

2017

Targeting Aberrant Glycosylation In Colon And Prostate Cancer With An Improved Synthetic Lectin Array

Tanya Hundal
University of South Carolina

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

 Part of the [Chemistry Commons](#)

Recommended Citation

Hundal, T.(2017). *Targeting Aberrant Glycosylation In Colon And Prostate Cancer With An Improved Synthetic Lectin Array*. (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/etd/4541>

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.

TARGETING ABERRANT GLYCOSYLATION IN COLON AND PROSTATE CANCER
WITH AN IMPROVED SYNTHETIC LECTIN ARRAY

by

Tanya Hundal

Bachelor of Science
Guru Nanak Dev University, 2009

Master of Science
Delhi University, 2011

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

2017

Accepted by:

John J. Lavigne, Major Professor

Mythreye Karthikeyan, Chair, Examining Committee

Qian Wang, Committee Member

Peisheng Xu, Committee Member

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

© Copyright by Tanya Hundal, 2017

All Rights Reserved.

DEDICATION

In memory of Dr. Maninder Singh Hundal: my first Chemistry teacher.

ACKNOWLEDGMENTS

When I joined this lab, I did not feel a day older than a child who was just learning how to walk. Ever since then, I have had some special people watching over influencing me and seeing me grow. I feel blessed to meet so many people who have been guiding beacons throughout my Ph.D. journey. The scope of this page would not cover how grateful I feel to have so many well-wishers and supporters. I would have to start by thanking Dr. Geeta Hundal, whose dream was that I get educated from the finest and achieve a Ph.D. one day. When I succeed, she would be the happiest human on the planet. I want to thank my younger brother Karan, who had to endure through all of my ‘mock-teaching classes’ growing up as kids. Little did I know that his passive reception would help me realize the love I have for teaching!

To my Ph.D. supervisor/mentor/Orgo-I “Rock star” Dr. John J. Lavigne, I humbly express my gratitude. I am so thankful for his guidance (on professional and on personal front) and patience (he winced every time I talked about boronic acids). He has taught me a valuable lesson on how Science is not all black and white and that everything in the middle is all that makes a difference.

To my committee members, Dr. Wang, Dr. Karthikeyan and Dr. Xu, for always educating me through discussions and challenging questions. To my teaching mentor, Dr. Sheryl Wiskur, for sharing her innumerable pedagogical skills with me. To my former lab mates, and colleagues Dr. Kathleen O’Connell, Dr. Erin Gatrone, Dr. Anna Veldkamp, for enduring through my bar graphs and just for being an amazing bunch. To my current

lab mates and friends, Rong, Steve, Ashley, Grace and Dan for their words of encouragement, humor and peanut butter cups.

I am very thankful to the close friends that I made here at USC, Naimah, Anusha, Shraddha, Bojidha and Vinya. Not only have they all been the greatest people to have as roommates and friends, but their personal journeys of thesis completion inspires this slug to finish as well! To all my close friends over the years: Naval, Neha, Varinda. Shivam and Ankur as they say here in the South, I thank all y'all!

ABSTRACT

Cancers of the colon and prostate, though treatable, necessitate early detection to improve patient outcomes. Current diagnostics, (visual methods or biopsies) besides being invasive, are subjective towards interpretation, thus decreasing accuracy. Alternatively, blood-based tests involving measuring of specific biomarkers (like CEA and PSA for colon and prostate cancers, respectively) are associated with high false-positive rates and are more useful for monitoring post-treatment patient health, thus driving efforts to identify better screening and diagnostic techniques.

Abnormal glycosylation of integral membrane and secreted glycoproteins is known to take place at the onset of many diseases, including cancer, and presents as the over, under or new occurrence of certain glycans. The aim of this study is to design synthetic lectins (SLs) that could discriminate cancer-associated glycans (CAGs) and to investigate global glycosylation changes associated with colon and prostate cancers. Further, the ability of an array of SLs to discriminate cells based upon their metastatic potential, demonstrates an alternative approach to detect colon and prostate cancers by looking at aberrant glycosylation changes rather than hunting for a specific biomarker.

Chapter 1 details on abnormal protein glycosylations. An account of how diols in glycans can be investigated using boronic acids mediated glycan sensors in the past. This follows introduction to Lavigne group peptide-boronic acid conjugates (Synthetic Lectins). A discussion of SL-array platform and its utility at discriminating proteins extracted from colon cell lines .

Chapter 2 describes various chemical modifications done to the SLs (peptide sequences and boronic acid moieties) in order to establish *structure and activity relationships* against purified glycoprotein analytes. Purified glycoproteins were used since the glycans displayed in them are also present in CAGs and have a positive disease correlation. The aim of these experiments is to gain an insight on chemical basis of SL-glycoprotein interactions. Several hypothesis were drawn on the design of new SLs. Positive charges on SLs helped pre-organization of SL-glycoprotein interaction.

Chapter 3 aims at incorporating new SLs into the pre-existing SL array to investigate colon and prostate cancer *in vitro*. The chemical make-up of the SLs that statistically contributed the most to discriminate cancer is evaluated. An extended SL array (combining old and new SLs), can discriminate normal, low metastatic and high metastatic states from secreted proteins of different prostate and colon cells with >99% accuracy. Metastatic potential of colon and prostate correlate with variation in i). number of Arginine (R) residues and ii). number of phenyl rings in SL peptide sequences.

Chapter 4 details on *quantifying sialic acid* content *in vitro* and investigating any correlations with the metastatic potential of colon and prostate cell lines. Further, several CAGs were classified using SL array with >99% accuracy. The positively charged amino acids (e.g., Arg) and those containing phenyl ring residues (e.g., Tyr, Phe) appear to be the principal factors involved in discriminating CAGs. Based on the relative importance that charged amino acids and phenyl boronic acids in SLs have at evaluating metastatic potential; several tissue specific SLs were found and their *peptide sequence homology* was studied.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGMENTS	iv
ABSTRACT	vi
LIST OF TABLES	xii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS.....	xvii
CHAPTER 1 INTRODUCTION AND BACKGROUND	1
1.0 OVERVIEW	1
1.1 COLON AND PROSTATE CANCER STATISTICS	1
1.2 COLORECTAL AND PROSTATE CANCER DETECTION AND DIAGNOSIS	2
1.3 ABERRANT GLYCOSYLATION: HALLMARK OF CANCER	4
1.4 LECTINS: CARBOHYDRATE BINDING PROTEINS	6
1.5 INTERACTIONS BETWEEN BORONIC ACIDS AND DIOLS	7
1.6 ADVENT OF BORONIC ACID FUNCTIONALIZED GLYCAN SENSORS	10
1.7 CROSS-REACTIVE SYNTHETIC LECTIN ARRAYS	14
1.8 OUR APPROACH: BORONIC ACID FUNCTIONALIZED SYNTHETIC LECTIN	15
1.9 PROPOSED SYNTHETIC LECTIN AND GLYCAN BINDING	16
1.10 SYNTHETIC LECTINS AND THEIR BINDING WITH CELL LINE PROTEINS	19

1.11 NON-COVALENT PROTEIN INTERACTIONS IN AQUEOUS ENVIRONMENT	22
1.12 REFERENCES	26
CHAPTER 2 : INVESTIGATION OF STRUCTURE-ACTIVITY RELATIONSHIP BETWEEN SYNTHETIC LECTINS AND GLYCO-PROTEINS	32
2.0 OVERVIEW	32
2.1 VARIATION OF INTRA-DIAMINO BUTANOIC ACID DISTANCE.....	33
2.2 VARIATION ON SL5 BY ALTERING BORONIC ACID ATTACHMENTS....	37
2.3 COMPARING DISSOCIATION CONSTANTS OF SL5 AND ITS MUTANTS	41
2.4 PEPTIDE SEQUENCE MUTATIONS ON POLAR AND NON-POLAR AMINO ACIDS.....	43
2.5 PEPTIDE SEQUENCE MUTATIONS ON POSITIVELY CHARGED AMINO ACIDS.....	47
2.6 CONCLUSIONS.....	51
2.7 FUTURE DIRECTIONS	53
2.8 EXPERIMENTAL PROCEDURES	55
2.9 REFERENCES	65
CHAPTER 3 : EMPLOYING DESIGNED SYNTHETIC LECTINS IN ARRAY TO EVALUATE METASTATIC POTENTIAL OF HUMAN COLON, AND PROSTATE CANCER <i>IN-VITRO</i>	66
3.0 OVERVIEW	66
3.1 INVESTIGATING COLON CANCER USING EXTENDED LECTIN ARRAY AND PROTEINS EXTRACTED FROM COLON CANCER CELL MEMBRANES	68
3.2 INVESTIGATING COLON CANCER USING EXTENDED LECTIN ARRAY AND SECRETED PROTEINS FROM COLON CELLS	72
3.3 IMPACT OF MANY LECTINS IN SL ARRAY ON TOTAL VARIANCE OF PCA MODEL.....	83
3.4 INVESTIGATING ROLE OF 2-PHENYL BORONIC ACIDS IN CLASSIFYING COLON CELLS ACCORDING TO THEIR METASTATIC POTENTIAL	84

3.5 INVESTIGATING PROSTATE CANCER USING EXTENDED LECTIN ARRAY AND SECRETED PROTEINS FROM PROSTATE CELLS.....	88
3.6 CONCLUSIONS.....	95
3.7 FUTURE DIRECTIONS	98
3.8 EXPERIMENTAL PROCEDURES.....	99
3.9 REFERENCES	108
CHAPTER 4 : <i>IN-VITRO</i> CHEMICAL TARGETS OF EXTENDED SL ARRAY AND ITS STRUCTURAL CORRELATIONS WITH CANCER TARGETS	111
4.0 OVERVIEW	111
4.1 SIALIC ACID QUANTIFICATION OF COLON AND PROSTATE CELLS...	112
4.2 INTERACTION OF SL-ARRAY WITH SECRETED COLON PROTEINS THAT CONTAIN OR ARE DEVOID OF SIALIC ACIDS.....	119
4.3 RESPONSE OF CANCER ASSOCIATED GLYCANS TO SL ARRAY	122
4.4 APPLICATIONS OF SL ARRAY TO DISCERN HUMAN PROSTATE TISSUE SAMPLES.....	127
4.5 DESIGNING FUTURE SL TARGETS FOR COLON AND PROSTATE CANCER	131
4.6 CONCLUSIONS.....	142
4.7 FUTURE DIRECTIONS	144
4.8 EXPERIMENTAL PROCEDURES.....	145
4.9 REFERENCES	156
BIBLIOGRAPHY:.....	157
APPENDIX A : LC-MS OF PEPTIDE SEQUENCES OF SELECT SLS.....	167
APPENDIX B : 22-UNIT SL ARRAY	179
APPENDIX C : TOP SEVEN SLS FOR ALL TISSUES TOGETHER	180
APPENDIX D : BAR GRAPHS OF OTHER 5 CAGS WITH 20 SLS	181

LIST OF TABLES

Table 1.1: Synthetic Lectins (SLs) with their peptide sequences that show binding with the proof of concept glycoproteins.	18
Table 1.2 Colon cell lines used for SL binding analysis; listed is the species, cell type and origin of the cell lines.	20
Table 2.1: Representing “SL5 positional mutants” to investigate impact of variation of intra-(Dab)* distance on SL5 binding with proteins.....	34
Table 2.2: Showing “SL5 Boronic acid mutants’	38
Table 2.3: showing few SL5 and its mutants involved in K_d investigation using PSM... ..	41
Table 2.4: Showing alanine mutations on polar and non-polar AA in SL5.....	43
Table 2.5: Showing SL5 Arginine (R) mutants.	48
Table 3.1: Designed SLs after structure and activity relationship evaluation.	67
Table 3.2: Human colon cell lines used with extended lectin array.	69
Table 3.3: SLs most significant at classifying proteins from colon cell membrane according to their metastatic potential.	69
Table 3.4: All 20 SLs in extended lectin array.	73
Table 3.5: SLs most significant at classifying proteins secreted from colon cell lines with metastatic potential.	75
Table 3.6: Cancer associated glycans in malignant tissues.....	89
Table 3.7: Human prostate cell lines used with extended SL array.....	89
Table 3.8: Showing most significant SLs at classifying proteins secreted from prostate cells according to their metastatic potential.....	91
Table 4.1: Illustrating total sialic acid residue count of secreted proteins extracted from media of cell lines.	114

Table 4.2: Showing Malignancy specific SL 'hits'. 141

LIST OF FIGURES

Figure 1.1: Some examples of cancer-associated glycans (CAGs) that are often involved in aberrant glycosylation during oncogenesis and disease progression. ^{14, 15, 20}	4
Figure 1.2: Showing two major types of protein glycosylation.....	5
Figure 1.3: Glycosylation reactions catalyzed by the action of glycosyltransferase causing several regulatory functions. ²¹	7
Figure 1.4: Interaction between boronic acids and 1,2- or 1,3- <i>cis</i> diols commonly present in glycans.	8
Figure 1.5: The equilibrium between phenylboronic acid and a diol lies towards boronate ester formation (2 &4) in basic aqueous solution.	8
Figure 1.6: Showing proposed equilibrium by Anslyn group by incorporating an aminomethyl group <i>ortho</i> to phenylboronic acid and facilitating binding of phenyl boronic acid with diols at physiological pH.	9
Figure 1.7: Structures of boronic acid functionalized sensors used for glycan sensing. (a) Yoon and Czarnik first fluorescent probe, (b) Lakowicz group's Intramolecular charge transfer (ICT) sensors, (c-d) Shinkai's and Wang's group (e) Photoinduced electron transfer (PET) sensors, (f-g) James' group electrochemical sensors.....	11
Figure 1.8: Animated view of cross-reactive sensors/receptors arranged into a sensor array.	14
Figure 1.9: General scheme of synthetic lectin (SL) attached to a polymeric resin bead.	16
Figure 1.10: Schematic representation of Boronic acid and SL interaction with a glycan or glycoprotein.	17
Figure 1.11: Detected response from SL1-5 binding with four purified proteins as measured by fluorescence.	19
Figure 1.12: Discrimination of healthy (◆), cancerous low metastatic (●) and metastatic (▲) colon cell lines using SL array shown by guided statistical grouping using linear discriminant analysis score plot.....	21
Figure 1.13: (A) Parameters used to identify CH- π interactions:	23

