMSO is toxic to cells so speed is required for this step. Have a cooled solution of 20% DMSO in FBS (this should have been filtered through a .2 µm filter. Trypsinize the cells according to the instructions above, inactivate the trypsin with media, and put into a centrifuge tube. Centrifuge the cell suspension for 4-5 minutes at 500 x g and remove the supernatant. The following steps you will want to use a p1000 pipette. Re-suspend the cell pellet in the usual full media. Add the amount you will need to have is .5 mL * the number of stocks you are making. Be sure to pipette the solution up and down in order to equally distribute the cells into all of the vials. Add .5 mL of the cell suspension to each cryovial and then add .5 mL of the 20% DMSO-FBS solution.
As soon as the DMSO is added the cells they must be frozen immediately. The best way to do this is use the Mr. Frosty container and place in the freezer. This will allow the temperature to reduce gradually. The Mr. Frosty contains isopropanol that needs to be changed every five uses. You can mark the top of the container with a marker in order to indicate which use it was.

*Starving the Cells*

When the cells reached no more than 70% confluence they were starved for 48 hours. The media did not include any supplements. This was due to some cell lines dying when antibiotics were present. For these first experiments, while we were unsure about the amount of media used ten T-75 or five T-125 flasks were used. This would give about 100 mL of media. In my notebook after the BCA it is seen that this gives back a tremendous amount of protein, much more than was needed for these or for repeat experiments. As little as 5 mL of media has been used. If the experiment requires cells in addition to the media the number of flasks needed is simply the number needed for cells. The cell pellet will be the limiting factor.

*Concentration of Media for Secreted Proteins*

After 48 hours the media was taken off of the cells and centrifuged to remove cell debris. The media was pipetted off and put in a new tube. The media was centrifuged in Amicon® 3KDa centrifuge tubes (EMD Millipore). This will be the longest step in the process. Only 10-15 mL will be spun through every twenty minutes. Keep this in mind when deciding on how many flasks to use. When the media is mostly spun through and there is only a small amount in the bottom of the filter (be sure not to let the filter go dry) it is a good idea to rinse PBS through in order to get rid of all of the media. If colored
media is being used when the media is no longer colored and is a true clear you have finished the wash step. If the media is colorless it will take a little more intuition to figure this out. There should be a small amount of liquid in the filter tube (around 1500 μL). Pipette this out with a p200 pipette and tip (yellow). Due to the filter size you want to make sure the pipette tip reaches the bottom because this is where the protein is. After pipetting the solution out in a 15 mL centrifuge tube add PBS and rinse the filter pipetting up and down multiple times.

Acetone Precipitation of Secreted Proteins from Media

The protein is then precipitated with ice-cold acetone stored in the -30 degrees Celsius freezer. Add four times the amount of acetone to the protein solution. For example, if there is 2 mL of protein 8 mL of acetone would be added. Shake the tube in order to see a white fluffy precipitate. At this point, the protein solution can be stored in the -30 degrees Celsius freezer overnight if you wish or for a minimum of an hour. The following day (or after some time) take the sample out of the freezer, spin it down, and remove the acetone. Allow the centrifuge tube to sit in the hood for 30 minutes to an hour to make sure the acetone is completely evaporated. At this point, the protein can be treated like any other protein sample. Usually (if moving on to the labeling step), the protein is re-suspended in carbonate buffer using only the amount necessary to completely dissolve (starting with about 1 mL each time). This will depend on the amount of protein recovered.

Membrane Proteins and Glycoproteins

The cells were scraped off the flask using a cell scraper. We use a scraper in this step and not the trypsin because trypsin would cut the surface glycan chains off. It is
good practice to put about 1 mL into the flask and scrape with cell scraper. Use pipette this solution off and rinse with a second mL of PBS. Scrape again and rinse the flask with PBS and the solution was centrifuged. Hold the flask up to the light to make sure there are not a lot of cells left behind. The PBS was taken off to leave behind a cell pellet. This can be stored in the -30 degree Celsius freezer for a bit of time or in the -80 degree Celsius freezer for an extended period of time. The plasma glycoproteins and proteins were extracted using the Qiagen® Plasma Membrane Protein Kit. The protein at the end is in elution buffer and must be spun into carbonate buffer purified using Amicon® 10K centrifuge (spin) tubes (EMD Millipore). The solution was put into the spin tube and spun for about 8 minutes at 4000g. The time spun may change depending on protein concentration, but you want to spin until it is at the “2” line on the tube. When the solution is at the “2” line, 1.5 mL of carbonate buffer is put in the exchange tube and spun for 5 minutes. Again it is checked to make sure the solution is at the “2” line. At this point, you want to turn the small tube inside over and spin out into a micro centrifuge tube in a bench-top centrifuge. Put in 300 µL of carbonate buffer and pipette up and down. It is advantageous to use a yellow tip again for this step in order to get all the way down in the filter. Turn tube over and spin again. Rinse tube with 2-100 µL increments of carbonate buffer.

Labeling of Membrane and Secreted Proteins/Glycoproteins

The membrane proteins/glycoproteins and secreted proteins were labeled separately with fluorescein isothiocyanate (FITC) in DMF and carbonate buffer. 2.8 mg of FITC was dissolved in 100 µL of DMF. 10 µL of the FITC solution was put into 200 µL of protein. The total volume of the solution was brought up to 600 µL by adding 400 µL
of carbonate buffer. The protein solution was tumbled for 1 hr at 37 degrees Celsius and purified using Amicon® 10K centrifuge (spin) tubes (EMD Millipore). The solution was put into the spin tube and spun for about 8 minutes at 4000g. The time spun may change depending on protein concentration, but you want to spin until it is at the “2” line on the tube. When the solution is at the “2” line, 1.5 mL of PBS is put in the exchange tube and spun for 5 minutes. Again it is checked to make sure the solution is at the “2” line. At this point, you want to turn the small tube inside over and spin out into a micro centrifuge tube in a benchtop centrifuge. Put in 300 µL of PBS and pipette up and down. It is advantageous to use a yellow tip again for this step in order to get all the way down in the filter. Turn tube over and spin again. If you are using secreted proteins you can repeat the rinse step since there is a large concentration of protein. If you are using membrane extracts you would want to be aware that there is not as big of a concentration and thus maybe rinse with 2-100 µL increments of PBS. After labeling a BCA assay was done to quantify the amount of protein.

**BCA Assay**

Thermo Scientific BCA kit was purchased from VWR.

**Sample Preparation**

If using secreted protein sample you may need to dilute your samples and run these with the concentrated sample to ensure the concentration falls within the range of the assay. If this BCA assay is done before labeling, the samples will be in carbonate buffer, thus your standards should be made in carbonate buffer. If this BCA is done after the labeling steps, the samples will be in PBS and thus your standards should be made in PBS.
Standard Preparation

There is a standard 2 mg/mL solution included with the kit. Put 100 µL of this solution into a small plastic tube. Put 50 µL of carbonate buffer in several other tubes. Then take 50 µL of the 2 mg/mL solution in the first tube. Be sure to vortex this several times (This is the first 1:2 dilution). Then take 50 µL of that solution and put it in the next tube. Continue this process. Make sure you also have a blank standard (either carbonate buffer or PBS depending on what step you are doing this).

Reagent Preparation

The directions instruct you to make the reagent to be made 50:1. 200 µL of reagent is required for each well so the amount of total reagent will need to be calculated (200 µL x # of wells). This can be done in a 15 mL centrifuge tube.

Well Preparation

Make sure the standards/samples are done in triplicate. Put 10 µL in each well of standards and samples. Then put 200 µL of reagent in all of the wells. Make sure to then mix each well with a pipette tip. The plate should then be placed in the oven for 30-45 minutes at 37 degrees Celsius. Periodically check your plate to see if the standards have changed. The plate is then read on the plate reader at 562 nm. Be sure not to leave your plate for a long period of time. It should be noted that this assay is not an endpoint assay and will continue to develop.

Synthesis of Synthetic Lectins

SLs are attached to Tentagel beads. The beads are weighed out and put in 10mL of DMF(dimethylformamide) for ten minutes in order to swell the beads. Meanwhile, Fmoc-Met-OH(10eq) is activated with HBTU(10eq) for 10min in 5% N-
methylmorpholine in DMF. The amino acid solution is added and tumbled for 1 hour. The solution is drained and the beads are washed with DMF, MeOH(methanol), DMF, MeOH. After the last washing, a ninhydrin test is performed on one or two beads to see if the amino acid has coupled. If coupled, the Fmoc group is removed using 10mL of 20% piperidine in DMF. That process is repeated manually until the amino acids are on or a peptide synthesizer can be used. The SL is then N-terminally acetylated using 10mL of CH₂Cl₂ containing 5% pyridine and 5% acetic anhydride.

**Removal of protecting group on DAB (diamino acid)**

Diamino acid (DAB) has an ivDde protecting group on it. In order to put on the boronic acids, these must be removed. The protecting group is removed by adding 10mL of 2% hydrazine monohydrate in DMF and tumbling for 1 hour.

**Addition of Phenyl Boronic Acids (PBA)**

2-formylphenyl boronic acid (8eq) is dissolved in 10% MeOH in DMF. Activated 3Å molecular sieves and the PBA solution are added to the beads and tumbled overnight at 37 degrees Celsius. Sodium borohydride (NaBH₄) (8eq) is added and gas was released over 30 min. The beads then tumble for 4 hours. The beads are washed with DMF and MeOH and the sieves are removed. The beads are then washed again.

**Removal of Acid Labile Groups**

In order to remove the acid labile groups the beads are tumbled with 10mL of 95% trifluoroacetic acid(TFA), 2.5% water, and 2.5% triisopropylsilane for 1 hours. After tumbling, the solution is taken off and washed with DMF, MeOH twice. The synthetic lectins can then be washed with PBS three times in order to cut down on PBS
wash time done before incubation. After washing with PBS it is important that the tubes of SLs be placed in the refrigerator and not the -30 degrees Celsius freezer.

**MALDI Sample Preparation**

**Phenyl Boronic Acid (PBA) Group Removal**

This step can be avoided if a small amount of beads are taken out before the boronic acids are added. Once removed, the rest of the SL synthesis steps are followed on the small amount of beads. However, if the boronic acids are still attached and a MALDI is needed to be run these groups will need to be removed.

Remove PBA groups by adding 100 mmol solution of \( H_2O_2 \) to the 1.5 mL centrifuge tube containing the beads. Put this tube at 50 degrees Celsius for 1 hour. After the hour, wash the beads with \( dH_2O \) three times. Add 50 \( \mu L \) of a 40 mg/mL solution of Cyanogen Bromide (CNBr) made in .1M HCl. Make sure all work with CNBr is done in the hood. Place this solution in the dark (usually your drawer) for 15-18 hours. Evaporate this solution using the speed vacuum centrifuge. There should be a small amount of white substance at the bottom of the centrifuge tube.