Figure 2.1: Showing impact of changing (Dab)* distance on SL5 with four purified proteins.....	35
Figure 2.2: Impact of mutating boronic acid in SL5 on binding with proteins.	39
Figure 2.3: Response of SL5 and its mutants with varying PSM concentration.	42
Figure 2.4: Showing response of SL5 peptide sequence mutants with four proteins.	45
Figure 2.5: SL5 Arginine positively charged mutants.....	49
Figure 2.6: Showing boronic acid mutants, which can be evaluated either by complementing with or by replacing 2-PBA.	54
Figure 3.1: Guided statistical plot (LDA) generated using extended lectin array with proteins extracted from cell membranes of four different colon cell lines.	70
Figure 3.2: Bar-graph showing output of extended lectin array with extracted proteins extracted from colon cell membranes of four different cell lines.	71
Figure 3.3: Unguided (PCA) plot showing output generated by extended lectin array with four colon cell line proteins (both cell membrane and cell-secreted) having different metastatic potentials.....	74
Figure 3.4: Showing unguided/PCA plot using 20-unit SL array on secreted proteins from colon cells.	76
Figure 3.5: Showing component loading values of PCA output generated by SL array containing 20 SLs with colon cell secreted proteins.....	79
Figure 3.6: Showing A: unguided/PCA plot of 20-unit SL array with secreted proteins from colon cells being impacted by B: Removal of Area ② SLs from PCA model; C: Removal of Area ① SLs from PCA model.	80
Figure 3.7: Showing normalized fluorescence intensity signal of SL7 and SL11-R8A over SL5-R1,5,11A (negative Arginine control SL).....	82
Figure 3.8: Showing unguided/PCA plot of SL array with secreted proteins from colon using array having A: 20 SLs. B: five “key discriminatory” SLs.....	83
Figure 3.9: LDA plots showing direct impact of boronic acid moiety in SLs included in array at classifying different protein samples from colon cells.	86
Figure 3.10: Alizarin red dye binding with SL5 and SL5-Dab.....	87
Figure 3.11: Showing A: unguided/PCA plot of extended 20-SL array with secreted proteins from prostate impacted by B: Positively charged SLs (charged +4 to +5) &	

SLs with lower number of phenyl rings (0-2); C: SLs with greater number of phenyl rings (4) and lower positive charge (0 to +2).....	92
Figure 3.12: Showing guided/LDA 3-class plots based on cell type of secreted proteins from prostate cell lines.....	94
Figure 3.13: Illustrating 3-D PCA plot of normal (green) and metastatic (red) cell lines.	98
Figure 4.1: Showing amount of total sialic acid isolated from colon cell secreted proteins on top of a PCA plot generated using an SL array.	116
Figure 4.2: Showing amount of total sialic acid isolated from prostate cell secreted proteins on top of a PCA plot generated using an SL array.....	117
Figure 4.3: Illustrates SL array mediated 4-class guided (LDA) plot of isolated proteins with and without sialic acid residues.	120
Figure 4.4: Illustrates unguided (PCA) plot of cell secreted proteins with and without sialic acids responding to a 20-unit SL array.....	121
Figure 4.5: Selective binding of Synthetic Lectins (SL).....	123
Figure 4.6: Guided statistical output of SL array with various CAGs.....	126
Figure 4.7: Showing unguided/ (PCA) plot discerning four patient-matched tissue samples having prostate cancer using 22 unit SL-array.....	129
Figure 4.8: Illustrating Split and Pool combinatorial method.....	134
Figure 4.9: Flowchart showing the incubation of different SL libraries with different healthy and metastatic dual-labeled protein analytes.....	135
Figure 4.10: Schematic depicting Edman degradation.	136
Figure 4.11: Binning chart representing population diversity of bead fluorescent intensity in 3-BA and 2-BA SL libraries	138
Figure 4.12: Binning chart representing population diversity of bead fluorescent intensity in 3-BA and 2-BA SL libraries	139
Figure 4.13: SL dual-dye library screens to investigate racial health disparities between African American (AA) and Euro-American (EA) prostate cancer patients.....	140

LIST OF ABBREVIATIONS

2-PBA	2-formyl phenyl Boronic Acid
4-PBA	4-formyl phenyl Boronic Acid
AA.....	Amino Acid
A.....	Alanine
BA.....	Boronic Acid
(BOC) ₂ O	Di-tert-butyl dicarbonate
BSA.....	Bovine Serum Albumin
BSM	Bovine Submaxillary Mucin
CAG	Cancer Associated Glycan
CHCA	α -Cyano-4-hydroxycinnamic acid
DMF.....	N,N'-Dimethylformamide
Dab.....	Diaminobutanoic Acid
(Dab)*	Diaminobutanoic Acid plus 2-formyl phenyl boronic acid
DSR.....	Discosoma sp. Red
F	Phenylalanine
FITC.....	Fluorescein isothiocyanate
Fmoc	Fluorenylmethyloxycarbonyl
G.....	Glycine
GFP.....	Green Fluorescent Protein
H&E	Haemotoxylin and Eosin staining

H ₂ O	Water
I	Isoleucine
ivDde	1-(4,4-Dimethyl-2,6-dioxocyclo-hexylidene)-3-methylbutyl
K	Lysine
K _d	Dissociation constant
L	Leucine
Le ^a	Lewis A
Le ^x	Lewis X
M	Methionine
MALDA	Linear Discriminant Analysis
MALDI	Matrix Assisted Laser Desorption/Ionization
MeOH	Methanol
MRBB	Methionine, Arginine, β-Alanine, β-Alanine
MS/MS	Mass Spectrometry/Mass Spectrometry
N	Asparagine
NBF	Neutral Buffered Formaldehyde
NaBH ₄	Sodium Borohydride
OVA	Ovalalbumin
PBA	Phenylboronic Acid
PBS	Phospho Buffer Saline
PCA	Principal Component Analysis
PSM	Porcine Stomach mucin
Q	Glutamine

R.....	Arginine
S.....	Serine
SL.....	Synthetic Lectin
sLe ^a	sialyl Lewis A
sLe ^x	sialyl Lewis X
T.....	Threonine
TFA.....	Thomsen Friedenreich Antigen
TIS.....	Triisopropylsilane
V.....	Valine
W.....	Tryptophan
Y.....	Tyrosine

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.0 OVERVIEW

Cancer, the rampant growth and invasion of abnormal cells in the body, is one of the leading causes of death not only in the United States but also worldwide. Colon cancer ranks as the third leading cause of death for both men and women. Whereas, prostate cancer is the second leading cause of death from cancer in men, here in the States. It is a well-known fact that the early diagnosis is a paramount for an optimal patient outcome. Post-translational modifications causing an advent of abnormal changes in glycosylation during the onset and progression of the disease, is a cancer hallmark. This anomalous glycosylation allows several glycans to serve as biomarkers. This chapter will shed some light upon the current cancer diagnostics and the necessity for the new ones. It will then discuss how polymeric boronic acid-peptide conjugates: also known as synthetic lectins (SLs) develop as glycan sensors. Later, how an array of such cross-reactive sensors are useful in designing a novel cancer diagnostic.

1.1 COLON AND PROSTATE CANCER STATISTICS

Cancer is one of the leading causes of morbidity and mortality worldwide, with approximately 14 million new cases in 2012.¹ American Cancer Society estimates that there will be 1,688,780 new cases of cancer and reports 600,920 deaths in the year 2017.

² In the United States, Colorectal cancer is the third deadliest in both sexes and prostate

cancer is the second deadliest of the cancers in men. In 2017, there will approximately 50,000 deaths and 27,000 deaths due to colorectal and prostate cancer respectively.²

1.2 COLORECTAL AND PROSTATE CANCER DETECTION AND DIAGNOSIS

WHO funded International Agency for Research on Cancer establishes that cancer mortality can reduce, provided there is an earlier detection and treatment available. Approximately 30-50% percent of cancers are preventable through screening and early detection strategies. There is a high five-year survival rate in case of colon cancer of 90% when the diagnosis reports the disease is localized within the bowel walls (Stage I). The rate of survival plummets to 68%, once cancer metastasizes in to the lymph nodes (stage III): thus necessitating early detection.³

There are some non-invasive colorectal cancer screening techniques too (gFOBT, FIT, stool DNA test), which look for blood in the stools, collectively have an accuracy rate of 38-51% hence they present high false-positive rates.⁴ On the other hand, invasive visual methods of screening, such as colonoscopy are also present.⁵ Colonoscopy although quite sensitive yet it has a set of complications, e.g. bleeding, lacerations, increased risk of infection and high price-point.⁶⁻⁸ Yet, colonoscopy provides a more accurate detection than the non-invasive procedures like FIT and gFOBT albeit associated with high false-negative rates.⁹ While the FIT has success with identifying advances neoplasms, the main disadvantage with the FIT over colonoscopy is that it has higher false-negative results in detecting T1 cancer versus T2- T4.¹⁰

There are invasive diagnostic techniques to determine prostate cancer, like Digital Rectal Exam (DRE) and Trans Rectal Ultrasound (TRUS). These involve monitoring abnormalities in the shape and size of the gland using touch and vision. The disadvantage

of TRUS is its invasive nature, leading to morbidity related to diagnosis due to infections, hematuria and haemospermia.¹¹ DRE lacked sensitivity in general especially at the onset of prostate cancer.^{12, 13} Increasing the specificity and sensitivity, while at the same time reducing the ‘invasiveness’ of a diagnostic procedure, has been a driving factor in the search for colorectal cancer biomarkers.

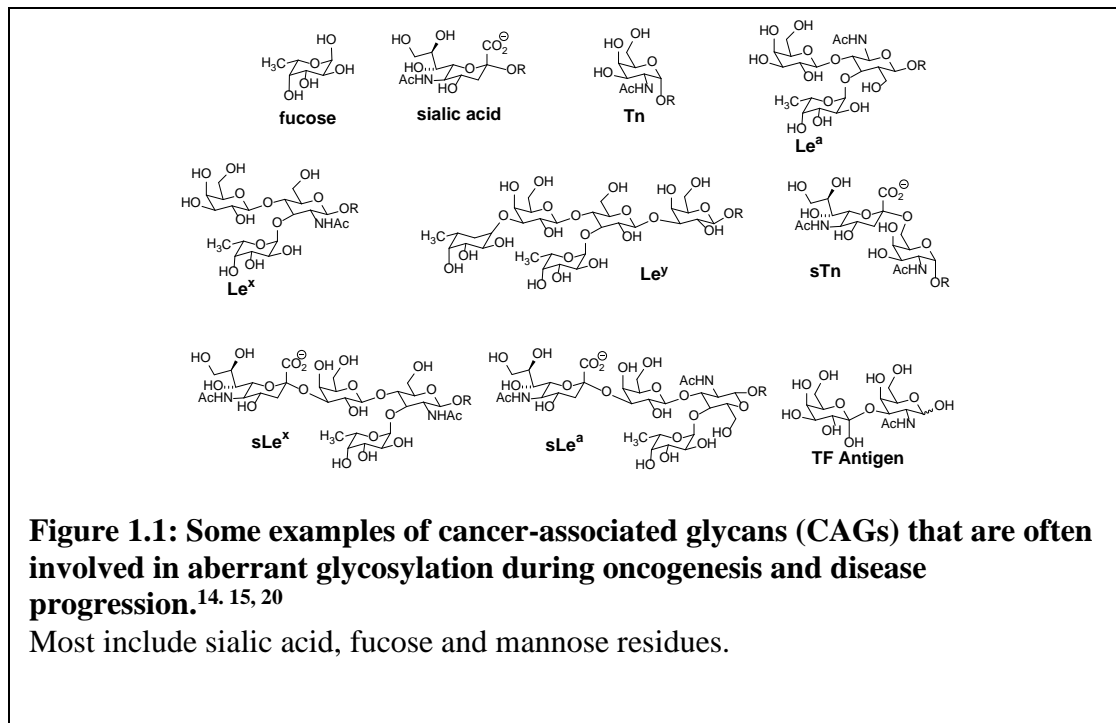
The other method of diagnosis is with use of biomarkers to diagnose and stage cancers. For example, Carcinoembryonic antigen (CEA) is a glycosylated cell-surface glycoprotein and in healthy patients, CEA levels can range from 2.5 to 5 ng/mL. The correlation of colorectal cancer with an increased CEA levels in patients with Stage I cancer is only 4%, while in patients with stage II cancer is 25%.¹⁴ Similarly there is CA 19-9 test, which measures the antibody against sialylated-Lewis a (sLe^a) (which typically increases in colorectal cancer) showed lack of disease specificity Both CEA and CA 19-9 tests are associated with high false-positive rates of 16 and 60% respectively.¹⁵ In 2000, the American Society of Clinical Oncology (ASCO) recommended that the CEA and CA 19-9 tests no longer be used as a diagnostic tool, but rather be used to monitor disease progression.¹⁶