**MS Sample**

Dissolve the sample obtained from the previous section in 10 \( \mu L \) of .1% TFA in \( H_2O \). In order to make the matrix, make a saturated solution of \( \alpha \)-cyano-4-hydroxy cinnamic acid (CHCA) in 50:50 acetonitrile/.1 % TFA in \( H_2O \). Add 1 \( \mu L \) of sample to the desired spot on the plate for MALDI. Add 1 \( \mu L \) of matrix to the same spot. Pipette the 2 \( \mu L \) of solution up and down a few times to get a good mix of matrix and sample. Let the plate dry and take obtain MALDI spectra.
**Incubation with Synthetic Lectins (SLs)**

Each synthetic lectin is weighed into a centrifuge tube (2mg) for the binding study. The beads were washed with PBS twice and PBS with 1% bovine serum albumin (BSA) for 15 minutes. This solution was removed and 1mL of .0002 mg/mL (50 fold dilution of a .1 mg/mL solution) of a particular protein is put in the tube and allowed to incubate overnight at room temperature. The solution was removed and then washed with PBS three times. After the last wash, the beads are re-suspended in about 500 µL of PBS to make the exchange to the microscope slide easier.

**Imaging the Beads**

The beads are imaged with a fluorescence microscope, Leica MZ 16F, using a GFP filter set. It is imperative that all of the SLs in each cell line (in this case SL 1-9) were imaged at the same exposure time. In order to do this, place each slide under the microscope and find the brightest SL. Once this is found for a particular sample, optimize this slide to have the largest green value without saturating the camera (no more than 190 to be safe, 255 is the absolute max). This number can be found by circling a bead in Adobe Photoshop® and pulling up the histogram information. After finding the appropriate exposure time, pictures can be taken with care that most beads are within the Region of Interest (ROI) and not touching one another, as these beads would be thrown out.

**Data Analysis**

After the pictures have been taken they must be converted into JPG or JPEG as the camera takes these pictures as TIFF files. There are protocols written in the lab to do this on a MAC or PC. After conversion into jpeg, the files can then be run through a code in
MATLAB® written by Dr. Greytak and modified by Anna Veldkamp. The code used for these experiments was image analysis anna. There is a protocol written in the lab that walks a user through the necessary steps to run the code. This code essentially picks out each bead and provides information about area, bright number, red number, green number, and blue number. It also gives information about where this bead was located in the picture so that you can identify a number for each bead. The data is processed in excel and outliers are removed (1.5 IQD) in order to get the array information for the SLs with each protein type. At this point, the data can be put into varying statistical models, the most common seen in this chapter are linear discriminant analysis (LDA), principal component analysis (PCA), and machine learning which is an extension of linear discriminant analysis (LDA).
CHAPTER 3: INVESTIGATING METASTATIC POTENTIAL WITH AN ISOGENIC CELL LINE

3.0 ABSTRACT

The utility of the synthetic lectin array has been shown for existing cell lines in both chapters 1 and 2, however, questions arise about the utility of the sensor array in a clinical setting. The cell lines used in the previous chapters had an established metastatic potential, meaning that these cell lines were either classified as healthy/normal, cancerous non-metastatic (lowly metastatic), or cancerous metastatic. In a clinical setting a patient may be undergoing metastatic changes as cancer progresses. It would be helpful to know the usefulness of the array in terms of diagnosing metastatic potential as it is changing. The first part of this chapter will discuss the metastatic reversion of an isogenic cell line, while the second part of this chapter will discuss the synthetic lectin array as a diagnostic in an *in vivo* longitudinal mouse study.

3.1 INTRODUCTION

Previously, the synthetic lectin array has been investigated for use in diagnosing colon cancer using cell lines. This was done using membrane bound proteins, referenced in Chapter 1, and secreted proteins, referenced in Chapter 2. This chapter will focus using the synthetic lectin array as a tool in investigating metastatic potential. This chapter will be split up into two parts. The first part will discuss investigating the reversion, in metastatic potential, of an isogenic cell line. The second part will discuss the use of the same isogenic cell line in a longitudinal mouse study.
3.2 PART 1: INVESTIGATING REVERSION OF METASTATIC POTENTIAL

Dr. Peña’s lab modified an existing cell line, CT26, \textit{in vivo}. CT26 is a lowly metastatic cell line with a metastatic rate of approximately 0-5%. CT26 cells were injected into a mouse cecum and some metastatic cells migrated to the liver, as is the case with most colon cancer. These cells were taken from the tumor sites upon the sacrificing of the mice and cultured. The new cells were then cultured and injected into the cecum of new mice. This process was repeated for three rounds in order to obtain what was called CT26-FL3, a highly metastatic cell line.

![Image](image.png)

Figure 3.1: A) Picture of the internal organs of a mouse. The cecum is highlighted which is where the cells were injected. B) Pictorial representation of the experimental procedure Dr. Pena’s lab used in order to make CT26-FL3 cells.

Figure 3.1a shows the internal organs of a mouse, noting that the cecum is where the cells were injected. Figure 3.1b depicts the process of in-vivo selection, which was how the Pena lab created a highly metastatic cell line, CT26-FL3, from a non-metastatic parent
line CT26. Dr. Peña’s group noticed, through invasion assays, that after 3-5 passages CT26-FL3 began to revert back to the metastatic potential of the parent cell line, CT26.

Table 3.1: Comparison of CT26 and CT26-FL3 cell lines. This table contains metastatic potential information for both cell lines, as well as metastatic potential information after 3-5 passages.

<table>
<thead>
<tr>
<th></th>
<th>CT-26</th>
<th>CT26-FL3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell Line</strong></td>
<td>Parent Cell Line</td>
<td>Modified Cell Line</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>Colon Cancer</td>
<td>Colon Cancer</td>
</tr>
<tr>
<td><strong>Metastatic Potential</strong></td>
<td>Non-metastatic 0-5%</td>
<td>Highly metastatic Approx. 95%</td>
</tr>
<tr>
<td><strong>Metastatic Potential after 3-5 passages</strong></td>
<td>0-5%</td>
<td>Reverts back to CT26 (0-5%)</td>
</tr>
</tbody>
</table>

Table 3.1 compares CT26 and CT26-FL3 in terms of metastatic potential prior to cell culture passages, as well as metastatic potential after 3-5 cell culture passages. As discussed in Chapter 1 glycans have a role in metastasis, particularly sLe\(^x\) has a role in the intravasation and extravasation of the cells from a primary tumor site to a secondary tumor site.\(^{21}\) With this in mind, the questions that were hoping to be answered were: does this reversion have something to do with glycosylation? and Can the metastatic reversion be probed with the synthetic lectin array?

3.3 PART 1: EXPERIMENTAL WORKFLOW

While the experimental section of this chapter will go into further detail about the specifics in terms of sample workup, it is important to highlight the general method for probing the above-mentioned questions. The same workflow that was used in Chapter 2 was employed in these experiments and is depicted in Figure 3.2.
These cells retrieved from the mouse were identified as CT26 +1, where the +# indicates the passage number. While the experimental workflow depicted above was followed, in the beginning of these experiments (the early passages) there were not enough cells in order to do a membrane extraction. When possible, a membrane extraction was done on the cells, however for this reason the chapter will mainly discuss the use of secreted proteins in the experimental as well as in the data analysis sections. As in Chapter 3, the SLs used for the experiments in this part were SLs 1-9. For convenience the sequences are listed again in this chapter in Table 3.2.
Table 3.2: Sequences for SLs 1-9.

<table>
<thead>
<tr>
<th>SL ID</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL 1</td>
<td>Ac-RGD<em>VTFD</em>RBRRM</td>
</tr>
<tr>
<td>SL 2</td>
<td>Ac-RTD<em>RFLD</em>VBRRM</td>
</tr>
<tr>
<td>SL 3</td>
<td>Ac-RSD<em>VTDD</em>RBRRM</td>
</tr>
<tr>
<td>SL 4</td>
<td>Ac-RRD<em>TQTD</em>QBRRM</td>
</tr>
<tr>
<td>SL 5</td>
<td>Ac-RAD<em>TRVD</em>VBRRM</td>
</tr>
<tr>
<td>SL 6</td>
<td>Ac-RTD<em>NRND</em>FBRRM</td>
</tr>
<tr>
<td>SL 7</td>
<td>Ac-RSD<em>YFTD</em>QBRRM</td>
</tr>
<tr>
<td>SL 8</td>
<td>Ac-RTD<em>YGND</em>NBBRM</td>
</tr>
<tr>
<td>SL 9</td>
<td>Ac-RTD<em>YQVD</em>ABBRM</td>
</tr>
</tbody>
</table>

The following section will discuss the experimental results as well as the associated discussion.

3.4 DATA ANALYSIS AND DISCUSSION

CT26-FL3 +1 was obtained from the Peña lab. This sample needed to be split in order for the subsequent passages to be grown. Secreted protein samples were from CT26-FL3 +2 through CT26-FL3 +10 with the exception of CT26-FL3 +6. There were issues with the cell culture of this passage; however, subsequent passages were grown without a problem. Recall, that each protein solution (from the passage) is incubated with SLs 1-9. Figure 3.3 depicts the Principal Component Analysis (PCA) of the SL array’s response with each protein sample. Each data point represents the response of each SL’s interaction with the fluorescently labeled protein. As discussed in Chapter 1, rather than looking at the fingerprint pattern of each SL’s response with each proteins sample it is easier to employ various statistical methods.

Principal Component Analysis is a statistical method employed throughout this chapter of the dissertation. It is similar to Linear Discriminant Analysis (LDA), discussed in Chapter 1, in that the program is attempting to group similar things together
and different things apart. However, the main difference between LDA and PCA is that PCA is unguided. Therefore, the statistical program does not know that CT26-FL3 +2 is different than CT26-FL3 +3.

Figure 3.3: Principal Component Analysis (PCA) of secreted proteins from CT26-FL3 passages.