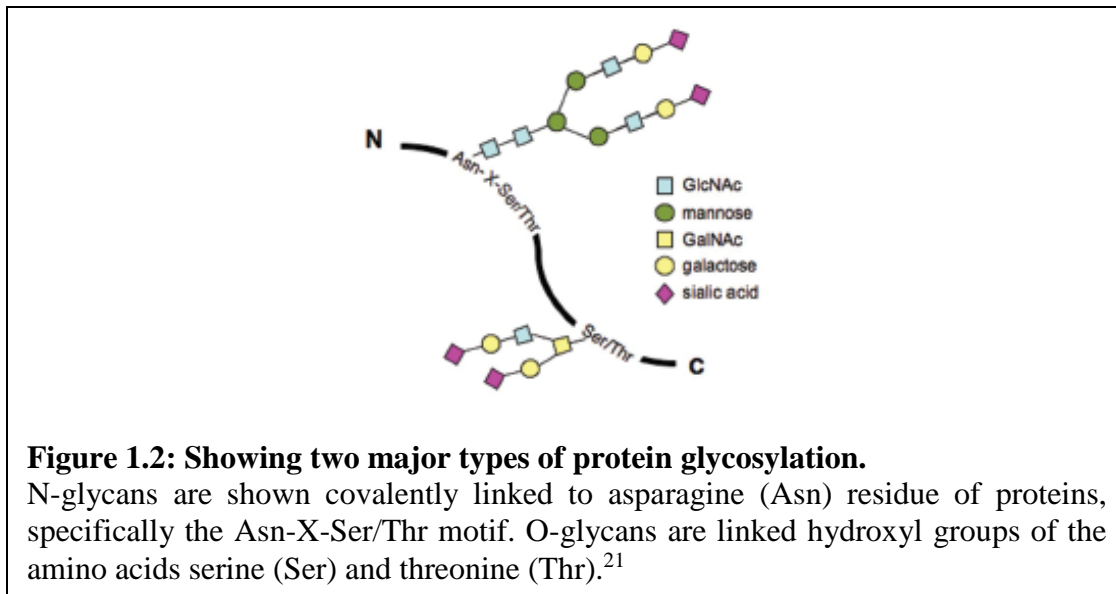
Along with DRE and TRUS, monitoring of a serum-based glycoprotein called Prostate-Specific Antigen (PSA) completes the diagnostic triad for prostate cancer.¹⁷ PSA levels above 4.0 ng/mL, meant some abnormality and the patient is recommended biopsy.¹⁸ Besides prostate cancer, age, level of activity and inflammation are also the reasons for elevated PSA, thus deeming it non-specific. PSA has notoriously high false-positive rate and only 25% of the men that undergo biopsies actually have cancer.¹⁹ Not only are cancer diagnostics associated with high false positive and false-negative rates,

most are also ineffective in early stages of cancer, when the disease is most treatable. This calls for a development of a detection procedure with high positive correlations and minimal invasion. Hence, the long-term goal of this project is the development of a sensitive and preferably easy to use sensor, with the ability to detect colon and prostate cancer related targets.

1.3 ABERRANT GLYCOSYLATION: HALLMARK OF CANCER

Aberrant glycosylation (term coined by Meezan et. al. in 1969) of integral cell membrane proteins and secreted glycoproteins occurs during carcinogenesis.^{20, 21} Over 70% of proteins (both membrane-bound and secreted) undergo glycosylation.²² Post-translational glycan modifications associated with cancer are termed as Cancer-Associated Glycans (CAGs) (**Figure 1.1**). Aberrant glycosylation also takes place at the onset of many diseases including cancers, inflammation, as well as lysosomal storage diseases. The glycans on the glycoproteins can be under, over, or neo-expressed.





This altered glycosylation occurs on both the membrane-bound proteins as well as the secretory proteins.²³ Aberrant glycosylation process takes place at the onset of cancer and continues as the disease progresses.²⁴ There are two sites for protein glycosylation: N-linked and O-linked glycosylation (**Figure 1.2**) For instance, sialyl Lewis X (sLe^x) and sialyl Lewis A (sLe^a) are CAGs that have increased abundance in colon cancer cells, they are few of the glycans looked at with the SL array.²³ There never is one CAG associated with one type of tissue malignancy, but a mélange of glycosylations alter, which change the overall glycan makeup by changing the amount of sialylation or fucosylation.²⁵ Taking this as a phenotypical characteristic of cancerous tissue.^{26, 27}, cells are probed with lectins (carbohydrate binding peptides), specifically synthetic lectins (SLs) to detect abnormal glycosylation.^{28, 29} For instance, sialyl Lewis X (sLe^x) and sialyl Lewis A (sLe^a) are CAGs that have increased abundance in colon cancer cells, they are few of the glycans looked at with the SL array.²⁹ While many glycoproteins are used to detect cancer (PSA and CEA), these diagnostics rely on detecting the amount of the glycoproteins themselves and not the changes in the glycan structures of the

glycoproteins. Harnessing the overall change in glycosylation pattern can prove to be a useful tool for cancer detection.

1.4 LECTINS: CARBOHYDRATE BINDING PROTEINS

CAGs can also be monitored using antibodies³⁰ and natural lectins^{31, 32}. 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 cell³¹, play a role in protein folding and cell-cell interaction.³³ They can also be involved in cellular functions like cell adhesion, and cell migration.²¹ **Figure 1.3** shows the impact of glycosylation on E-cadherin and integrin (common glycoproteins involved in cell adhesion). Different type of glycosylations instigate different form of cellular functions. Lectins have high selectivity for a particular target, natural lectins and natural lectin arrays have been of interest in research in a variety of research areas (endocytosis, adhesion, EMT), but, we will limit this discussion to cancer-related areas. Studies are looking at lectins as diagnostic tools. One such study investigated natural lectin microarrays as a means for distinguishing glycosylation patterns in serum from 24 different patients (model including 10 normal patients, 8 patients with chronic pancreatitis and 6 patients with pancreatic cancer).

This microarray included five different natural lectins, however, due to the specificity of lectins only five 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 (including glycoprotein enrichment followed by RP-HPLC to isolate glycoproteins before they are blotting them on a microarray and analyzed using lectins). This study also found several glycoproteins that over-expressed sialylation and fucosylation.^{34, 35}

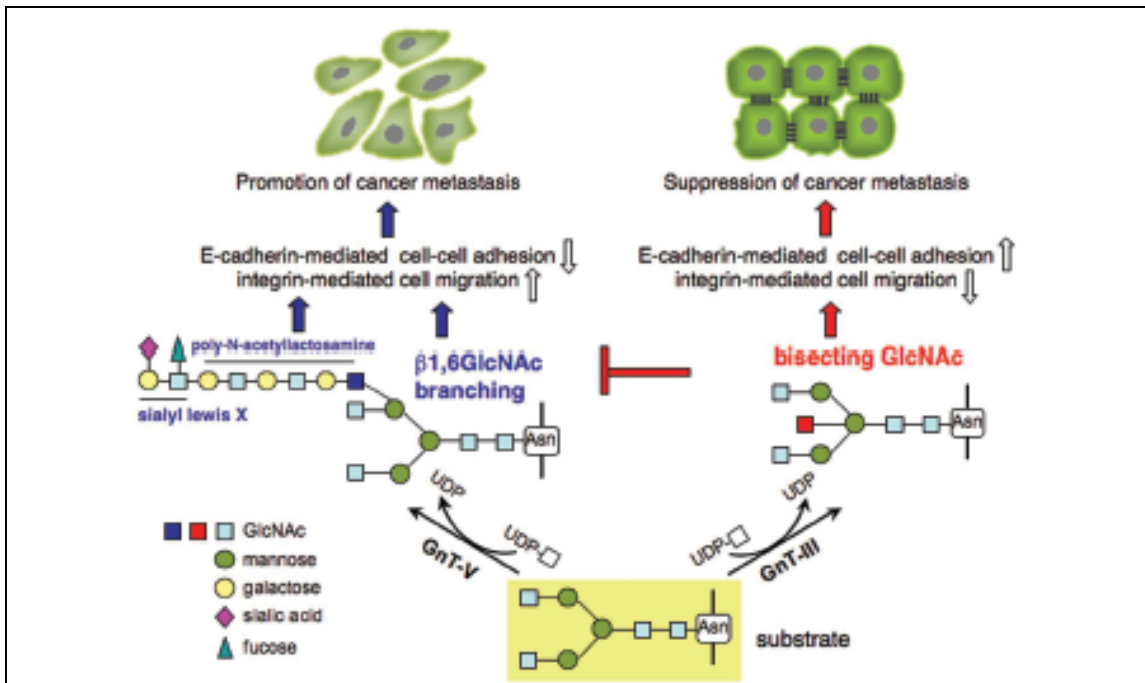


Figure 1.3: Glycosylation reactions catalyzed by the action of glycosyltransferase causing several regulatory functions.²¹

Enhanced expression of GnT-V in epithelial cells results in a loss of cell-cell adhesion, increasing integrin-mediated cell migration. In contrast, overexpression of GnT-III strengthens cell-cell interaction and downregulates integrin-mediated cell migration, which may contribute to the suppression of cancer metastasis.

Natural lectins are expensive, unstable and give incomplete results. Natural lectins tend to lose binding specificity when they are extracted out of their native protein conformation. Besides this, the number of glycan motifs and the number of glycan linkages severely outweigh the number of known natural lectins. For instance, the accuracy of an assay developed for Carcinoembryonic Antigen (CEA) is only 4% and 25%, respectively in the diagnosis of colorectal cancer of stage I and II respectively, thus making it an unsuitable candidate for screening and more useful for prognosis of colon cancer.¹⁴

1.5 INTERACTIONS BETWEEN BORONIC ACIDS AND DIOLS

In an attempt to cross-over the previously stated disadvantages of using natural lectins, boronic acids have been utilized in several different glycan sensors because

boronic acids are known to form covalent yet reversible boronate esters with 1,2- and 1,3-diols, which are found on glycans, to produce cyclic boronate esters. The interaction between boronic acids and diols commonly found on glycans

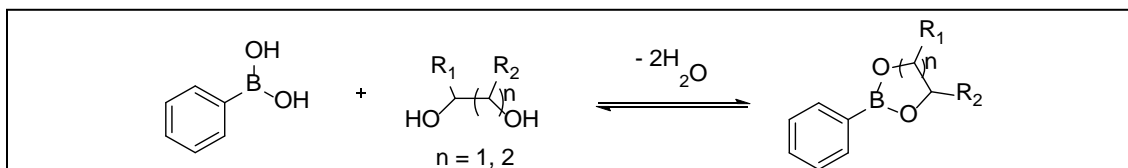


Figure 1.4: Interaction between boronic acids and 1,2- or 1,3-*cis* diols commonly present in glycans.

This leads to formation of proposed reversible and covalent boronate ester bond between boronic acid found on SLs and *cis* (1, 2- or 1, 3-) diols of the CAGs.

is illustrated (**Figure 1.4**). When using boronic acids in glycan sensors, it is important to understand what influences the boronic acid-diol interaction. The equilibrium exists between phenylboronic acid and a diol in solution and illustrated in (**Figure 1.5**).

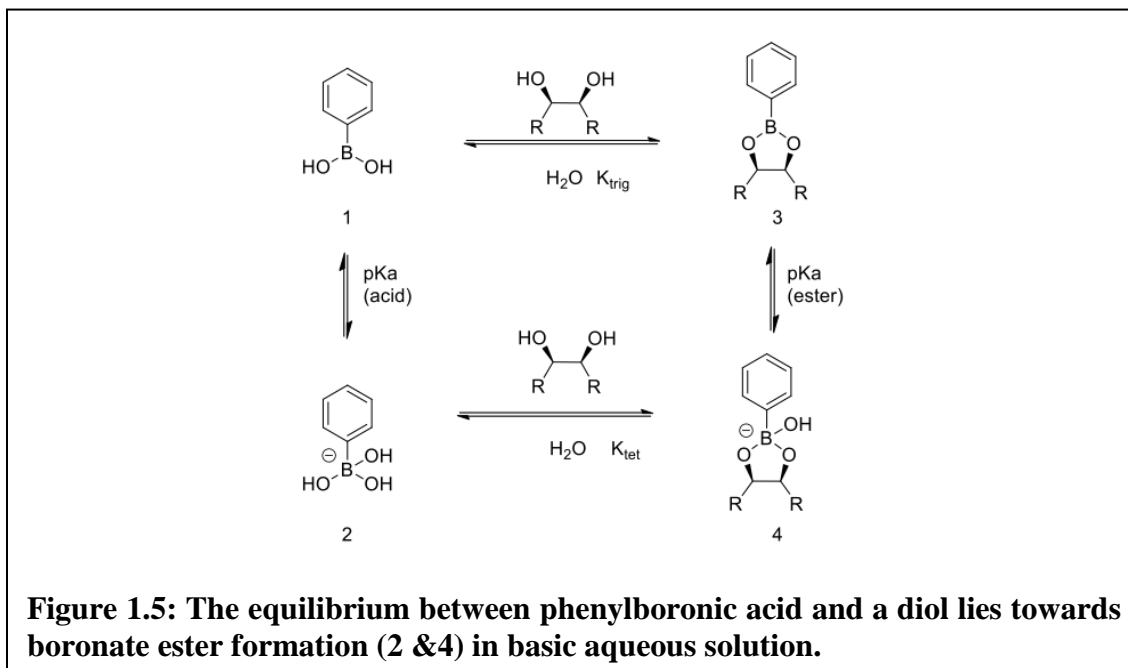


Figure 1.5: The equilibrium between phenylboronic acid and a diol lies towards boronate ester formation (2 &4) in basic aqueous solution.