Principal Component Analysis was used in this case because unlike the cell lines used in Chapter 1 and Chapter 2, we did not know the metastatic potential of each passage. The only information that was known is that the Peña lab saw a metastatic potential reversion around passage 3-5. Therefore, PCA could be used to understand how the different samples grouped. With this method it was possible to extract information about the samples and how they are grouped. In looking at Figure 3.3, it can be seen that there are two distinct populations. CT26-FL3 passages +5, +7, +8, +9, and +10 are one population, while CT26-FL3 +2, +3, and +4 are the other population.
Figure 3.4 is the same PCA plot from Figure 3.3 with the addition of the black circles indicating the initial groupings that were seen. These groupings correlate to the observations seen by the Peña lab. Recall, the metastatic reversion took place somewhere between 3-5 passages. These groupings place CT26-FL3 +5 through CT26-FL3 +10 together and CT26-FL3 +2 through CT26-FL3 +4 together. With the invasive information from the Peña lab, it can be assumed initially that passages +2 through +4 would be highly metastatic with +5 through +10 as non-metastatic.

To look at these groupings further, linear discriminant analysis (LDA) was used. Recall, this is the guided statistical method in which the computer program creates a model based on the information given about each group. Figure 3.5 shows the LDA plot of this data.
All of the data depicted in the PCA plot was from secreted proteins. Therefore, the healthy cell line data is based on the sensor array’s interaction with secreted proteins from CCD-841 CON, a healthy/normal colon cell line. The information put into the model was that CT26-FL3 +2-CT26-FL3 +4 was identified as cancerous metastatic and CT26-FL3 +5-CT26-FL3 +10 was identified as cancerous metastatic. The model created is what will be called 3 class, which means each point was identified in the program as healthy, cancerous non-metastatic, and cancerous metastatic. The N=116 for this data set and the leave-one-out cross-validation accuracy was 98%. This is obviously a high classification accuracy, however, it can be argued that this was to be expected as the groupings were decided based on where they fell in the PCA plot (which was unguided).

To validate these groupings, sensor array data was added that included known metastatic (LoVo) and non-metastatic (HCT 116 and HT29) secreted proteins. This LDA plot is seen in Figure 3.6.
Adding this information will allow the visualization of the CT26-FL3 passages along with the cell line data to validate the groupings that were identified in the PCA graph. Again, this plot was generated by using a three-class classification accuracy: healthy/normal, cancerous non-metastatic (lowly), and cancerous metastatic. It can be seen that the CT26-FL3 passages, whether assigned cancerous non-metastatic or cancerous metastatic are grouping with the secreted protein of that metastatic potential. The N=149 for this plot and the leave-one-out classification accuracy was 97%.

3.5 CONCLUSIONS AND FUTURE WORK

The information presented in the first part of this chapter indicates that the reversion may be seen from the glycosylation standpoint. The array can is able to group the 9 passages into two populations. With the metastatic reversion information from the Peña lab, it is possible to extrapolate the metastatic potential after subsequent passaging.

There are a variety of directions this project could be taken. The first would be to repeat these experiments beginning with CT26-FL3 +1 and passing through CT26-FL3
+10. To externally verify the metastatic potential of each passage, invasion assays could be performed. Dr. Peña’s lab has performed these assays, however, the invasion assays would need to be re-done in conjunction with the cell culture passages and sensor array data. Figure 3.7 depicts a general protocol of an invasion assay as well as general results.

Figure 3.7: Invasion assay. a) pictorial representation of an invasion assay being performed. b) Both pictures are the bottom of the trans-membrane well of an invasion assay. The left picture is a photograph of a non-invasive cell line after an invasion assay and the right picture is a highly invasive cell line after an invasion assay.

Figure 3.7a depicts the overall scheme of an invasion assay. A trans-membrane well is used that has a layer of matrigel in it. This matrigel mimics the basement membrane that a cell would need to migrate through in the body. The cells are starved for 24 hours and placed in the top of the well. A chemo attractant is placed in a chamber below the well, in this case, it would be media with FBS added, as the cells that can move would move to where the supplement is. The cells are incubated or 24 hours. The picture on the right of Figure 3.7a mimics the result of the invasion assay with a mixture of invasive and non-invasive cell lines. In this case, the green cells would be the cells with the invasive properties. After the incubation the cells are stained and imaged.
Figure 3.7b\textsuperscript{51} is a photograph and what the results would look like when comparing an invasive cell line (right) with a non-invasive cell line (left). Ideally, these experiments would verify what the sensor array data and the PCA plot indicate, two populations indicating a change in metastatic potential from the glycosylation pattern point of view.

Another addition to this project could be the use of natural lectins. What is assumed is that because the SL sensor array is binding with sugars there has to be a different response because of aberrant glycosylation. While extremely selective, natural lectin arrays could be used to discover which glycans are changing between the passages (if there is a natural lectin to detect this sugar). Because there is such a high correlation between sLe\textsuperscript{x} and metastasis this could be one sugar to potentially investigate. The natural lectin information could then be combined with the SL array information in order to create a better picture of what is happening throughout the metastatic reversion from the glycosylation side of things.

3.6 PART 2: INVESTIGATING METASTATIC POTENTIAL IN AN IN-VIVO MOUSE STUDY

The second part of this chapter will focus on using CT26-FL3, highly metastatic cell line, and CT26, a lowly metastatic cell line, for a longitudinal mouse study. The aim of this study was to determine if the sensor array could detect glycosylation changes from what is secreted by the tumor into the vasculature. It was also to determine if the sensor array can differentiate the cancerous from non-cancerous mice and when the disease became metastatic. The experimental workflow section of this chapter will discuss the basic set-up and collections, while the experimental section will discuss the intricate details of each part of this experiment.
3.7 PART 2: EXPERIMENTAL WORKFLOW

It should be mentioned that Dr. Peña’s lab did the mouse work for all of the experiments discussed. While I was present for most of the blood draws and the final sacrificing and weighing, I am grateful to the Dr. Peña and all those in her lab for taking on the workload of the mouse study.

The study involved the use of fifteen mice. Table 3.3 provides the four types of mice as well as the condition of each mouse. The mice used were Balb/c mice.

Table 3.3: Mouse conditions and information for longitudinal mouse study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Number Of Mice</th>
<th>Cells Injected</th>
<th>Metastatic Potential</th>
<th>Condition Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>3</td>
<td>None</td>
<td>N/A</td>
<td>None</td>
</tr>
<tr>
<td>Sham</td>
<td>4</td>
<td>None</td>
<td>N/A</td>
<td>Surgery, no injection</td>
</tr>
<tr>
<td>CT26</td>
<td>4</td>
<td>CT26</td>
<td>0-5%</td>
<td>Surgery, CT26 injected</td>
</tr>
<tr>
<td>CT26-FL3</td>
<td>4</td>
<td>CT26-FL3</td>
<td>&gt;95%</td>
<td>Surgery, CT26-FL3 injected</td>
</tr>
</tbody>
</table>

The control mice had no procedures done to them. Because aberrant glycosylation occurs with inflammation the sham mice were included in the study. The sham mice had the same incision as the CT26 and CT26-FL3 mice, however, they did not have any cells injected, but rather were opened up and closed back up. The CT26 mice had surgery and the lowly metastatic cells injected, while the CT26-FL3 mice had surgery and the highly metastatic cells injected. All of the mice were subjected to a pre-bleed before any surgery happened. This blood draw happened at the eye of the mice on average every 5-7 days. The blood was spun down and the serum samples were given to
the Lavigne lab. Table 3.4 shows the exact days of the experiment and what associated event took place. The experiment lasted a total of 47 days. It should be noted that one of the CT26 mice died during the experiment.

Table 3.4: Specific blood draws and surgery days for the longitudinal mouse study.

<table>
<thead>
<tr>
<th>Day</th>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5-29-15</td>
<td>Blood Draw</td>
</tr>
<tr>
<td>N/A</td>
<td>6-1-15</td>
<td>Surgery</td>
</tr>
<tr>
<td>2</td>
<td>6-5-15</td>
<td>Blood Draw</td>
</tr>
<tr>
<td>3</td>
<td>6-10-15</td>
<td>Blood Draw</td>
</tr>
<tr>
<td>4</td>
<td>6-18-15</td>
<td>Blood Draw</td>
</tr>
<tr>
<td>5</td>
<td>6-25-15</td>
<td>Blood Draw</td>
</tr>
<tr>
<td>6</td>
<td>7-2-15</td>
<td>Blood Draw</td>
</tr>
<tr>
<td>7</td>
<td>7-9-15</td>
<td>Blood Draw</td>
</tr>
<tr>
<td>8</td>
<td>7-14-15</td>
<td>Blood Draw (sacrifice)</td>
</tr>
</tbody>
</table>

For ease of explanation, the "day" will refer to the blood draw as on the day of the surgery no blood was taken. Each blood draw from the mouse resulted in about 50 µL of serum except for the sacrifice blood draw, which was about 150 µL of serum as there was no risk of harming the animal by taking too much blood.

For reasons summarized in Chapter 2, one of the major concerns initially was being sure that the albumin was removed from the blood samples. With this in mind the first step in the experiment was to use a Pierce™ albumin depletion kit (Product #85160) in order to remove the albumin. Following this, the sample was spun into carbonate buffer, fluorescently labeled with FITC, and a BCA assay was done in order to quantify protein concentration. After the BCA, the sample was then spun into PBS and a BCA was done again to determine the amount of sample needed to use a 50-fold dilution of .01 mg/mL.
Unfortunately, after a large portion of the mouse samples had been depleted it was noticed that the kit purchased was not meant for mouse samples. It had said that it would not be effective for mouse samples. Rather, than starting over and using more of the mouse sample, the SL array response with both experimental set-ups: using the albumin depletion kit and not using the albumin depletion kit. Ideally, if the kit is not effective with mouse samples then it will not affect the sample’s response with the synthetic lectin array. This experiment was done in order to know if the “albumin depletion” had to be done to keep everything consistent or if the following samples could be run without this step. It should be noted that taking out the albumin depletion would significantly reduce sample work-up. Figure 3.8 is a flow chart of both experimental options.

![Experimental flow chart of both experimental options for the mouse serum samples. A) represents the experimental work-up if the albumin depletion kit is not used. B) represents the experimental work-up if the albumin depletion kit is used.](image-url)

Figure 3.8b was initial experimental set up. The albumin was depleted using the depletion kit. After the kit was used, the serum was in a wash buffer. The solution was
then exchanged using Amicon<sup>®</sup> 10K centrifuge (spin) tubes (EMD Millipore) in order to have the serum sample in carbonate buffer. The sample was labeled with FITC and the solution was exchanged to be PBS prior to a BCA assay and incubation with SLs 1-9. Figure 3.8a depicts the second experiment designed to validate the albumin depletion kit use. In this step, the serum sample was diluted directly into 600 µL of carbonate buffer. The sample was labeled with FITC and the solution was exchanged to be PBS prior to a BCA assay and incubation with SLs 1-9 exactly as it was done in Figure 3.8b. Figure 3.9 shows the fingerprint pattern for each experiment (A and B) for the mouse 3-M1. This mouse was one of the CT26-FL3 mice. This sample was chosen because there was a large amount of serum from this draw.

![Figure 3.9: Fingerprint patterns from SLs 1-9. 3M1 is the mouse used. 3M1-A is experiment A from the flow chart above where no albumin depletion was performed. 3M1-B is experiment B from the flow chart above where albumin depletion was performed.](image)

3M1-B is the fingerprint pattern for the protocol that used the albumin depletion kit, while 3M1-A is the fingerprint pattern for the protocol that did not use the albumin depletion kit. The error bars represent the standard deviation for each SL. It should be noted that all percent standard deviations were below 15%. Looking at the fingerprint
patterns above, it can be seen that 3-M1A and 3-M1B have relatively similar patterns. With this information, it was decided that the experiment could move on without the use of the albumin depletion kit. The advantage of this was that the albumin depletion was a time consuming step and only a few samples could be done at a time with one person. Within the data analysis section, it will discuss which samples have been processed with the albumin depletion kit and which samples have not if it pertains to the graph presented.

3.8 DATA ANALYSIS AND DISCUSSION

With such a large number of samples [(15 mice)(8 blood draws) = 120 serum samples], there had to be an overall plan for how the serum samples would be processed. It was decided to initially investigate the first day and the last day of the experiment. This would compare the blood draw from the pre-bleed, which was before any surgery (day 1), with the blood draw from the final sacrifice day (day 8). This would roughly be about 30, recall one mouse died during the experiment, so in actuality it was 29 samples. As this was the first set up samples that were worked with all of them went through all of the samples went through protocol B, which involved the use of the albumin depletion kit. In the graphs this will be listed as day 1 and day 8, as it is in Table 3.4. This seemed an obvious place to start, as this should be the biggest difference in glycosylation patterns. Aberrant glycosylation begins at the onset of cancer and continues to change as the disease progresses. Figure 3.10 shows the PCA plot of the SL array’s response with all day 1 (pre-bleed) and day 8 (sacrificing bleed) samples.
Each data point represented in the plot is from the response of SL 1-9 with that sample. The values used were raw bright numbers from the sensor array. Principal component analysis was chosen because it is an unguided method. We did not know what samples represented what stage of the disease. It was assumed that the CT26-FL3 would be metastatic by the end of the experiment; however, we did not want to assume any metastatic potential. All of the day 8 (at sacrificing) draws have black lines around the data points, while all of the day 1 (pre-surgery) draws do not. The information in this graph can be difficult to tell which groups should be grouped together. Therefore, Figure 3.11 will be used to clarify this.
Figure 3.11: PCA plot of day 1 (pre-surgery bleed) and day 8 (sacrificing bleed). The boxes around CT26-FL3-8 and CT26-8 represent that these two groups should be clustering together, while all of the unboxed samples should also be clustering together. This ideally would create two populations. This is created with bright raw numbers.

This is the same PCA plot from Figure 3.10, however boxes have been added to clarify the groupings that were expected. Ideally, it was expected that there would be two populations.

The first grouping would be all of the CT26-8 and CT26-FL3-8 (both of these are boxed in the legend) data points. This grouping was expected because all of these mice would have had cancer, whether metastatic or not metastatic. The second grouping would be the unboxed sample IDs in the legend. This would include the control (day 1 and 8), all of the sham (day 1 and 8), CT26 (day 1), and CT26-FL3 (day 1). It was expected that the day 1 samples should all be grouped together because this was the pre-bleed before any surgery and/or injection of cancerous cells occurred. It was also
expected that the control and sham day 8 would stay with the day 1 (pre-bleed) samples because these mice had not any cancerous cell injection.

While the majority of the points were grouped in the way hypothesized, there were two data points (both denoted with an arrow pointing to them)-CT26-FL3-8 and CT26-1 that seem to be outside of the trend that was expected. As stated before, the numbers used for this analysis was using the raw bright numbers extracted from the MATLAB experiment. At this time, the lab was attempting to figure out which numbers to use from the sensor array as the numbers given from MATLAB are given with different channels (red, green, blue, and bright). With this in mind, bright normalized was run to look at the PCA plot given. This plot is shown in Figure 3.12.

Figure 3.12: PCA plot of day 1 (pre-surgery bleed) and day 8 (sacrificing bleed). The boxes around CT26-FL3-8 and CT26-8 represent that these two groups should be clustering together, while all of the unboxed samples should also be clustering together. This ideally would create two populations. This is created with normalized bright numbers.
The samples that were expected to group together are boxed as in the last plot and all other samples were expected to be together for the reasons stated above. When using the normalized bright numbers, there is one point, denoted with an arrow, which is grouping in a place that was not expected. This point, C26-FL3-8, should be a blood draw from a highly metastatic mouse at the sacrifice day, however, is grouping near the control mice from the last day.

Upon first analysis of the data, it looked as though the SL array could distinguish between the non-cancerous draws and the cancerous draws, however, it was a very limited data set. With this in mind, it was decided that it would make the most sense to choose one of the mid-experiment draws to look at the results in a similar way. This would include the addition of day 4 (mid-way through the experiment), approximately twenty days into the experiment. PCA is still being used because it is an unguided statistical method. The plot below uses SL sensor array data that is the normalized bright numbers. The plot is shown in Figure 3.13.
Figure 3.13: PCA plot of day 1 (pre-surgery bleed-points without outlines), day 4 (near midpoint of the experiment-points with dashed outlines), and day 8 (sacrificing bleed-points with black outline). The boxes around CT26-FL3-8 and CT26-8 represent that these two groups should be clustering together. This is created with normalized bright numbers.

The day 1 points (pre-surgery bleed) are the data points without the outlines. The day 4 points (close to midway through the experiment) have the dashed lines around them, while the day 8 points (sacrifice bleed) are the points with the solid black line around them. The assumption was still made that the CT26-FL3 day 8 and the CT26 day 8 (both are boxed in the plot above) samples should group together. It was also assumed that all of the pre-surgery points should group together (all points that are not outlined). These were the same assumptions made with the previous plots. As can be seen in the plot above, the difficulty begins when adding in the day 4 points. To simplify this process, the plot shown in Figure 3.13 is depicted again in Figure 3.14 with arrows drawn to the points that are in question.
Figure 3.14: PCA plot of day 1 (pre-surgery bleed-points without outlines), day 4 (near midpoint of the experiment-points with dashed outlines), and day 8 (sacrificing bleed-points with black outline). The boxes around CT26-FL3-8 and CT26-8 represent that these two groups should be clustering together. Arrows are drawn to the points that information is not known about where these points should cluster. This is created with normalized bright numbers.

It can be inferred that the sham day 4 and control day 4 (green points with dashed lines as the outline), should cluster with all pre-surgery points as well as sham day 9 and control day 8. The points that have arrows drawn to them are the points in question where they are grouping. It can be seen that there are significantly more points that are in question that in Figure 3.12.

The main disadvantage for the designed experiment is that there was no external validation of what stage each mouse was in. For instance the CT26-4 point towards the right of the plot looks as if it is grouping with the points that would be without cancer (sham, control, and pre-surgery). The problem is that it is not known whether this particular mouse had cancer at the time or if in fact the cells had not grown in the mouse.
The problem with the CT26-FL3 mice is that when the blood draws were taken it was unknown whether the cells had metastasized. All samples from Day 1, 2, 4, 6, and 8 were analyzed, however, the complications seen in the above graph were multiplied.

3.9 CONCLUSIONS

The intent of this experiment was to evaluate the SL array in terms of being able to distinguish between healthy, cancerous non-metastatic, and cancerous metastatic mice. This would have given detailed information about when the SL array could detect a change in metastatic potential. All of the previous experiments with the SL array had been used on cell lines in which the metastatic potential was already established and not changing in vivo.

The longitudinal mouse study done Part 2 of Chapter 3 indicated promise when comparing the pre-bleed experiments with those from the sacrificing bleed. With the exception of one point, it appeared as though all of the points containing pre-bleed as well as the sacrificing bleed for the sham and control mice were a population, while the CT26 and CT26-FL3 day 8 mice were another population. Difficulty began to arise when adding in additional days without a way to externally validate the cancer stage in the mice. Five days were analyzed (Days 1, 2, 4, 6, 8), however the data did not appear to group in any way. The additional days were not analyzed due to a cost-benefit analysis of the data that was being generated. The following section will describe the future work to make this experiment valuable.

3.10 FUTURE WORK

While the initial Day 1 and Day 8 experiments showed promise, the primary problem to address in doing an additional longitudinal study is that of external validation.
In order to test the array’s sensitivity in detecting cancer and the metastatic changes, the experiment would need a way to look at where the cancerous tumor is and has this tumor metastasized.

Dr. Marj. Peña’s has the CT26 and CT26-FL3 cells that contain a fluorescent tag. The Center for Colon Cancer Research (CCCR) at the University of South Carolina has the instruments necessary for imaging whole mouse specimens with anesthesia. In an ideal experiment, each blood draw from a particular mouse would have a corresponding picture of that mouse which would depict the fluorescent cancerous cells. This would allow for the researcher to be able to “quantify” the cancer and correlate this to the SL array’s response. With this information, it would no longer be a question why a CT26-FL3 mouse was clustering with a CT26 sample. The picture could give information about if the cells had metastasized or not. This information is invaluable in determining the efficacy of the SL array. This future experiment could also include using an albumin removal kit that was effective in mouse samples. In addition, a piece of the final tumor could be given to the investigator in order to look at the differences in the tumor response form the serum response. Based on the proof-of-concept experiment in Chapter 2, it can be inferred that these responses would be different, however, it would be interesting to see in an in vivo experiment if this was true.

3.11 EXPERIMENTAL FOR CT26-FL3 REVERSION STUDY (CHAPTER 3 PART 1)

Cell Culture

CT26-FL3 cell line used a media made up of Dulbecco’s Modified Eagle Medium (DMEM) purchased from VWR (45000-734), 10% FBS (50mL) purchased from VWR (45000-734) and 1% (5mL) antibiotics/antimycotics purchased from VWR (45000-616).
Media Preparation

Remove 55mL of media from the bottle and place in a centrifuge tube. You can use this later for starving cells if needed. Add the FBS and antibiotics/antimycotics in with a pipette. Do not pour this in. After everything has been added to the media use a larger pipette and pipette up and down in order to thoroughly mix the media. You can do this cold if you are not using the media for anything at this time.