The equilibrium is very dependent on pH. At lower pH, the acid and diol stay unbound since any boronic ester that forms, contains an empty p-orbital in Boron. At higher pH, (when pH increases more than pK_a of phenylboronic acid: 8.8) boronic acid

stays as a boronate anion (in **Figure 1.5**, $1 \rightarrow 2$ or $3 \rightarrow 4$). The boronate anion has a tetrahedral geometry, making the bond angles around the boron similar to the bond angle of the cyclic boronate ester. Therefore, at physiological pH, phenylboronic acid does not form boronate ester and this limits its use in carbohydrate detection sensors.³⁶

In order to use phenyl boronic acids on sensors at physiological pH, the Anslyn group investigated intramolecular boron-nitrogen interactions. In neutral aqueous solutions, the incorporation of an *ortho*-aminomethyl group on to phenylboronic acid facilitated boronic ester formation with diols.³⁷ The amino methyl group interacts with the empty p-orbital on the boron transforming the geometry into the favored tetrahedral conformation.

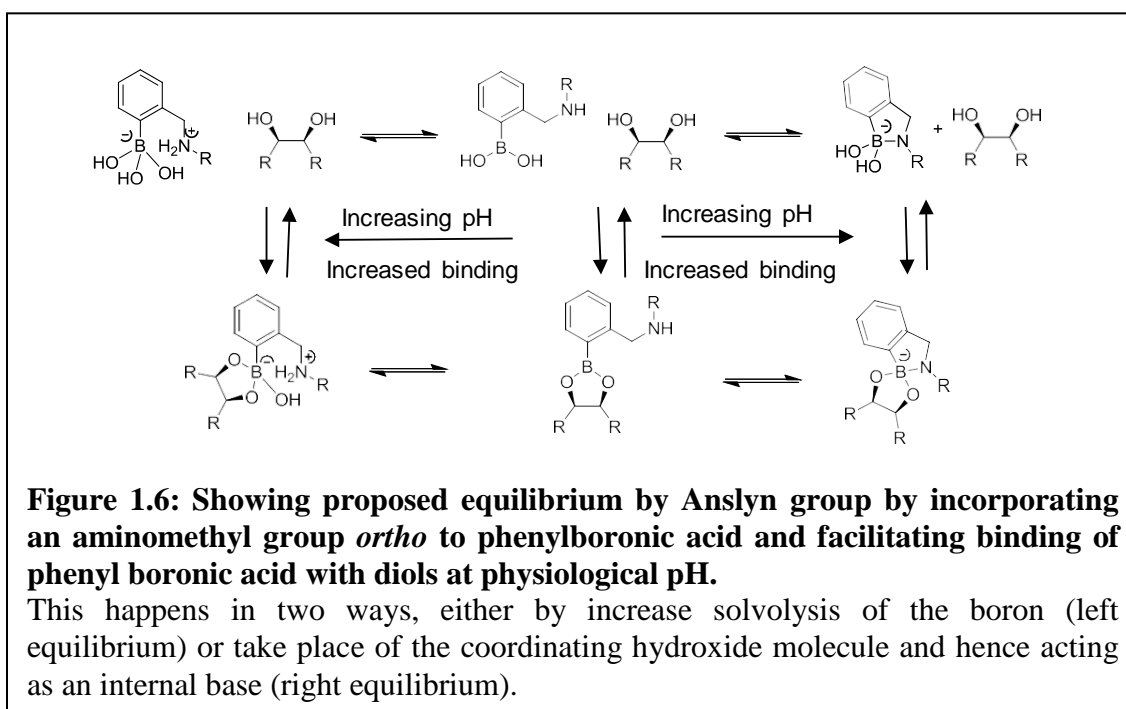


Figure 1.6: Showing proposed equilibrium by Anslyn group by incorporating an aminomethyl group *ortho* to phenylboronic acid and facilitating binding of phenyl boronic acid with diols at physiological pH.

This happens in two ways, either by increase solvolysis of the boron (left equilibrium) or take place of the coordinating hydroxide molecule and hence acting as an internal base (right equilibrium).

The Anslyn group proposes two reasons for this increase in boronate ester formation. First, the neutral aqueous stability of the boron is promoted because the nitrogen can act as an internal Lewis base and increasing boron's solvolysis. Second, due to this nitrogen behaving as a base, it donates lone pair to the empty p-orbital to form a weak co-

ordinate/dative bond between the nitrogen and the boron. Both of these explanations facilitate the boronate ester formation at physiological pH (**Figure 1.6**).³⁷

1.6 ADVENT OF BORONIC ACID FUNCTIONALIZED GLYCAN SENSORS

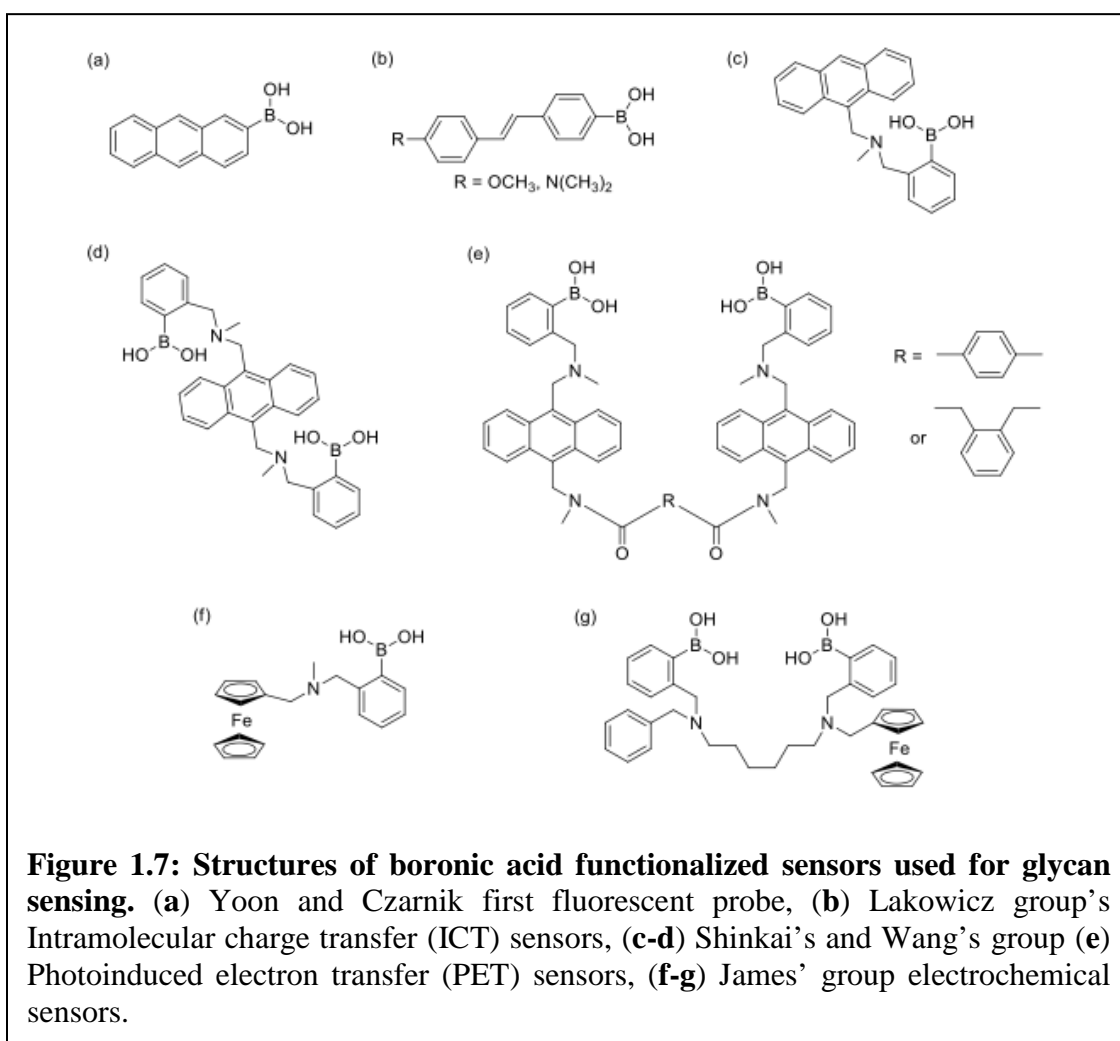
Over the years, several phenylboronic acid functionalized sensors were developed for different types of saccharide sensing, specifically keeping an interest of detection of D-glucose. It is mentioned in previous sections, that sensing glycans can be valuable when trying to diagnose and prognose cancer. The specific recognition of boronic acid for cis-diols provides selectivity in sensing. Some of the structures of different phenyl boronic acid functionalized sensors that employed fluorescent, colorimetric, and electrochemical methods are in (**Figure 1.7**).

Yoon and Czarnik generated the first aqueous fluorescent chemo-sensing probe using anthrylboronic acid (**Figure 1.7a**).³⁸ When the anthrylboronic acid was bound to fructose, the anionic boronate ester formed and fluorescence intensity decreased. The resulting cyclic ester of anthrylboronic acid and fructose had pKa of 5.9. The causation of decrease in fluorescence intensity was proposed to be by the boronate ester formation, which quenched the fluorescence due to electron transfer from anthracene ring.

Several other groups began to work on development of fluorescent glycan-sensors with similar intramolecular charge transfer (ICT) complexation. For example, boronic acid substituted stilbenes were synthesized as glycan sensors by DiCesare-Lakowicz group (**Figure 1.7b**).³⁹ The boronate anion acted as the electron-donor and the photo-induced CT would complete when the fluorophore contained an electron-withdrawing R group (e.g. -CN). Therefore, whenever the glycan formed boronate ester, the resulting boronate anion would register blue shift due to CT. When the boronic acid converted to the

boronate ester in the addition of a sugar at high pH, the boron does not accept electrons anymore and this loss of charge transfer disrupts the fluorescent properties.

Shinkai's group was the first to design Photoinduced Electron Transfer boronic acid sensors for sugars (**Figure 1.7c**).⁴⁰ They introduced N-methyl-o-(aminomethyl) appendage on to the phenylboronic acid as the recognition motif and connected to a fluorophore anthracene through a methylene spacer. The amine being *ortho* to the boronic acid allowed for the interaction between the boronic acid and diols at neutral pH.



The amine-boron interaction, besides assisting formation of boronate ester also dictated the molecule's ability to partake in fluorescence. Without the sugar binding, the

sensor remains in an “off” state due to the nitrogen’s lone pair. While binding the sugar, the nitrogen coordinates with the boron, causing formation of boronate ester, which prevented electron transfer and induced fluorescence quenching. These sensors exhibited some selectivity for sugars with it being the selective for D-Fructose and D-Glucose equally. The molecule c in **Figure 1.7** would lose selectivity when the amounts of boronic acid or the geometry around the boron was changed. Several other PET based B-N molecules were made where **Figure 1.7d** represents yet another example of an anthracene-based system, this time containing di-boronic acids.⁴¹ It had a high selectivity for D-Glucose over other glycans.

Binghe Wang’s group used a similar strategy to the Shinkai’s PET mechanism and developed a series of di-boronic acid sensors. These molecules had different linkers between two anthracene fluorophores (**Figure 1.7e**).⁴² Wang’s group found that the compounds with the para-benzene linker showed moderate selectivity for sialyl Lewis X in aqueous solution. Di-boronic acid sensor was found to show the strongest fluorescence enhancement upon binding with sLe^x. After binding *in situ* to sLe^x, the sensor enabled labeling of hepatocellular carcinoma (HEPG2) cells selectively via binding sLe^x over Ley expressing HEP3B cells and COS7 cells, either of which do not express sLe^x.⁴³ Non-sLe^x-expressing cells remained un-labeled.

James’s group developed electrochemical sensors, which contained electro-active ferrocene attached to mono-boronic acid (**Figure 1.7f**) or di-boronic acid (**Figure 1.7g**). Overall, the di-boronic acid sensor was more selective for D-glucose and D-galactose compared to the mono-boronic acid sensor.

As discussed earlier, the recognition of more complex saccharides can play a valuable role in the detection of many different diseases. This is due to the cis-diols that are found on glycans and glycoproteins and their ability to interact with boronic acids. The variety of glycan conjugates and glycoproteins that show up during oncogenesis, create the dire need for glycan-sensing and therefore, boronic acids have been incorporated into many biosensors for their application in cancer diagnostics.

Borono-lectins are the bio-sensing probes that were developed by Hall and Anslyn groups. They are boronic acid conjugated short peptide sequences. The selectivity of these sensors has been greatly improved over the years. One important issue of these sensors was their biocompatibility. In order to optimize these sensors, they needed to have good water solubility, low toxicity, and greater stability than the other glycan sensors. Most of the fluorescent boronic acid sensors had poor water solubility and therefore needed organic co-solvents for sugar binding studies. Wang's group after enjoying the success of anthracene based PET sensors, developed different sensors that contained quinoline and naphthalene instead. These structures improved water solubility and stability. When binding to glycans, the fluorescence of these sensors changed. The changes were specific to each glycan. However, these polycyclic aromatic compounds are known to be carcinogenic, hence could not be used for sensing *in vivo*.⁴⁴⁻⁵⁰

While all of these sensors were promising in glycan sensing, they still had limitations. In order to overcome these limitations, Hall used a library technique to find sensors that targeted the TF antigen, which is a cancer-associated glycan. He used a peptide backbone due to its biocompatible properties. The peptide backbone also had the potential to have secondary interactions with the glycans, which could enhance binding affinity and

selectivity. Hall found that peptides functionalized with boroxoles or boronic acids had more selectivity for the TF antigen compared to peptides without those groups.⁵¹

1.7 CROSS-REACTIVE SYNTHETIC LECTIN ARRAYS

Previously, the discussion involved application of natural lectins binding and recognizing specific glycan residues with high affinity. The specificity of this interaction is analogous to a lock-key model, which is similar to an antigen/antibody interaction. Since, one specific glycan does not correspond to one type of cancer, hence oncogenesis involves a gross change in glycosylation. Therefore, 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. (Figure 1.8).⁵² shows a signal output after interaction of single analyte with many receptors.

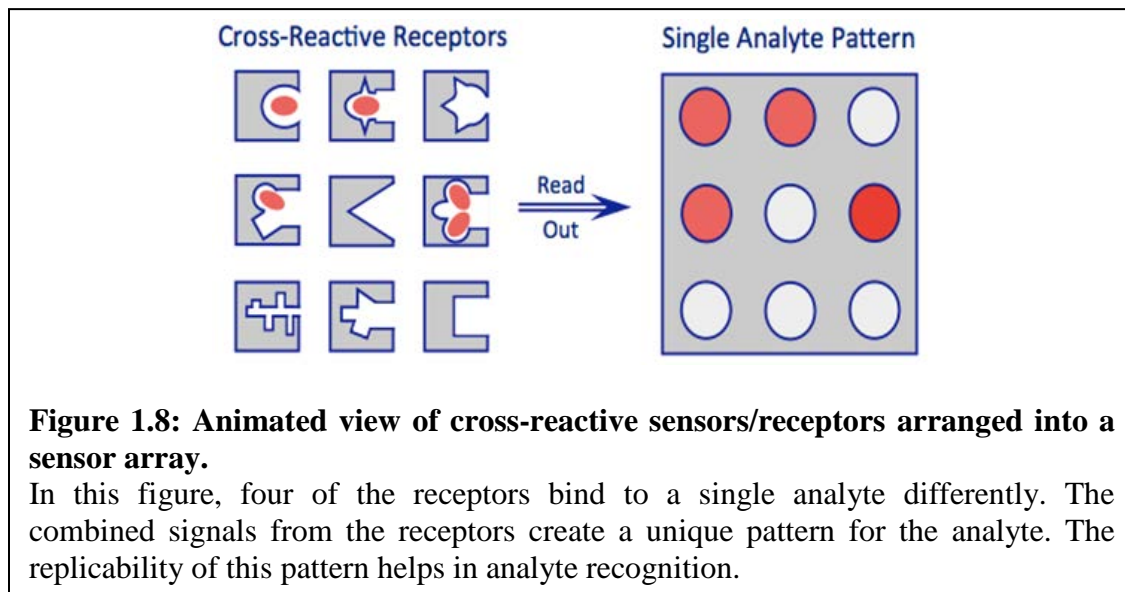


Figure 1.8: Animated view of cross-reactive sensors/receptors arranged into a sensor array.

In this figure, four of the receptors bind to a single analyte differently. The combined signals from the receptors create a unique pattern for the analyte. The replicability of this pattern helps in analyte recognition.

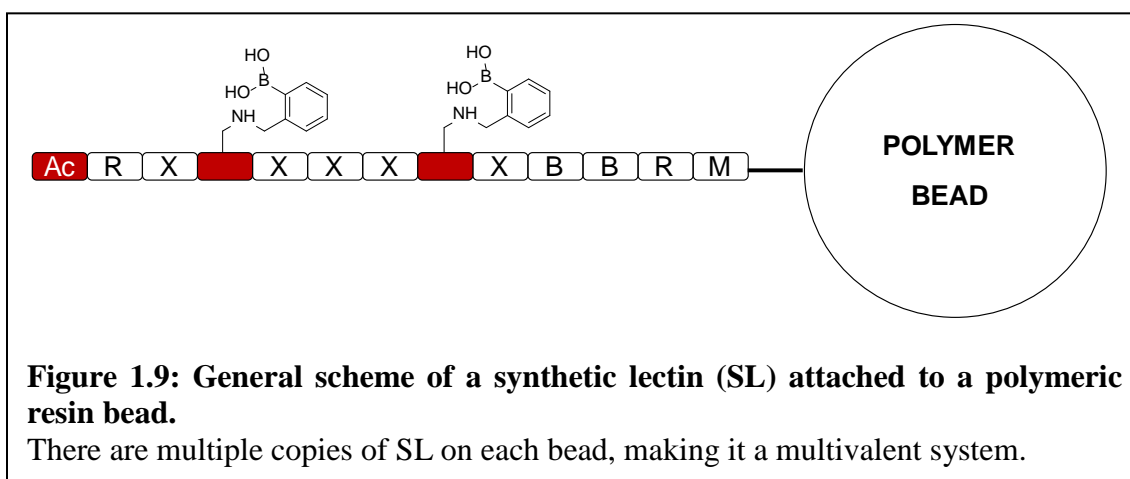
Although aberrant glycosylation is a characteristic of cancer, there are no specific glycans that have been identified for a specific cancer type. Therefore, instead of looking for a specific biomarker, we should be looking for a global pattern of glycosylation.

Hence monitoring changes in the abundance of more than one glycan makes it more useful for the diagnosis of abnormal cells. The basis of utilization of SL sensor array is the principle of cross-reactivity. As explained earlier, cross-reactive sensor arrays are composed of multiple ‘receptors’ (SLs in this case), with each receptor interacting differently with different glycans. Information from these binding interactions, as a whole, leads to the formation of a reproducible unique pattern for the “analyte” (glycans in this case). This approach enables the discrimination of cells into normal, cancerous or metastatic groups. These sensor arrays resemble, in principle, the electronic nose⁵³ or electronic tongue⁵⁴ sensors, which involve pattern recognition systems. Each sensor only has to be incrementally different to create a fingerprint that maximizes variation of the array response binding to different analytes.

1.8 OUR APPROACH: BORONIC ACID FUNCTIONALIZED SYNTHETIC LECTIN

Keeping natural lectin arrays and their binding interactions with glycans in mind, one can synthesize lectins and then generate a synthetic lectin array to investigate binding to glycans. In doing so, the cross-reactive property of the lectin array is harnessed, while avoiding an elaborate usage of too many natural lectins. The Lavigne lab synthesized Polymeric Lectins and its array. These synthetic lectins (SLs) are short peptide chains having with Phenyl Boronic acids attached on to them as a side-chain functionality. These peptide-boronic acid conjugates have several copies of them on each polymeric resinous bead (typically 300 μ m size). The advantages to using synthetic lectins are that they are easy to make, biocompatible, and have low toxicity. **(Figure 1.9)** depicts a general synthetic lectin on a resin. It depicts one synthetic lectin on a bead; however, there is more than one (bead capacity on the resin 0.3 μ M/mg). The C-terminus of the

peptide starts with a Methionine (M, for orthogonal binding), followed by Arginine (R) and two β -Alanine (B, as spacers). The resin-MRBB unit is followed by eight X amino acid residues, where X represents a random amino acid chosen (natural as well as non-natural). The non-natural amino acid which is commonly used is diaminobutanoic acid, represented as Dab. Dab contains $-\text{NH}_2$ on its side chain and each peptide chain contains two Dab units. The boronic acids (2-formyl phenyl boronic acid) are attached on to this side chain via reductive amination. As seen in **Figure 1.9** there are two phenyl boronic acid units in each peptide sequence. The N-terminus of the SL sequence is usually acylated (Ac).

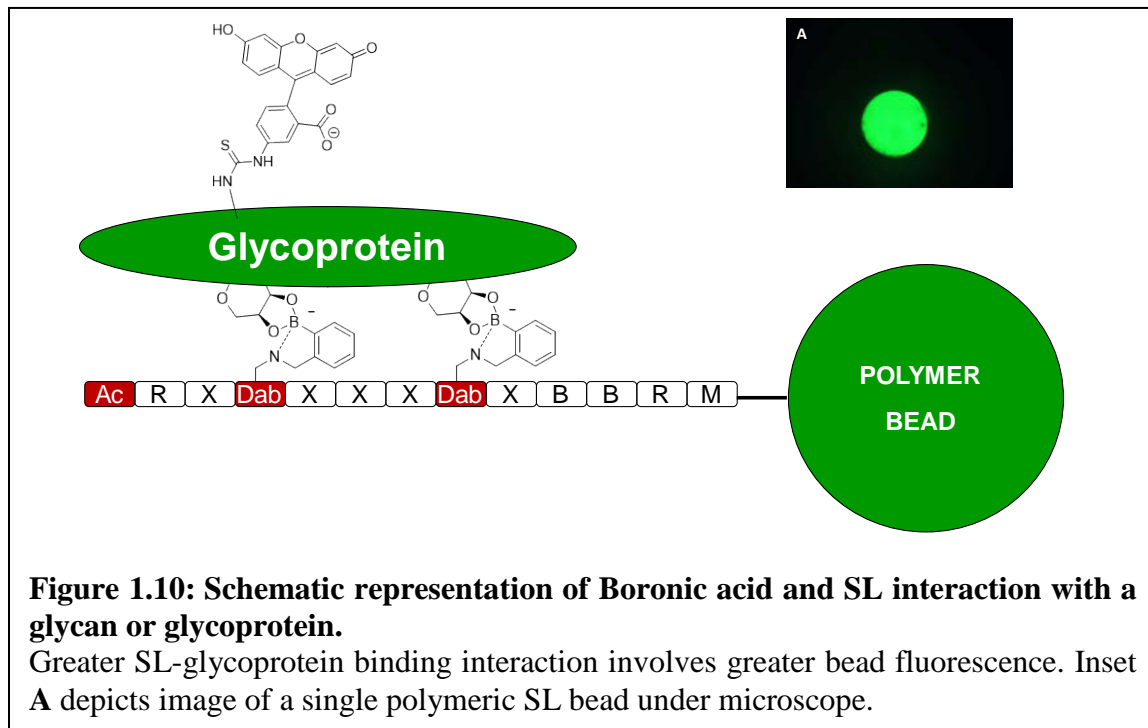


1.9 PROPOSED SYNTHETIC LECTIN AND GLYCAN BINDING

Several other research groups have targeted carbohydrates/ CAGs using boronic acids in their receptors⁵⁵⁻⁶⁰ because they form covalent yet reversible boronate esters with *cis* diols (1,2- or 1,3-) present on CAGs.⁶¹⁻⁶³ Our results indicate that boronic acid incorporation causes enhancement of SL-sugar interaction because of the covalent nature of its bond formation i.e. a boronate ester linkage.²⁸ Since purified and quantified CAGs were very expensive for initial studies, the SL-glycan binding were studied by making

use of glycoproteins from chickens, cows and pigs. The glycoproteins used were albumin from chicken ovalalbumin (OVA) and, mucins of porcine stomach (PSM) and bovine submaxillary glands (BSM). The mucins were selected because they are enriched in high mannose, hybrid and complex N-linked and O-linked glycans; all usually being present in cancer related targets.²³ Similarly, OVA also includes complex N-linked glycans which also make up the core of CAGs. Bovine Serum Albumin (BSA) is a negative glycan control protein since it remains un-glycosylated.

These glycoproteins/ proteins were fluorescently tagged and incubated with SLs. It is established that higher the SL-glycoprotein binding, higher would be the fluorescence of polymeric SL bead (**Figure 1.10**).



Synthetic lectin library design, optimization and peptide sequencing of the synthetic lectins have been discussed in previous Lavigne et. al. work²⁹, they will also be discussed later in chapter 4.

From the previously established work, several SLs were found to have higher selectivity for a particular *proof-of-concept* glycoprotein (e.g. SL5 for PSM), have been identified as SL ‘hits’. There are several SLs which are cross-reactive i.e. recognize more than one glycoprotein in a unique manner (e.g. SL1). The earliest of selective hits and cross-reactive SLs have been identified and their sequence is depicted in (**Table 1.1**).