Thawing Stocks

Typically, the initial thaw of a stock will want to be done in a T25 flask. Once you know that cell viability is high for a cell stock, you can thaw into a T75 flask. If you notice the T25 is confluent fast, the next time you can thaw into a T75. Pre-warm your full media (all additives added) to 37°C. Make sure you have everything you need to thaw the cells in your hood ready to go, as DMSO is toxic to cells so they need to be thawed and thawed in a larger volume of media.

Place about 5mL of warm media into a 15mL Falcon tube. Pipette small amounts of warm media into the cryogenic vial. You will need to pipette a small amount in and then pipette up and down. As some of the cell stock thaws, place it into the falcon tube with the other warm media to dilute the DMSO. Keep adding the warm media and removing the DMSO/cell solution to media until all of the stock has thawed. The following day remove the media and add fresh media. Be careful to not do this too soon after. Dead cells and live cells look different yet they both float (dead cells will look like broken dark pieces where live unattached cells will look like nice healthy spheres). Usually, cells were split 2-3 times before running any experiments to make sure these were healthy and viable.
**Feeding the Cells**

Feeding the cells requires the pipetting of old media off before putting new media on. I would suggest changing the media every couple of days when not splitting the cells. The pink media will begin to turn orange when the cells are either too confluent or the media needs to be changed (you should not let it get to this point). The media should be warm (37°C) for this step.

**Splitting Cells**

The trypsin/EDTA media should be warmed in incubator to 37°C before beginning the process of splitting the cells. The cells should be at 70-75% confluence prior to splitting. Take the existing media off of the cells and put into the waste container. The same pipette may be used as long as it does not touch the outside of the flask or any part of the hood. Wash the back of the flask with 1-2 mL of the trypsin/EDTA solution then pipette this off and put into waste. This will remove the dead cells and excess media that may not have come off with the first media removal. Then (with a fresh pipette) add trypsin/EDTA to the flask. The amount needed will depend on the flask size, but you want enough of the solution to cover the bottom of the flask when horizontal. This will ensure that all cells will have the solution over them. Place the flask back into the incubator (37°C).

The amount of time necessary to trypsinize the cells will depend on the cell line. In the beginning, check the cells every 1-2 minutes. You can either hold the flask up to the light or look under the microscope to see if the cells are attached. If you know that your cell line is hearty you can rub the back of the flask to speed up this process or in extreme cases hit the flask against your hand. The cells should be floating when the have
completely detached. Be mindful not to leave too many cells behind, as this will deplete your cell number.

Wash the cells that have detached to the bottom of the flask with fresh warm media. This amount will depend on the size flask used. You may need to rinse the flask a couple of times to ensure all of the cells have been removed. Add this solution to a 15 mL centrifuge tube and centrifuge for 4-5 minutes at 500 x g. Be sure to not spin at too high of a rate as this will damage the cells. Remove the supernatant and put in the waste container. Add 1-2 mL of fresh media to the cell pellet. Use a small pipette (p1000) and pipette this solution up and down. This will ensure there are not clumps of cells when adding to a new flask. Add the appropriate amount of microliters to each flask depending on the split you want and how many flasks you need. If there are extra cells these can be placed in the waste container. Add the media to the flask to make it to the total volume for each flask type (5 mL for T25, 10 mL for T75, and 20 mL for T125). The following day check to make sure the cells have attached and there are not too many floating cells. You can change the media if you would like using warmed media.

Making Stocks

Only make stocks from cells that are actively growing and no more than 75-80% confluent. If you are unsure, split the cells again and makes stocks from the next passage. One T75 flask should result in 3-5 stocks, however, adjust this number depending on how many cells you want in your stock. I tend to put more in so that it is a better chance of recovery. Cell stocks are frozen in cryovials that can be found in the stockroom. Label each vial with you initials, cell line, and cell passage.
Make sure everything is ready in the hood prior to these steps. DMSO is toxic to cells so you must move quickly. Have a cooled solution of 20% DMSO in FBS (this should have been filtered through a .2 µm filter. Trypsinize the cells according to the instructions above, inactivate the trypsin with media, and put into a centrifuge tube. Centrifuge the cell suspension for 4-5 minutes at 500 x g and remove the supernatant. The following steps you will want to use a p1000 pipette. Re-suspend the cell pellet in the usual full media. Add the amount you will need to have is .5 mL * the number of stocks you are making. Be sure to pipette the solution up and down in order to equally distribute the cells into all of the vials. Add .5 mL of the cell suspension to each cryovial and then add .5 mL of the 20% DMSO-FBS solution.

As soon as the DMSO is added the cells they must be frozen immediately. The best way to do this is use the Mr. Frosty container and place in the freezer. This will allow the temperature to reduce gradually. The Mr. Frosty contains isopropanol that needs to be changed every five uses. Mark the top of the container with a marker in order to indicate which use it was.

Starving the Cells

When the cells reached no more than 70% confluence, they were starved for 48 hours. The media did not include any supplements. This was due to some cell lines dying when antibiotics were present. For these first experiments, the maximum amount of media was used that one flask of cells could reasonably be split. In most cases one T-75 flask was split into five T-75 flasks. One of these flasks would continue to grow further passages, while the four other flasks would be starved for the experiments. In my notebook after the BCA it is seen that even with these flasks sizes, there was always more
protein that what was needed. The only different to this set up was the first passage. CT26-FL3 +2 had less flasks because the initial cells given were in a petri dish rather than a full T-75 flask. I have used as little as 5 mL of just media. If you are doing an experiment where cells are being used in addition to the media the number of flasks needed is simply the number needed for cells.

Concentration of Media for Secreted Proteins

After 48 hours the media was removed from the flask and centrifuged to remove cell debris. The media was pipetted off and put in a new tube. The media was centrifuged in Amicon® 3K centrifuge tubes (EMD Millipore). This will be the longest step in the process. Only 10-15 mL will be spun through every twenty minutes. Keep this in mind when deciding on how many flasks to use. When the media is mostly spun through and there is only a small amount in the bottom of the filter (be sure not to let the filter go dry) it is a good idea to rinse PBS through in order to get rid of all of the media. If colored media is being used when the media is no longer colored and is a true clear you have finished the wash step. If the media is colorless it will take a little more intuition to figure this out. There should be a small amount of liquid in the filter tube (around 1500 µL). Pipette this out with a p200 pipette and tip (yellow). Due to the filter size you want to make sure the pipette tip reaches the bottom because this is where the protein is. After pipetting the solution out in a 15 mL centrifuge tube add PBS and rinse the filter pipetting up and down multiple times.

Acetone Precipitation of Secreted Proteins from Media

The protein is then precipitated with ice-cold acetone stored in the -30 degrees Celsius freezer. Add four times the amount of acetone to the protein solution. For
example, if there is 2 mL of protein 8 mL of acetone would be added. Shake the tube in order to see a white fluffy precipitate. At this point, the protein solution can be stored in the -30 degrees Celsius freezer overnight if you wish or for a minimum of an hour. The following day (or after some time) take the sample out of the freezer, spin it down, and remove the acetone. All the centrifuge tube to sit in the hood for 30 minutes to an hour to make sure the acetone is completely evaporated. At this point, the protein can be treated like any other protein sample. Usually (if moving on to the labeling step), the protein is re-suspended in carbonate buffer using only the amount necessary to completely dissolve (starting with about 1 mL each time). This will depend on the amount of protein recovered.

*Membrane Proteins and Glycoproteins*

The cells were scraped off the flask using a cell scraper. The flask was rinsed with PBS and the solution was centrifuged. The PBS was taken off to leave behind a cell pellet. This can be stored in the -30 degree Celsius freezer for a bit of time or in the -80 degree Celsius freezer for an extended period of time. The plasma glycoproteins and proteins were extracted using the Qiagen® Plasma Membrane Protein Kit. The protein at the end is in elution buffer and must be spun into carbonate buffer purified using Amicon® 10K centrifuge (spin) tubes (EMD Millipore). The solution was put into the spin tube and spun for about 8 minutes at 4000g. The time spun may change depending on protein concentration, but you want to spin until it is at the “2” line on the tube. When the solution is at the “2” line, 1.5 mL of carbonate buffer is put in the exchange tube and spun for 5 minutes. Again it is checked to make sure the solution is at the “2” line. At this point, you want to turn the small tube inside over and spin out into a micro centrifuge
tube in a benchtop centrifuge. Put in 300 µL of carbonate buffer and pipette up and down. It is advantageous to use a yellow tip again for this step in order to get all the way down in the filter. Turn tube over and spin again. Rinse tube with 2-100 µL increments of carbonate buffer.

**Labeling of Membrane and Secreted Proteins/Glycoproteins**

The membrane proteins/glycoproteins and secreted proteins were labeled separately with fluorescein isothicyanate (FITC) in DMF and carbonate buffer. 2.8 mg of FITC was dissolved in 100 µL of DMF. 10 µL of the FITC solution was put into 200 µL of protein. The total volume of the solution was brought up to 600 µL by adding 400 µL of carbonate buffer. The protein solution was tumbled for 1 hr at 37 degrees Celsius and purified using Amicon® 10K centrifuge (spin) tubes (EMD Millipore). The solution was put into the spin tube and spun for about 8 minutes at 4000g. The time spun may change depending on protein concentration, but you want to spin until it is at the “2” line on the tube. When the solution is at the “2” line, 1.5 mL of PBS is put in the exchange tube and spun for 5 minutes. Again it is checked to make sure the solution is at the “2” line. At this point, you want to turn the small tube inside over and spin out into a micro centrifuge tube in a benchtop centrifuge. Put in 300 µL of PBS and pipette up and down. It is advantageous to use a yellow tip again for this step in order to get all the way down in the filter. Turn tube over and spin again. If secreted proteins are being used, repeat the rinse step since there is a large concentration of protein. If membrane extracts are being used be aware that there is not as big of a concentration and thus maybe rinse with 2-100 µL increments of PBS. After labeling a BCA assay was done to quantify the amount of protein.
**BCA Assay**

Thermo Scientific BCA kit was purchased from VWR.

Sample Preparation

If using secreted protein sample you may need to dilute your samples and run these with the concentrated sample to ensure the concentration falls within the range of the assay. If this BCA assay is done before labeling, the samples will be in carbonate buffer, thus your standards should be made in carbonate buffer. If this BCA is done after the labeling steps, the samples will be in PBS and thus your standards should be made in PBS.

Standard Preparation

There is a standard 2 mg/mL solution included with the kit. Put 100 µL of this solution into a small plastic tube. Put 50 µL of carbonate buffer in several other tubes. Then take 50 µL of the 2 mg/mL solution in the first tube. Be sure to vortex this several times (This is the first 1:2 dilution). Then take 50 µL of that solution and put it in the next tube. Continue this process. Make sure you also have a blank standard (either carbonate buffer or PBS depending on what step you are doing this).