Table 1.1: Synthetic Lectins (SLs) with their peptide sequences that show binding with the proof of concept glycoproteins.			
(Dab)* means diaminobutanoic acid, the linker amino acid with phenyl boronic acid on its side chain. Each sequence has 3 arginines hence an overall charge of +3. Ac stands for acylation of N-terminus of the peptide sequence.			
SL Hit	Sequence	Overall SL charge	Type
SL1	Ac-RG(Dab)*VTF(Dab)*R-BBRM-resin	+3	cross-reactive
SL2	Ac-RT(Dab)*RFL(Dab)*V-BBRM-resin	+3	moderately selective
SL3	Ac-RS(Dab)*VTT(Dab)*R-BBRM-resin	+3	moderately selective
SL4	Ac-RR(Dab)*TQT(Dab)*Q-BBRM-resin	+3	moderately selective
SL5	Ac-RA(Dab)*TRV(Dab)*V-BBRM-resin	+3	highly selective

From the analysis, five SLs (designated SL1-SL5) were validated as hits, out of which SL1 was highly cross-reactive towards all four glycoproteins, whereas SL2 and SL5 showed selectivity towards a specific glycoprotein. A percentage change in binding, (**Figure 1.11**) was obtained by dividing the fluorescence of a similarly sized set of SL library from the difference between the fluorescence value of re-synthesized SL and fluorescence value of the library. (**Eq 1**)

$$\% \Delta \text{ Fluorescence} = \frac{\text{SL fluorescence intensity} - \text{library fluorescence intensity}}{\text{library fluorescence intensity}} \quad \text{Eq 1}$$

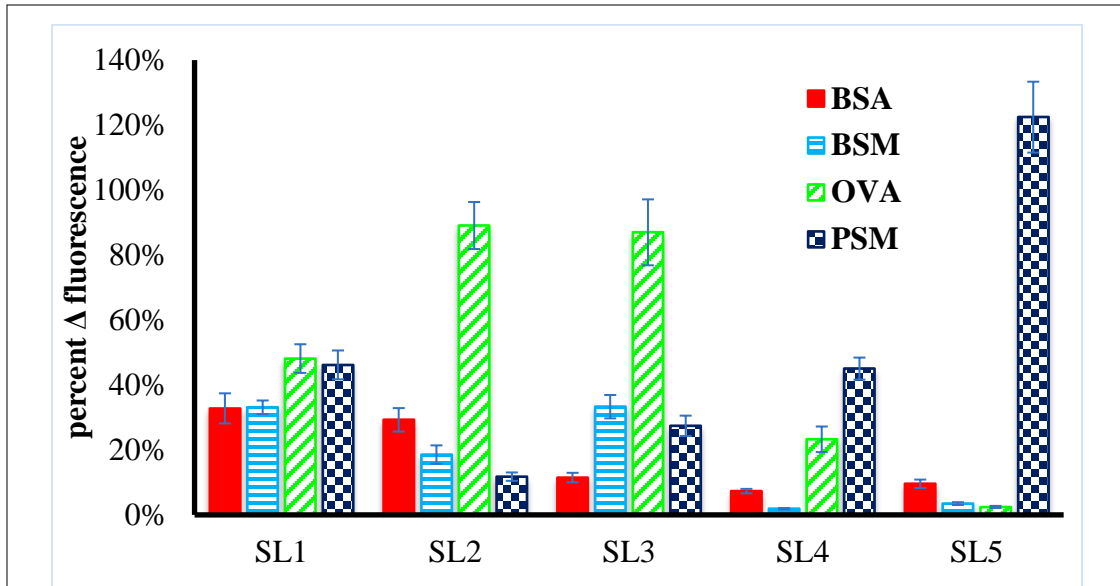


Figure 1.11: Detected response from SL1-5 binding with four purified proteins as measured by fluorescence.

Certain SLs reported high selectivity for a specific protein (e.g. SL5 for PSM). SL1 remained cross-reactive for all four proteins. A percentage change in fluorescence obtained by using the average fluorescence of a set of SL library beads as a control. (The library control used because it contained all selective and cross-reactive elements that could have interfered with the assessment of binding). Signals represent average percentage change in luminosity with error bars displaying ± 1 standard deviation.

The library was used as a control because it contained all selective and cross-reactive elements that could have interfered with the assessment of our binding.²³

1.10 SYNTHETIC LECTINS AND THEIR BINDING WITH CELL LINE PROTEINS

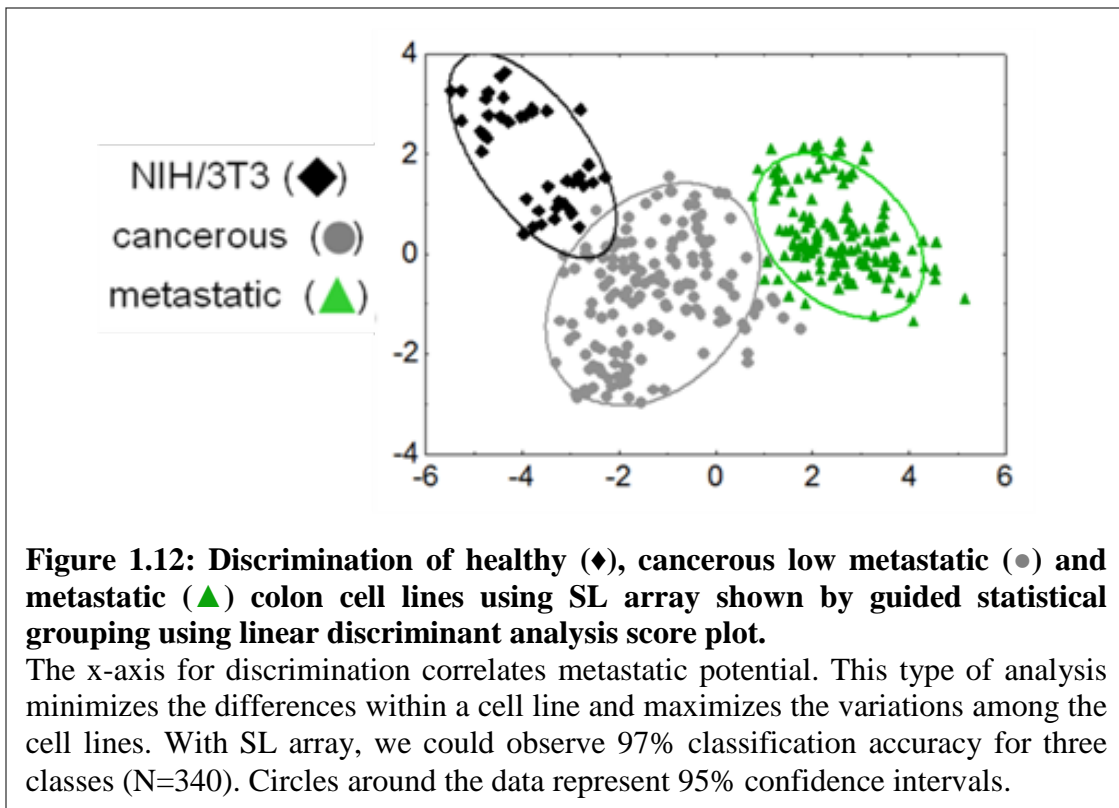
A four-component lectin array was assembled, which included both selective and cross-reactive SLs (SL1, SL3, SL4 and SL5). This array could discriminate between five common CAGs (TF antigen, Le^a, Le^x, sLe^a and sLe^x) with 95% accuracy.²³ (figure not included and this work will be discussed in detail later in chapter 4). Being that the SL array response was able to discriminate between structurally similar polysaccharides (e.g. sLe^a and sLe^x), 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. It will mean whether the cell line is considered healthy/normal, cancerous non-metastatic/lowly-metastatic, and highly metastatic. (Table 1.2) lists the cell lines used in the experiment described.

Table 1.2 Colon cell lines used for SL binding analysis; listed is the species, cell type and origin of the cell lines.			
Cell Line name	Species type	Type	Cancerous tissue area
3T3/NIH	Mouse	Healthy	N/A
CT-26	Mouse	Low metastatic	colon
CT-26-FL1	Mouse	High metastatic	colon
CT-26-FL3	Mouse	High metastatic	colon
HT-29	Human	Low metastatic	colon
HCT-116	Human	Low metastatic	colon
LoVo	Human	High metastatic	colon

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 from colon cancer, it had metastasized to its secondary site, lymph nodes. CT26-F1 and CT26-FL3 are isogenic mouse cell lines meaning that they are derived from the same cell line, CT26. For clarity, CT26 is non-metastatic/lowly metastatic with being only 0-5% metastatic. CT26-F1 is metastatic with a rate of 50% metastatic, and CT26-FL3 is highly metastatic with a rate of being 95% metastatic. The cell lines were grown according to ATCC guidelines and the membrane proteins/glycoproteins were extracted using the Qiagen® membrane extraction kit. Later, these proteins were fluorescently tagged using

FITC and incubated with SL array containing only four lectins. The beads were analyzed using fluorescence microscopy and imaged. The bright values were obtained using Adobe Photoshop® as described earlier in this chapter. 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. LDA is a statistical program that tries to minimize the separation within the same group, while maximizing the differences between other groups. LDA is a guided statistical method, which means that the program knows which



group each data point should be in. (Figure 1.11) is the LDA plot generated from the data obtained by incubating four SLs with six cell lines.

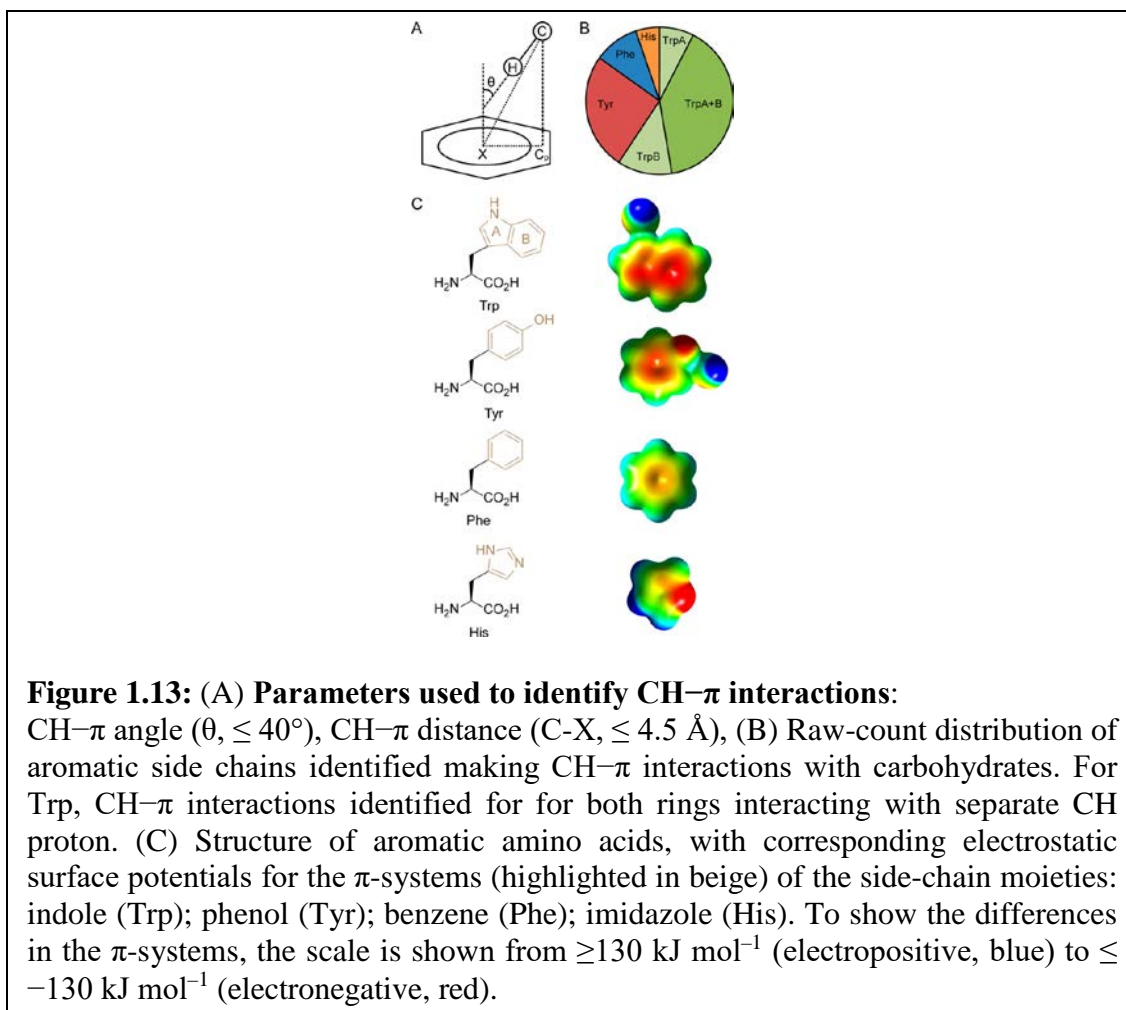
The fingerprint pattern was obtained and LDA was used to better distinguish the similarities and differences in the array response. Interestingly, the four SLs selected for this study were selected against ‘proof-of-concept’ glycoproteins. None of those

glycoproteins were present in the cancer cell lines studied, thus illustrating the ability of array to classify glycoproteins for which it was not originally designed. This four-component SL array was assembled, and included both selective and cross-reactive SLs (SL1, SL3, SL4 and SL5). They were able to discriminate classify seven different colon cancer cell lines in three classes (healthy/cancerous/metastatic) with 97% accuracy²³

1.11 NON-COVALENT PROTEIN INTERACTIONS IN AQUEOUS ENVIRONMENT

The strongest of the non-covalent interactions include ionic and hydrogen bonding, but while investigating proteins in aqueous solutions, these interactions become irrelevant.⁶⁴ Weaker or “non-conventional” hydrogen bonding interactions, like CH- π interactions become stronger and more specific. CH- π interactions (first postulated by Tamres in 1952) involve a CH group, which acts as the hydrogen bond donor (soft acid) and the electrons of a π -system act as hydrogen acceptor (soft base).⁶⁵ Nishio et. al. reported that roughly 29% of all known organic crystal structures in the Cambridge database containing a phenyl ring and an alkane region show an evidence of a CH- π interaction. There exists a remarkable difference between the nature of a traditional Pauling postulated strong hydrogen bond interactions and CH- π interactions.⁶⁶

The conventional hydrogen bonds mainly contain electrostatic forces, whereas “typical” CH- π interactions have a major contribution from London dispersion forces, electrostatic forces and directional polarization.⁶⁷ However, the magnitude of CH- π interactions, is considerably smaller than that of a conventional H-bond. In gas phase, an isolated and conventional neutral H-bond in chloroform solution has an enthalpy gain that ranges from 1 to 7 kcal·mol⁻¹. Conversely, the enthalpy of a single CH- π interaction is less than one kcal·mol⁻¹.⁶⁵



Despite of the low enthalpic contribution, the total energy provided by CH- π interactions is significantly greater and thus is due to the participation of multiple CH and/or π -groups acting together in-sync (multivalency). The occurrence of multiple CH- π interactions in a supramolecular assembly renders them quite important in terms of total energy and preferred structure, but they do not have a significant role in controlling relative orientations of molecules.

Consequently, CH- π interactions exist in a variety of phenomena including molecular recognition, crystal engineering, supramolecular adhesives and the structure of many biomolecules.⁶⁸⁻⁷⁰ In (Figure 1.13) Hudson et. al. describe the parameters used to define the strength of CH- π interactions between C-H donor of carbohydrates and pi-system of

aromatic amino acids in proteins. While studying CH- π binding, they collected all data (Protein Data Bank) for any interaction between atoms of amino acids with and any atom of a carbohydrate moiety within 4.0 Å distance. Using this dataset and applying CH- π measurement parameters (Figure 1.13A), they found out that aromatic residues contacted carbohydrates in the order tryptophan (Trp) \gg tyrosine (Tyr), Phenylalanine (Phe) $>$ histidine (His) (**Figure 1.13B**).⁷¹

Brandl et. al. examined a set of 1154 protein structures from the Protein Data Bank for any CH-donor and pi-acceptor type interactions. They found out that the most prominent interactions were between aliphatic CH donors and aromatic pi-acceptors and interactions between aromatic CH donors and aromatic pi-acceptors. About half of all Phenylalanine (Phe, F) and Tyrosine (Tyr, Y) π -rings, were involved as acceptors in CH- π -interactions. Aromatic CH groups as well as aliphatic side-chains of the long, extended amino acid residues Lys, Arg and Met, and the Pro ring were preferred as donors.⁷²

Besides CH- π interaction, there also exists cation- π type interactions that contribute to protein folding as well. In gas phase the cation- π interaction energies between aromatic amino acids (Phe, Tyr, and Trp) and the interaction energies of protonated amino acids (Arg (+) and Lys (+)) are in the range from -9 to -18 kcal·mol⁻¹. In aqueous solutions, the cation- π energies of H₃O (+) and protonated amino acids are less affected by solvation effects.⁷³

The next chapter (**chapter 2**) describes various chemical modifications done to the SLs (peptide sequences and boronic acid moieties) in order to establish *structure and activity relationships* against purified glycoprotein analytes. The aim of these experiments is to gain an insight on chemical basis of SL-glycoprotein interactions.

Using what was learnt in chapter 2, several hypotheses were drawn on the design of new SLs.

Chapter 3 aims at incorporating new SLs into the pre-existing SL array to investigate colon and prostate cancer *in vitro*. The chemical makeup of the SLs which statistically contributed the most to the cancer discrimination, is evaluated.

Chapter 4 details on **quantifying sialic acid** content *in vitro* and investigating any correlations with the metastatic potential of colon and prostate cell lines. Based on the relative importance that charged amino acids and phenyl boronic acids in SLs have at evaluating metastatic potential; several tissue-specific SLs were found and their **peptide sequence homology** was studied.

1.12 REFERENCES

1. Ferlay, J.; Soerjomataram, I.; Ervik, M.; Dikshit, R.; Eser, S.; Mathers, C. Cancer Incidence and Mortality Worldwide: IARC. *International Agency for Research on Cancer GLOBOCAN 2012*.
2. Society, A. C., Cancer Facts and Figures **2017**, pp 1-76.
3. Ries, L.; Melbert, D.; Krapcho, M. SEER Cancer Statistics Review, 1975–2004, *National Cancer Institute, 2007*.
4. Zuckerman, D.R.; Rockey, D.C. A prospective multicenter evaluation of new fecal occult blood tests in patients undergoing colonoscopy. *Am. J. Gastroenterol.* **2000**, 95, 1331-1338.
5. Singh, H. N. Z.; Demers, A.A.; Kliewer, E.V.; Mahmud, S.M.; Bernstein, C.N. The reduction in colorectal cancer mortality after colonoscopy varies by site of the cancer. *Gastroenterology* **2010**, 139, 1128-1137.
6. Levi, Z. B., S.; Vilkin, A.; Bar-Chana, M.; Lifshitz, I.; Chared, M.; Maoz, E.; Niv, Y. A higher detection rate for colorectal cancer and advanced adenomatous polyp for screening with immunochemical fecal occult blood test than guaiac fecal occult blood test, despite lower compliance rate. A prospective, controlled, feasibility study. *International J. of Cancer* **2011**, 128 (10), 2415-2424.
7. Gupta, S. S., D.A.; Doubenia, C.A.; Anderson, D.S.; Lukejohn, D.; Deshpande, A.R.; Elmunzer, B.J.; Laiyemo, A.O.; Mendez, J.; Somsouk, M.; Allison, J.; Bhuket, T.I.; Geng, Z.; Green, B.B.; Itzkowitz, S.H.; Martinez, M.E., Challenges and possible solutions to colorectal cancer screening for the underserved. *J. of the National Cancer Institute* **2014**, 106 (4).
8. Gupta, S. H., E.A.; Rockey, D.C.; Hammons, M.; Koch, M.; Carter, E.; Valdez, L.; Tony, L.; Ahn, C.; Kashner, M.; Argenbright, K.; Tiro, J.; Geng, Z.; Pruitt, S.; Skinner, C., Comparative effectiveness of fecal immunochemical test outreach, colonoscopy outreach, and usual care for boosting colorectal cancer screening among the underserved. *JAMA Internal Medicine* **2013**, 173 (18), 1725-1732.
9. Quintero, E. C., A.; Bujanda, L.; Cubiella, J. et al, Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. *New England Journal of Medicine* **2012**, 366, 697-706.
10. Han–Mo Chiu, Y. C. L., Chia–Hung Tu, Chien–Chuan Chen, Ping–Huei Tseng, Jin–Tung Liang, Chia–Tung Shun, Jaw–Town Lin, Ming–Shiang Wu, Association between early stage colon neoplasms and false-negative results from the fecal immunochemical test. *Clinical Gastroenterology and Hepatology* **2013**, 11 (7), 832-838.
11. Postma, R.; Schroder, F. H.; van Leenders, G. J.; Hoedemaeker, R. F.; Vis A. N.; Roobol, M. J. Cancer detection and cancer characteristics in the European Randomized Study of Screening for Prostate Cancer (ERSPC) – Section Rotterdam. A comparison of two rounds of screening. *Eur Urol.* **2007**, 52, 89–97.

12. Schröder, F. H.; van der Crujisen-Koeter, I.; de Koning H. J.; Vis A. N.; Hoedemaeker, R. F.; Kranse, R.; Prostate cancer detection at low prostate specific antigen. *J Urol.* **2000**, *163*, 806–812.
13. Schroder, F. H.; van der Maas, P.; Beemsterboer, P.; Kruger, A. B.; Hoedemaeker, R.; Rietbergen, J. Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst.* **1998**, 1817–1823.
14. Fakhri, M.G.; Aruna, P. CEA monitoring in colorectal cancer: what you should know. *Oncology* **2006**, *20*, 579-587.
15. Passerini, R.; Cassatella, M. D.; Boveri, S.; Salvatici, M.; Radice, D.; Zorzino, L.; Galli, C.; Sandri, M. T. The Pitfalls of CA19-9: Routine Testing and Comparison of Two Automated Immunoassays in a Reference Oncology Center. *Am J Clin Pathol* **2012**, *138* (2): 281-287.
16. Bast, R. C., Jr.; Ravdin, P.; Hayes, D. F.; Bates, S.; Fritsche, H., Jr.; Jessup, J. M.; Kemeny, N.; Locker, G. Y.; Mennel, R. G.; Somerfield, M. R. 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol.* **2001**, *19* (6), 1865-78.
17. Candas, B.; Cusan, L.; Gomez, J. L.; Diamond, P.; Suburu, R. E.; Lévesque, J. Evaluation of prostatic specific antigen and digital rectal examination as screening tests for prostate cancer. *Prostate* **2000**, *45*, 19–35.
18. Brawer, M. K.; Lange, P. H. Prostate-specific antigen in management of prostatic carcinoma. *Urology* **1989**, *33* (5), 11–16.
19. Barry, M. J., Prostate-Specific–Antigen Testing for Early Diagnosis of Prostate Cancer. *New England Journal of Medicine* **2001**, *344* (18), 1373-1377.
20. Hollingsworth, M.A; Swanson, B.J. Mucins in cancer: Protection and control of the cell surface. *Nature Rev. Cancer* **2004**, *4*, 45-60.
21. Gu, J.; Taniguchi, N. Potential of N-glycan in cell adhesion and migration as either a positive or negative regulator, *Cell Adh Migr.* **2008**, *2*, (4), 243–245.
22. Dell, A.; Galadari, A.; Sastre, F.; Hitchen, P., Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *International Journal of Microbiology* **2010**, *2010*, 1-14.
23. Dube, D.H.; Bertozzi, C.R. Glycans in cancer and inflammation, potential for therapeutics and diagnostics. *Nature Reviews* **2005**, *4*, 477-488.
24. Munkley, J. E.; Hallmarks of glycosylation in cancer. *Oncotarget* **2016**, *7* (23), 35478-35483.
25. Taniguchi, N., Korekane, H. Branched N-glycans and their implications for cell adhesion, signaling and clinical applications for cancer biomarkers and in therapeutics. *BMB Rep.* **2011**, *44* (12), 772-81.

26. Meezan, E.; Wu, H.C.; Black, P.H.; Robbins, P.W. Comparative studies on the carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. *J. Am. Chem. Soc.* **1969**, *8*, 2518-2524.
27. Turner, G.A. N-glycosylation of serum proteins in disease and its investigation using lectins, *Clinica Chimica Acta.* **1992**, *3*, 149–171.
28. Bicker, K.L.; Sun, J.; Lavigne, J.J. Boronic acid functionalized peptidyl synthetic lectins: combinatorial library design, Peptide sequencing, and selective glycoprotein recognition. *ACS Comb. Sci.* **2011**, *13*, 232-243.
29. Bicker, K.L.; Sun, J.; Lavigne, J.J. Synthetic lectin arrays for the detection and discrimination of cancer associated glycans and cell lines. *Chem. Sci.* **2012**, *3*, 1147-1156.
30. Magnani, J.L. *Arch. Biochem. Biophys.* **2004**, *426*, 122-131.
31. Mody, R.; Joshi, S.; Chaney, W. Use of lectins as diagnostic and therapeutic tools for cancer. *J. Pharm. & Toxic. Methods* **1995**, *33*.
32. Shui, W.; Li, Z. Glycoproteomic analysis of tissues from patients with colon cancer using lectin microarrays and nanoLC-MS/MS. *Mol. BioSyst.* **2013**, *9*, 1877.
33. Berg, J. M.; Stryer, L. Biochemistry. In Biochemistry, 5th ed.; W.H. Freeman: New York **2002**, (5).
34. Patwa, T. H. Z., J.; Andersone, M.A.; Simeone, D.M.; Lubman, D.M., Screening of Glycosylation Patterns in Serum Using Natural Glycoprotein Microarrays and Multi-Lectin Fluorescence Detection. *Anal. Chem.* **2006**, *78*, 6411-6421.
35. Zhao, J.; H. Z., J.; Andersone, M.A.; Simeone, D.M.; Lubman, D.M., Glycoprotein Microarrays with Multi-Lectin Detection: Unique Lectin Binding Patterns as a Tool for Classifying Normal, Chronic Pancreatitis and Pancreatic Cancer Sera *Anal. Chem.* **2006**, *78*, 6411-6421.
36. Oshovsky, G. V.; Reinhoudt, D. N.; Verboom, W., Supramolecular Chemistry in Water. *Angew, Chem. Int. Ed.* **2007**, *46* (14), 2366-2393.
37. Collins, B. E.; Sorey, S.; Hargrove, A. E.; Shabbir, S. H.; Lynch, V. M.; Anslyn, E. V., Probing Intramolecular B–N Interactions in Ortho-Aminomethyl Arylboronic Acids. *J. Org. Chem.* **2009**, *74* (11), 4055-4060.
38. Yoon, J.; Czarnik, A. W., Fluorescent chemosensors of carbohydrates. A means of chemically communicating the binding of polyols in water based on chelation-enhanced quenching. *J. Am. Chem. Soc.* **1992**, *114* (14), 5874-5875.
39. DiCesare, N.; Lakowicz, J. R., Spectral Properties of Fluorophores Combining the Boronic Acid Group with Electron Donor or Withdrawing Groups. Implication in the Development of Fluorescence Probes for Saccharides. *The Journal of Physical Chemistry A* **2001**, *105* (28), 6834-6840.
40. James, T. D.; Sandanayake, K. R. A. S.; Iguchi, R.; Shinkai, S., Novel Saccharide-Photoinduced Electron Transfer Sensors Based on the Interaction of Boronic Acid and Amine. *Journal of the American Chemical Society* **1995**, *117* (35), 8982-8987.