Reagent Preparation

The directions instruct you to make the reagent to be made 50:1. 200 µL of reagent is required for each well so the amount of total reagent will need to be calculated (200 µL x # of wells). This can be done in a 15 mL centrifuge tube.

Well Preparation

Make sure the standards/samples are done in triplicate. Put 10 µL in each well of standards and samples. Then put 200 µL of reagent in all of the wells. Make sure to then
mix each well with a pipette tip. The plate should then be placed in the oven for 30-45
minutes at 37 degrees Celsius. Periodically check your plate to see if the standards have
changed. The plate is then read on the plate reader at 562 nm. Be sure not to leave your
plate for a long period of time. It should be noted that this assay is not an endpoint assay
and will continue to develop.

*Synthesis of Synthetic Lectins*

SLs are attached to Tentagel beads. The beads are weighed out and put in 10mL
of DMF(dimethylformamide) for ten minutes in order to swell the beads. Meanwhile,
Fmoc-Met-OH(10eq) is activated with HBTU(10eq) for 10min in 5% N-
methylmorpholine in DMF. The amino acid solution is added and tumbled for 1 hour.
The solution is drained and the beads are washed with DMF, MeOH(methanol), DMF,
MeOH. After the last washing, a ninhydrin test is performed on one or two beads to see
if the amino acid has coupled. If coupled, the Fmoc group is removed using 10mL of
20% piperidine in DMF. That process is repeated manually until the amino acids are on
or a peptide synthesizer can be used. The SL is then N- terminally acetylated using
10mL of CH₂Cl₂ containing 5% pyridine and 5% acetic anhydride.

*Removal of protecting group on DAB (diamino acid)*

Diamino acid (DAB) has an ivDde protecting group on it. In order to put on the
boronic acids, these must be removed. The protecting group is removed by adding 10mL
of 2% hydrazine monohydrate in DMF and tumbling for 1 hour.

*Addition of Phenyl Boronic Acids (PBA)*

2-formylphenyl boronic acid (8eq) is dissolved in 10% MeOH in DMF. Activated 3Å molecular sieves and the PBA solution are added to the beads and tumbled
overnight at 37 degrees Celsius. Sodium borohydride (NaBH₄) (8eq) is added and gas was released over 30 min. The beads then tumble for 4 hours. The beads are washed with DMF and MeOH and the sieves are removed. The beads are then washed again.

**Removal of Acid Labile Groups**

In order to remove the acid labile groups the beads are tumbled with 10mL of 95% trifluoroacetic acid(TFA), 2.5% water, and 2.5% triisopropylsilane for 1 hours. After tumbling, the solution is taken off and washed with DMF, MeOH twice. The synthetic lectins can then be washed with PBS three times in order to cut down on PBS wash time done before incubation. After washing with PBS it is important that the tubes of SLs be placed in the refrigerator and not the -30 degrees Celsius freezer.

**MALDI Sample Preparation**

Phenyl Boronic Acid (PBA) Group Removal

This step can be avoided if a small amount of beads are taken out before the boronic acids are added. Once removed, the rest of the SL synthesis steps are followed on the small amount of beads. However, if the boronic acids are still attached and a MALDI is needed to be run these groups will need to be removed.

Remove PBA groups by add 100 mmol solution of H₂O₂ to the 1.5 mL centrifuge tube containing the beads. Put this tube at 50 degrees Celsius for 1 hour. After the hour, wash the beads with dH₂O three times. Add 50 µL of a 40 mg/mL solution of Cyanogen Bromide (CNBr) made in .1M HCl. Make sure all work with CNBr is done in the hood. Place this solution in the dark (usually your drawer) for 15-18 hours. Evaporate this solution using the speed vacuum centrifuge. There should be a small amount of white substance at the bottom of the centrifuge tube.
MS Sample

Dissolve the sample obtained from the previous section in 10 μL of .1% TFA in H₂O. In order to make the matrix, make a saturated solution of α-cyano-4-hydroxy cinnamic acid (CHCA) in 50:50 acetonitrile/.1 % TFA in H₂O. Add 1 μL of sample to the desired spot on the plate for MALDI. Add 1 μL of matrix to the same spot. Pipette the 2 μL of solution up and down a few times to get a good mix of matrix and sample. Let the plate dry and take obtain MALDI spectra.

Incubation with Synthetic Lectins (SLs)

Each synthetic lectin is weighed into a centrifuge tube (2mg) for the binding study. The beads were washed with PBS twice and PBS with 1% bovine serum albumin (BSA) for 15 minutes. This solution was removed and 1mL of .0002 mg/mL (50 fold dilution of a .1 mg/mL solution) of a particular protein is put in the tube and allowed to incubate overnight at room temperature. The solution was removed and then washed with PBS three times. After the last wash, the beads are re-suspended in about 500 μL of PBS to make the exchange to the microscope slide easier.

Imaging the Beads

The beads are imaged with a fluorescence microscope, Leica MZ 16F, using a GFP filter set. It is imperative that all of the SLs in each cell line (in this case SL 1-9) were imaged at the same exposure time. In order to do this, place each slide under the microscope and find the brightest SL. Once this is found for a particular sample, optimize this slide to have the largest green value without saturating the camera (no more than 190 to be safe, 255 is the absolute max). This number can be found by circling a bead in Adobe Photoshop® and pulling up the histogram information. After finding the
appropriate exposure time, pictures can be taken with care that most beads are within the Region of Interest (ROI) and not touching one another, as these beads would be thrown out.

Data Analysis

After the pictures have been taken they must be converted into JPG or JPEG as the camera takes these pictures as TIFF files. There are protocols written in the lab to do this on a MAC or PC. After conversion into jpeg, the files can then be run through a code in MATLAB® written by Dr. Greytak and modified by Anna Veldkamp. The code used for these experiments was image analysis anna. There is a protocol written in the lab that walks a user through the necessary steps to run the code. This code essentially picks out each bead and provides information about area, bright number, red number, green number, and blue number. It also gives information about where this bead was located in the picture so that you can identify a number for each bead. The data is processed in excel and outliers are removed (1.5 IQD) in order to get the array information for the SLs with each protein type. At this point, the data can be put into varying statistical models, the most common seen in this chapter are linear discriminant analysis (LDA), principal component analysis (PCA), and machine learning which is an extension of linear discriminant analysis (LDA).

3.12 EXPERIMENTAL FOR MOUSE SERUM STUDIES (CHAPTER 3 PART 2)

Mouse Serum Specific Protocols for Protocol A (No albumin depletion kit)

Sample Preparation

There is a limited amount of sample preparation. However, there are a few things to keep in mind when the samples are received. It is possible that there are glass pieces
from the capillary tubes in the smaller tube so make sure this is removed. Some of the samples may contain tissue parts or blood. Centrifuge these and then only take the serum. Label the tube accordingly. Generally, there is about 50 µL of serum for each sample received.

*Carbonate Buffer*

There is a lot of protein present in a serum sample. To reduce the dilutions needed for a BCA assays, less serum can be taken here. It will also leave serum in case further experiments need to be done or if the SL experiment needs to be re-done. Safely, 15 µL can be taken (I have used as little as 5 µL). Dilute serum to 600 µL of carbonate buffer in a micro centrifuge tube.

*BCA Assay*

Thermo Scientific BCA kit was purchased from VWR.

Sample Preparation

If using secreted protein sample you may need to dilute your samples and run these with the concentrated sample to ensure the concentration falls within the range of the assay. If this BCA assay is done before labeling, the samples will be in carbonate buffer, thus your standards should be made in carbonate buffer. If this BCA is done after the labeling steps, the samples will be in PBS and thus your standards should be made in PBS.

Standard Preparation

There is a standard 2 mg/mL solution included with the kit. Put 100 µL of this solution into a small plastic tube. Put 50 µL of carbonate buffer in several other tubes. Then take 50 µL of the 2 mg/mL solution in the first tube. Be sure to vortex this several
times (This is the first 1:2 dilution). Then take 50 µL of that solution and put it in the next tube. Continue this process. Make sure you also have a blank standard (either carbonate buffer or PBS depending on what step you are doing this).

Reagent Preparation

The directions instruct you to make the reagent to be made 50:1. 200 µL of reagent is required for each well so the amount of total reagent will need to be calculated (200 µL x # of wells). This can be done in a 15 mL centrifuge tube.

Well Preparation

Make sure the standards/samples are done in triplicate. Put 10 µL in each well of standards and samples. Then put 200 µL of reagent in all of the wells. Make sure to then mix each well with a pipette tip. The plate should then be placed in the oven for 30-45 minutes at 37 degrees Celsius. Periodically check your plate to see if the standards have changed. The plate is then read on the plate reader at 562 nm. Be sure not to leave your plate for a long period of time. It should be noted that this assay is not an endpoint assay and will continue to develop.

Labeling

Depending on the concentration from the BCA assay, only label a portion of the sample in carbonate buffer. The serum was labeled with fluorescein isothicyanate (FITC) in DMF and carbonate buffer. 2.8 mg of FITC was dissolved in 100 µL of DMF. 10 µL of the FITC solution was put into 200 µL of protein. The total volume of the solution was brought up to 600 µL by adding 400 µL of carbonate buffer. The protein solution was tumbled for 1 hr at 37 degrees Celsius and purified using Amicon® 10K centrifuge (spin) tubes (EMD Millipore). The solution was put into the spin tube and spun for about
8 minutes at 4000g. The time spun may change depending on protein concentration, but you want to spin until it is at the “2” line on the tube. When the solution is at the “2” line, 1.5 mL of PBS is put in the exchange tube and spun for 5 minutes. Again it is checked to make sure the solution is at the “2” line. At this point, you want to turn the small tube inside over and spin out into a micro centrifuge tube in a benchtop centrifuge. Put in 300 μL of PBS and pipette up and down. It is advantageous to use a yellow tip again for this step in order to get all the way down in the filter. Turn tube over and spin again. If you are using secreted proteins you can repeat the rinse step since there is a large concentration of protein. If you are using membrane extracts you would want to be aware that there is not as big of a concentration and thus maybe rinse with 2-100 μL increments of PBS. After labeling a BCA assay was done to quantify the amount of protein.

*Incubation with Synthetic Lectins (SLs)*

Each synthetic lectin is weighed into a centrifuge tube (2mg) for the binding study. The beads were washed with PBS twice and PBS with 1% bovine serum albumin (BSA) for 15 minutes. This solution was removed and 1mL of .0002 mg/mL (50 fold dilution of a .1 mg/mL solution) of a particular protein is put in the tube and allowed to incubate overnight at room temperature. The solution was removed and then washed with PBS three times. After the last wash, the beads are re-suspended in about 500 μL of PBS to make the exchange to the microscope slide easier.

*Imaging the Beads*

The beads are imaged with a fluorescence microscope, Leica MZ 16F, using a GFP filter set. It is imperative that all of the SLs in each cell line (in this case SL 1-9)
were imaged at the same exposure time. In order to do this, place each slide under the microscope and find the brightest SL. Once this is found for a particular sample, optimize this slide to have the largest green value without saturating the camera (no more than 190 to be safe, 255 is the absolute max). This number can be found by circling a bead in Adobe Photoshop® and pulling up the histogram information. After finding the appropriate exposure time, pictures can be taken with care that most beads are within the Region of Interest (ROI) and not touching one another, as these beads would be thrown out.

Data Analysis

After the pictures have been taken, convert into JPG or JPEG as the camera takes these pictures as TIFF files. There are protocols written in the lab to do this on a MAC or PC. After conversion into jpeg, the files can then be run through a code in MATLAB® written by Dr. Greytak and modified by Anna Veldkamp. The code used for these experiments was image analysis anna. There is a protocol written in the lab that walks a user through the necessary steps to run the code. This code essentially picks out each bead and provides information about area, bright number, red number, green number, and blue number. It also gives information about where this bead was located in the picture so that you can identify a number for each bead. The data is processed in excel and outliers are removed (1.5 IQD) in order to get the array information for the SLs with each protein type. At this point, the data can be put into varying statistical models, the most common seen in this chapter are linear discriminant analysis (LDA), principal component analysis (PCA), and machine learning which is an extension of linear discriminant analysis (LDA).
**Mouse Serum Specific Protocols for Protocol B (using albumin depletion kit)**

**Sample Preparation**

There are a few things to keep in mind when the samples are received. It is possible that there are glass pieces from the capillary tubes in the smaller tube so make sure this is removed. Some of the samples may contain tissue parts or blood. Centrifuge these and then only take the serum. Label the tube accordingly. Generally, there is about 50 µL of serum for each sample received.

**Albumin Depletion Kit Protocol**

The kit used is the Pierce Albumin Kit (Product #45160). Take about 25 µL (or half of each sample if there is less than 25 µL) for the albumin depletion kit steps. The excess will ensure that there is enough protein even though there will be some protein loss at each step. The 25 µL of sample will get diluted with 75 µL of wash buffer from the kit. If there is less than 25 µL still only dilute with 75 µL of wash buffer. The protocol in the kit was followed exactly. In the wash step the protocol was followed exactly. In the last step four total washes were done. The protocol calls for 4-5.

**Carbonate Buffer Exchange from Wash Buffer**

This protocol is much the same as in all other experiments. After the albumin depletion was steps there is about 200-300 µL. All was taken to spin into carbonate buffer purified using Amicon® 10K centrifuge (spin) tubes (EMD Millipore). The solution was put into the spin tube and spun for about 8 minutes at 4000g. The time spun may change depending on protein concentration, but you want to spin until it is at the “2” line on the tube. When the solution is at the “2” line, 1.5 mL of carbonate buffer is put in
the exchange tube and spun for 5 minutes. Again it is checked to make sure the solution is at the “2” line. At this point, you want to turn the small tube inside over and spin out into a micro centrifuge tube in a benchtop centrifuge. Put in 300 µL of carbonate buffer and pipette up and down. It is advantageous to use a yellow tip again for this step in order to get all the way down in the filter. Turn tube over and spin again. Rinse tube with 2-100 µL increments of carbonate buffer.

BCA Assay

Thermo Scientific BCA kit was purchased from VWR.

Sample Preparation

If using secreted protein sample you may need to dilute your samples and run these with the concentrated sample to ensure the concentration falls within the range of the assay. If this BCA assay is done before labeling, the samples will be in carbonate buffer, thus your standards should be made in carbonate buffer. If this BCA is done after the labeling steps, the samples will be in PBS and thus your standards should be made in PBS.

Standard Preparation

There is a standard 2 mg/mL solution included with the kit. Put 100 µL of this solution into a small plastic tube. Put 50 µL of carbonate buffer in several other tubes. Then take 50 µL of the 2 mg/mL solution in the first tube. Be sure to vortex this several times (This is the first 1:2 dilution). Then take 50 µL of that solution and put it in the next tube. Continue this process. Make sure you also have a blank standard (either carbonate buffer or PBS depending on what step you are doing this).
Reagent Preparation

The directions instruct you to make the reagent to be made 50:1. 200 µL of reagent is required for each well so the amount of total reagent will need to be calculated (200 µL x # of wells). This can be done in a 15 mL centrifuge tube.

Well Preparation

Make sure the standards/samples are done in triplicate. Put 10 µL in each well of standards and samples. Then put 200 µL of reagent in all of the wells. Make sure to then mix each well with a pipette tip. The plate should then be placed in the oven for 30-45 minutes at 37 degrees Celsius. Periodically check your plate to see if the standards have changed. The plate is then read on the plate reader at 562 nm. Be sure not to leave your plate for a long period of time. It should be noted that this assay is not an endpoint assay and will continue to develop.

Labeling

Depending on the concentration from the BCA assay, only label a portion of the sample in carbonate buffer. The serum was labeled with fluorescein isothiocyanate (FITC) in DMF and carbonate buffer. 2.8 mg of FITC was dissolved in 100 µL of DMF. 10 µL of the FITC solution was put into 200 µL of protein. The total volume of the solution was brought up to 600 µL by adding 400 µL of carbonate buffer. The protein solution was tumbled for 1 hr at 37 degrees Celsius and purified using Amicon® 10K centrifuge (spin) tubes (EMD Millipore). The solution was put into the spin tube and spun for about 8 minutes at 4000g. The time spun may change depending on protein concentration, but you want to spin until it is at the “2” line on the tube. When the solution is at the “2” line, 1.5 mL of PBS is put in the exchange tube and spun for 5 minutes. Again it is
checked to make sure the solution is at the “2” line. At this point, you want to turn the small tube inside over and spin out into a micro centrifuge tube in a benchtop centrifuge. Put in 300 µL of PBS and pipette up and down. It is advantageous to use a yellow tip again for this step in order to get all the way down in the filter. Turn tube over and spin again. If you are using secreted proteins you can repeat the rinse step since there is a large concentration of protein. If you are using membrane extracts you would want to be aware that there is not as big of a concentration and thus maybe rinse with 2-100 µL increments of PBS. After labeling a BCA assay was done to quantify the amount of protein.

*Incubation with Synthetic Lectins (SLs)*

Each synthetic lectin is weighed into a centrifuge tube (2mg) for the binding study. The beads were washed with PBS twice and PBS with 1% bovine serum albumin (BSA) for 15 minutes. This solution was removed and 1mL of .0002 mg/mL (50 fold dilution of a .1 mg/mL solution) of a particular protein is put in the tube and allowed to incubate overnight at room temperature. The solution was removed and then washed with PBS three times. After the last wash, the beads are re-suspended in about 500 µL of PBS to make the exchange to the microscope slide easier.

*Imaging the Beads*

The beads are imaged with a fluorescence microscope, Leica MZ 16F, using a GFP filter set. It is imperative that all of the SLs in each cell line (in this case SL 1-9) were imaged at the same exposure time. In order to do this, place each slide under the microscope and find the brightest SL. Once this is found for a particular sample, optimize this slide to have the largest green value without saturating the camera (no more
than 190 to be safe, 255 is the absolute max). This number can be found by circling a bead in Adobe Photoshop® and pulling up the histogram information. After finding the appropriate exposure time, pictures can be taken with care that most beads are within the Region of Interest (ROI) and not touching one another, as these beads would be thrown out.

*Data Analysis*

After the pictures have been taken they must be converted into JPG or JPEG as the camera takes these pictures as TIFF files. There are protocols written in the lab to do this on a MAC or PC. After conversion into jpeg, the files can then be run through a code in MATLAB® written by Dr. Greytak and modified by Anna Veldkamp. The code used for these experiments was image analysis anna. There is a protocol written in the lab that walks a user through the necessary steps to run the code. This code essentially picks out each bead and provides information about area, bright number, red number, green number, and blue number. It also gives information about where this bead was located in the picture so that you can identify a number for each bead. The data is processed in excel and outliers are removed (1.5 IQD) in order to get the array information for the SLs with each protein type. At this point, the data can be put into varying statistical models, the most common seen in this chapter are linear discriminant analysis (LDA), principal component analysis (PCA), and machine learning which is an extension of linear discriminant analysis (LDA).
CHAPTER 4:

INVESTIGATING METASTATIC POTENTIAL USING STATISTICAL METHODS

4.0 ABSTRACT

Statistical methods are an important tool in many fields. This chapter will discuss the use of statistical methods to gain insight into the biochemistry of the sensor array. Specifically, it will use linear discriminant analysis and the application of training and test models to look at the use of 3T3, a mouse fibroblast cell line, as a healthy control. It will also discuss the use of reference beads in investigating the limitations and parameters of the fluorescence microscopy that is used in the synthetic lectin array experiments.

4.1 INVESTIGATION INTO A HEALTHY CONTROL

As mentioned in Chapter 1, the synthetic lectin sensor array was used to classify seven different cell lines by metastatic potential. The seven cell lines used are again listed in Table 4.1. Note that there are cell lines that are of mouse and human species. There is also a range of metastatic potentials.
Table 4.1: Cell lines discriminated with Synthetic Lectin array. The associated metastatic potentials are listed as well as the secondary metastatic site when applicable.

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Classification</th>
<th>Cancer Type</th>
<th>Origin</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>3T3/NIH</td>
<td>Healthy</td>
<td>N/A</td>
<td>Mouse</td>
<td></td>
</tr>
<tr>
<td>CT-26</td>
<td>Cancerous, Non-Metastatic</td>
<td>Colon</td>
<td>Mouse</td>
<td></td>
</tr>
<tr>
<td>HCT116</td>
<td>Cancerous, Non-Metastatic</td>
<td>Colon</td>
<td>Human</td>
<td></td>
</tr>
<tr>
<td>HT29</td>
<td>Cancerous, Non-Metastatic</td>
<td>Colon</td>
<td>Human</td>
<td></td>
</tr>
<tr>
<td>LoVo</td>
<td>Metastatic</td>
<td>Colon</td>
<td>Lymph Node</td>
<td>Human</td>
</tr>
<tr>
<td>*CT26-F1</td>
<td>Metastatic</td>
<td>Colon</td>
<td>Liver</td>
<td>Mouse</td>
</tr>
<tr>
<td>*CT26-FL3</td>
<td>Metastatic</td>
<td>Colon</td>
<td>Liver</td>
<td>Mouse</td>
</tr>
</tbody>
</table>

* Denotes the cell line is isogenic

As can be seen from this table the healthy control used was NIH/3T3 (ATCC® CRL-1658™), which is a mouse fibroblast cell line. Previous linear discriminant analysis has depicted that the cell line data was bimodal, or in two clusters. Figure 4.1 depicts a linear discriminant analysis done with the cell lines listed above using 3T3/NIH as a control.

![Figure 4.1: LDA array response with cell lines of varying metastatic potential. This analysis was a three-class classification based on metastatic potential (i.e., healthy, cancerous non-metastatic (lowly metastatic), cancerous metastatic (highly metastatic).](image)

Classification Accuracy = 97%
N=340
0.6% “false negative” rate
0.8% “false-positive” rate
Recall that LDA is a guided statistical method meaning that the computer program has the information regarding which metastatic potential each cell line falls into. The analysis done above in Figure 4.1 was a three-class classification indicating that each cell line would classify as one of three groups (healthy/normal, cancerous non-metastatic (lowly metastatic), and cancerous metastatic (highly metastatic). It can be seen when looking at the 3T3 data points (black diamonds) that there is a bimodal quality to the data.

Initially, there were a few different hypotheses for the data would appear this way. The first had to do with the cell line and whether this data was from the same cell stock. It was hypothesized that perhaps the cells being from two different stocks would cause the bimodal trend. It was also thought that perhaps this was due to variability between different researchers methodology. In order to answer these questions a significant amount of work was done to look through old experiments and compile all of the necessary information as well as the images from the incubation of 3T3 with the sensor array. It should be mentioned that the fluorescent intensity data collected from the images represented in Figure 4.1 was collected manually using Adobe® Photoshop. All of the data used in the lab now is found by running the images in MATLAB® through a program written by Dr. Andrew Greytak. With this in mind, all of the previous images used for this data were run through the program in MATLAB®.

There were additional cell lines added to the data set represented in Figure 4.1. It was decided that additional healthy cell lines would be added as well as addition replicated of 3T3. Other prostate and breast cell lines were added to give additional information to the statistical program. With this in mind, another researcher had 3T3 data
that was added to this set-AAV 3T3. This data set is represented as the green squares. Additionally, in order to put another healthy cell into this data set MCF 10A (ATCC® CRL-10317™) was added. This is a human mammary gland tissue.\textsuperscript{53} PC-3 (ATCC® CRL-1435™), a cancerous metastatic prostate cell line was also added. This cell line was derived from the bone, a metastatic site.\textsuperscript{54} MCF7 (ATCC® HTB-22™), a metastatic breast cell line derived from a pleura effusion.\textsuperscript{55}

Figure 4.2 depicts the LDA plot for a three-class analysis of the sensor array data from these cell lines. Recall that the three-class analysis is the three groupings of healthy, cancerous non-metastatic (lowly metastatic), and cancerous metastatic (highly metastatic). The green data points represent the healthy cell line data, while yellow data points represent the cancerous non-metastatic cell line data. The red data points represent the cancerous metastatic cell line data. The N=433 for this analysis and the leave one-out-classification accuracy was 78%. At this point in the lab only SLs 1, 3, 4, and 5 and thus that is the sensor array data used in this plot.

![LDA plot](image)

**Figure 4.2:** LDA plot of N=433 and the leave-one-out classification accuracy was 85%.
While the classification accuracy was relative high (78%), but lower than the analysis represented in Figure 4.1 (97%), this was not the focus of running the analysis. The goal of this analysis was to gain insight into the bimodal nature of the healthy cell lines. It was discovered that the bimodal data for 3T3 was actually two biological replicates done by the same researcher. The green triangles (SJ 3T3 Rep 1) and the green Xs (SJ 3T3 Rep 2) represent this data. Originally, it was thought that two different researchers generated this data. It was also discovered that the cell line data was from the same stock of 3T3. This ruled out the two hypotheses discussed previously. The additional 3T3 data added (AAV 3T3- represented by the green squares) also depicted some bimodal properties. The additional healthy cell line added (MCF 10A) clustered near one group of the 3T3 data. There was no literature precedent that explained the 3T3 data trend.

In order to further study the data set, training/test statistical analysis was performed using Statistica®. This is an application of linear discriminant analysis. This analysis involves having a portion of the data be a training set, i.e. a set of data that the computer knows information about, in our case which metastatic potential the cell line falls in. This would be the guided part of the analysis. The test set is a group of data that the computer program does not know to which group this belongs. The part of the analysis is unguided. These analyses can be done in a variety of different ways in varying the percentage of data that is the training and test set. Previous training/test analysis done in the lab used half of the data to create a model and the other half was put in as the test set. For the analysis that will be discussed below, a more stringent analysis was done. One cell line was removed, a model was created with the other cell lines (test set), and then
the cell line was reinserted to determine the classification accuracy. This was done in two different cases: three-class and two-class. Three-class is defined as healthy, cancerous non-metastatic (lowly metastatic), and cancerous metastatic (highly metastatic), while two-class is defined as healthy and cancerous.

Figure 4.3 represents the results from the training and test analysis for all cell lines in Figure 4.2. The blue colored bars are the three-class classification accuracies for the test set (one cell line), while the green colored bars are the two-class classification accuracies for the test set. To better explain this analysis one cell line, CT26, will be explained in detail. The training set for the CT26 analysis would be all cell lines besides CT26, while the test set is CT26. When the model created was three-class (healthy, cancerous non-metastatic, and cancerous metastatic), CT26 had test classification accuracy between 80-85%. When the model was recreated as a two-class classification (healthy and cancerous), this classification accuracy was between 95-100%. The analysis demonstrates that all classification accuracies improve from three-class to two-class. This result is often seen in the sensor array data. It is common in the data for the sensor array to have more overlap between the cancerous non-metastatic and cancerous metastatic data. As can be seen from this analysis 3T3 has a low classification accuracy. In referring to the data points in Figure 4.2 this was expected as the bimodal properties of 3T3 group a portion of the data with the cancerous non-metastatic data points.
Figure 4.3: Training/Test analysis for all cell lines represented in Figure 4.2. The training set was all cell lines except one, while the test set was that individual cell line. The three-class classification accuracies are colored blue, while the two-class classification accuracies are colored green.

To investigate how 3T3 affects the training/test analysis of the data another analysis was done for each cell line, however, this time 3T3 was left out of the training model. Figure 4.4 depicts the classification accuracies for this analysis. MCF 10A was left out because without 3T3 in the analysis there would not be another healthy cell line to put in the model when testing MCF 10A.
Figure 4.4: Class Training/Test analysis for most of the cell lines represented in Figure 4.2. The training set was all cell lines except one, while the test set was that individual cell line. The three-class classification accuracies are colored blue, while the two-class classification accuracies are colored green. 3T3 was left out of this analysis to look at its affect on the data and MCF 10A was removed because there would not be another healthy cell line in the model if this was removed to test.

It is interesting to note that all of the classification accuracies (except three-class PC3) increase when 3T3 is removed from the model. Also all of the two-class classification accuracies are 100% for every cell line analyzed.

Another researcher’s sensor array data from 3T3 was found and a training test analysis was done for this cell line with the cell lines included in Figure 4.2. The test data was all of the cell lines while the test data was this researcher’s 3T3 data.
Figure 4.5: Training test plot of 3T3 with the cell lines. As can be seen, the test data should overlap with the healthy data, however, it overlaps with the cancerous non-metastatic data points (yellow diamonds).

Being that this data was from the sensor array’s interaction with 3T3, a normal cell line the green triangles with black borders should have grouped with the other healthy cell lines (green triangles), however it can be seen that the test data overlaps with the cancerous non-metastatic (yellow diamonds). The classification accuracy for this analysis was 0%.

3T3 is commonly used as a control for a variety of studies, however, with this information the lab decided to move in the direction of using a healthy human colon cell line, CCD-841-CoN.45

4.2 MICROSCOPE STUDIES WITH FLUORESCENCE REFERENCE BEADS
At a certain point in lab issues began to arise with the microscope bulb and fluorescent microscope. For this reason, reference beads, FocalCheck™ Fluorescent Microspheres Kit, 15µm, were bought from Invitrogen™. These were reference beads that were mounted on slides. While the issues with the bulb and microscope were eventually
worked out, the reference beads were then used for a time study and bleaching study. The first study that will be discussed is the microscope time study. The information given when the pictures are run through MATLAB® is luminosity values from 0 to 255. The data is given in terms of the four channels: red, green, blue, and bright. Bright is a weighted average of the three other channels. When imaging beads, the optimal experimental set-up is that for a particular analyte that all SLs (1-9) are run at the same time. For instance, the exposure time for each image of each SL of CT26 is 500ms. The problem that can occur is that one SL is significantly brighter or dimmer than another. Occasionally this would lead to different exposure times. Recall that after 255 the camera is saturated. The time study was done for the reference beads from 500ms to 5s. This study was done in order to determine if the luminosity at 1s was the value at 500ms doubled. The result for this study is depicted in Figure 4.6.

![Figure 4.6: Time Study done with reference beads to investigate the change in luminosity with exposure time.](image-url)
As can be seen by the plot in Figure 4.6, the correlation between exposure time and luminosity is linear, it is not a slope of 1. In order to calculate the luminosity change with respect to time, the $y$-intercept has to be taken into account which would indicate that the amount of dye present would have to be known.

A luminosity study was also conducted with the reference beads. The intent of the study was to determine if the luminosity would stay constant. The study was run from 0 minutes to 5 minutes. The results are shown below in Figure 4.7.

![Figure 4.7: Luminosity study over 5 minutes elapsed. The luminosity remains consistent at least over 5 minutes.](image)

### 4.3 CONCLUSIONS AND FUTURE DIRECTIONS

Statistical analyses are powerful tools to understand the biochemistry behind the SL array response. This chapter demonstrated the usefulness of linear discriminant analysis with a training/test component in investigating a healthy cell line control, 3T3. It was determined that a healthy human colon cell line would most likely be the best option.
This chapter also discussed the use of reference beads to investigate the microscope. Being that chronologically most of this data occurred prior to the data collection in Chapter 2 and 3 there are a limited number of future directions. A healthy colon cell lines, CCD-841-CON, was used in subsequent experiments. While the microscope will always prove to be rate limiting in terms of data collection, current lab members are investigating the use of flow cytometry for use with the SL sensor array.
REFERENCES