41. James, T. D.; Samankumara Sandanayake, K. R. A.; Shinkai, S., Chiral discrimination of monosaccharides using a fluorescent molecular sensor. *Nature* **1995**, *374* (6520), 345-347.
42. Yang, W.; Gao, S.; Gao, X.; Karnati, V. V. R.; Ni, W.; Wang, B.; Hooks, W. B.; Carson, J.; Weston, B., Diboronic acids as fluorescent probes for cells expressing sialyl lewis X. *Bioorganic & Medicinal Chemistry Letters* **2002**, *12* (16), 2175-2177.
43. Yang, W.; Fan, H.; Gao, X.; Gao, S.; Karnati, V. V. R.; Ni, W.; Hooks, W. B.; Carson, J.; Weston, B.; Wang, B., The First Fluorescent Diboronic Acid Sensor Specific for Hepatocellular Carcinoma Cells Expressing Sialyl Lewis X. *Chemistry & Biology* **2002**, *11* (4), 439-448.
44. Arimori, S.; Ushiroda, S.; Peter, L. M.; Jenkins, A. T. A.; James, T. D., A modular electrochemical sensor for saccharides. *Chemical Communications* **2002**, (20), 2368-2369.
45. Zhang, Y.; Gao, X.; Hardcastle, K.; Wang, B., Water-Soluble Fluorescent Boronic Acid Compounds for Saccharide Sensing: Substituent Effects on Their Fluorescence Properties. *Chemistry – A European Journal* **2006**, *12* (5), 1377-1384.
46. Yang, W.; Yan, J.; Springsteen, G.; Deeter, S.; Wang, B., A novel type of fluorescent boronic acid that shows large fluorescence intensity changes upon binding with a carbohydrate in aqueous solution at physiological pH. *Bioorganic & Medicinal Chemistry Letters* **2003**, *13* (6), 1019-1022.
47. Stones, D.; Manku, S.; Lu, X.; Hall, D. G., Modular Solid-Phase Synthetic Approach To Optimize Structural and Electronic Properties of Oligoboronic Acid Receptors and Sensors for the Aqueous Recognition of Oligosaccharides. *Chemistry – A European Journal* **2004**, *10* (1), 92-100.
48. Gao, X.; Zhang, Y.; Wang, B., Naphthalene-based water-soluble fluorescent boronic acid isomers suitable for ratiometric and off-on sensing of saccharides at physiological pH. *New Journal of Chemistry* **2005**, *29* (4), 579-586.
49. Gao, X.; Zhang, Y.; Wang, B., New Boronic Acid Fluorescent Reporter Compounds. 2. A Naphthalene-Based On-Off Sensor Functional at Physiological pH. *Organic Letters* **2003**, *5* (24), 4615-4618.
50. Gao, X.; Zhang, Y.; Wang, B., A highly fluorescent water-soluble boronic acid reporter for saccharide sensing that shows ratiometric UV changes and significant fluorescence changes. *Tetrahedron* **2005**, *61* (38), 9111-9117.
51. Pal, A.; Bérubé, M.; Hall, D. G., Design, Synthesis, and Screening of a Library of Peptidyl Bis(Boroxoles) as Oligosaccharide Receptors in Water: Identification of a Receptor for the Tumor Marker TF-Antigen Disaccharide. *Angewandte Chemie International Edition* **2010**, *49* (8), 1492-1495.
52. Lavigne, J. J. A., E. V. Sensing a paradigm shift in the field of molecular recognition: from selective to differential receptors. *Angew. Chem Int. Ed* **2001**, *40*, 3118-3130.
53. Brattoli, M.; Gennaro, G.D. Odour detection methods: olfactometry and chemical sensors, *Sensors* **2011**, *115*, 5290-5322.

54. Lavigne, J. J.; Anslyn, E.V. Solution-based analysis of multiple analytes by a sensor array: toward the development of an “electronic tongue”. *J. Am. Chem. Soc.* **1998**, *120*, 6429-6430.
55. Yang, W.; Gao, S.; Weston, B. Diboronic acids as fluorescent probes for cells expressing sialyl lewis X. *Bioorg. Med. Chem. Lett.*, **2002**, *12*, 2175–2177.
56. Yang, W.; Carson, J.; Weston, B.; Wang, B. The first fluorescent diboronic acid sensor specific for hepatocellular carcinoma cells expressing sialyl Lewis X. *Chem. Biol.* **2004**, *11*, 439–448.
57. Springsteen, G.; Wang, B. Alizarin red as a general fluorescent reporter for studying the binding of boronic acids and carbohydrates. *Chem. Commun.* **2001**, 1608–1609.
58. Anslyn, E.V.; Lavigne, J.J. Boronic acid based peptidic receptors for pattern-based saccharide sensing in neutral aqueous media, an application in real-life samples. *J. Am. Chem. Soc.* **2007**, *129*, 13575-13583.
59. Pal, A.; Hall, D.G. Design, Synthesis and screening of a library of peptidyl bis(boroxoles) as oligosaccharide receptors in water: Identification of a receptor for the tumor marker TF-Antigen disaccharide. *Angew. Chem. Int. Ed.* **2010**, *49*, 1492-1495.
60. Shinkai, M.D.; Boronic acids in saccharide recognition. *Royal Society of Chemistry Publishing*; **2006**, 14-16.
61. Wang, W.; Gao, X.; Wang, B. Boronic acid-based sensors. *Current Organic Chemistry* **2002**, *6*, 1285-1317.
62. Moertel, C. CEA monitoring among patients in multi-institutional adjuvant G.I therapy protocols. *Ann. Surg.* **1982**, *196* (2), 162-169.
63. Wang, B. Water-soluble fluorescent boronic acid compounds for saccharide sensing. *Chem. Eur. J.* **2006**, *12*, 1377–1384.
64. Dougherty, D.A. Cation- π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science*, **1996**, *271*, 163-168.
65. Nishio, M.; Hirota, M.; Umezawa, Y. The CH/ π Interaction. Evidence, Nature, and Consequences; Wiley-VCH: New York, NY, USA, **1998**.
66. Aragay, G.; Hernández, D.; Verdejo, B.; Escudero-Adán, E.C.; Martínez, M.; Ballester, P. Quantification of CH- π Interactions Using Calix(4)pyrrole Receptors as Model Systems. *Molecules*. **2015**, *20*, 16672-16686.
67. Fink, K.; Boratyński, J. Noncovalent cation- π interactions--their role in nature. *Postepy. Hig. Med. Dosw. (Online)*. **2014**, *68*, 1276-86.
68. Zondlo, N.J. Aromatic-proline interactions: Electronically tunable ch/ π interactions. *Acc. Chem. Res.* **2013**, *46*, 1039–1049.
69. Nishio, M.; Umezawa, Y.; Fantini, J.; Weiss, M.S.; Chakrabarti, P. CH/ π hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12648–12683.

70. Ballester, P.; Biros, S.M. CH- π and π - π interactions as contributors to the guest binding in reversible inclusion and encapsulation complexes. In *The Importance of Pi-Interactions in Crystal Engineering*; John Wiley & Sons, Ltd: Chichester, West Sussex, UK, **2012**, 79–107.
71. Hudson, K. L.; Bartlet, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. Carbohydrate-Aromatic Interactions in Proteins. *J Am Chem Soc.* **2015**, *137*(48), 15152-60.
72. Brandl, M.; Weiss, M.S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. C-H. π -interactions in proteins. *J Mol Biol.* **2001**, *307*, 357-77.
73. Wang, Q. Y.; Lu, J.; Liao, S. M.; Du, Q. S.; Huang, R.B. Unconventional interaction forces in protein and protein-ligand systems and their impacts to drug design. *Curr Top Med Chem.* **2013**, *13* (10), 1141-51.

CHAPTER 2:

INVESTIGATION OF STRUCTURE-ACTIVITY RELATIONSHIP BETWEEN SYNTHETIC LECTINS AND GLYCO-PROTEINS

2.0 OVERVIEW

In order to investigate virulence of cancers, as discussed in the previous chapter, an array of SLs was made (section 1.10). Some of the SLs in the array were ‘cross-reactive’ for different proteins (i.e. they bound to different analytes), and some SLs were ‘selective’ (towards a specific analyte). When the output of all the SLs was taken together, including the ‘cross-reactive SLs, they help us in generation of unique ‘fingerprint’ for a particular cancer.¹ The applicability of an SL array lies in the generation of unique fingerprints for healthy and cancerous samples.

Although, it is unclear that which protein/glycoprotein structure is causing the intended discrimination in patterns, the focus of this chapter is to investigate the factors/sites that are important, on the synthetic lectin (SL) which enable it to bind with proteins to generate a representative “signature” pattern. To determine the chemical rationale behind SL-protein binding, the SL undergoes structural alteration. After altering the SL, we observe the impact of the proposed modifications on the binding of that SL to the proteins. SL5 is selected for the alteration/mutation studies, because of its ability to bind selectively to one protein (porcine stomach mucin, PSM) over the others (i.e. Bovine

submaxillary mucin (BSM), Ovalalbumin (OVA) and Bovine serum albumin (BSA)).² SL5 underwent the following structural modifications as listed below, including:

1. Variation of intra-diamino butanoic acid distance (ref: section 2.1)
2. Altering the boronic acid attachments (ref: sections 2.2 and 2.3)
3. Altering the peptide sequence: by replacing amino acid with Alanine (A) on to
 - I. Other amino acids (both polar (Dab), Threonine (T) and non-polar Valine (V)) (ref: section 2.4)
 - II. Positively charged amino acids (e.g. Arginine (R)) (ref: section 2.5)

2.1 VARIATION OF INTRA-DIAMINO BUTANOIC ACID DISTANCE

The peptide sequence of SL5 contains two diamino butanoic acids (Dab) which have $-NH_2$ on their side chains and they act as the binding site for two 2-formyl phenyl boronic acids (2-PBA) respectively. The $-NH_2$ of the Dab and the carbonyl carbon of 2-PBA bind to each other via reductive amination. Dab is a non-natural amino acid with two less methylene groups than Lysine on its side chain.³ Dab is preferred over Lysine because of its shorter side chain providing a pre-organization to bind SL to diols.⁴ Together, Dab and 2-PBA abbreviate as (Dab)*. The two (Dab)* in SL5 are three amino acids apart. We want to investigate the proposed covalent bonding between 2-PBA on SLs and diols on the proteins. In order to do so, we study the impact on protein binding by varying distance between two 2-PBA units, by changing the intra-(Dab)* distance. Doing so, we aim to understand whether spreading the two boronic acids far apart improves the SL/diol interaction or confining the two boronic acids near each other serves to limit the covalent interaction. Several 'SL5 positional mutants' (**Table 2.1**)

were prepared In these mutants, the intra-(Dab)* distance was reduced up to one and stretched up to a maximum of six AA distance. These seven mutants along with the original SL5 (as a reference) were tested by incubating each one of them with four different purified proteins, and imaging under fluorescent microscope (**Figure 2.1**)

Table 2.1: Representing “SL5 positional mutants” to investigate impact of variation of intra-(Dab)* distance on SL5 binding with proteins.					
Mutant name	Dab* space	+	Peptide sequence after BBRM-resin	Fold selectivity PSM/BSA	Fold selectivity PSM/BSM
(Dab)*4,6	1	3	Ac-RAT(Dab)*R(Dab)*VV	2.6	1.8
(Dab)*3,6	2	3	Ac-RA(Dab)*TR(Dab)*VV	2.5	1.9
(Dab)*4,7	2	3	Ac-RAT(Dab)*RV(Dab)*V	2.3	1.6
(Dab)*3,7(SL5)	3	3	Ac-RA(Dab)*TRV(Dab)*V	2.2	1.6
(Dab)*3,8	4	3	Ac-RA(Dab)*TRVV(Dab)*	2.9	1.8
(Dab)*2,7	4	3	Ac-R(Dab)*ATRV(Dab)*V	2.6	1.8
(Dab)*2,8	5	3	Ac-R(Dab)*ATRVV(Dab)*	2.8	1.8
(Dab)*1,8	6	3	Ac-(Dab)*RATRVV(Dab)*	2.3	1.9

Grey indicates the positions of diamino butanoic acid functionalized with boronic acid (Dab)* present on the peptide sequence. The numbers in the mutant names refer to positions of (Dab)* moieties while counting from N-terminus of the peptide. Ac stands for Acetylation of N-terminus of peptide sequence. B stands for β-Alanine AA. Fold selectivity describes propensity of mutant to bind to one protein over the other. BSA is a no-glycosylated protein serving as a negative control. BSM has the closest glycan homology to PSM.

The experiment was conducted multiple times and the following discussion is the result of them together. As mentioned in previous chapter, SL5 binds well to PSM (SL5 was selected out of a SL library pool against PSM protein) and it is reflected in **Figure 2.1A** under the response of PSM with eight SLs. Also, looking at the fold selectivity of PSM/BSM from **Table 2.1**, it can be seen that PSM response with the eight SLs is nearly

twice as much as their response with BSM. It indicates that changing the (Dab)* distance made no significant change to the binding of SL5 with PSM. This can be explained by the fact that all of the mutants have three Arginine (R) residues, giving the SL a +3 charge as shown in **Table 2.1**. Further, looking at the overall binding paradigm of SLs

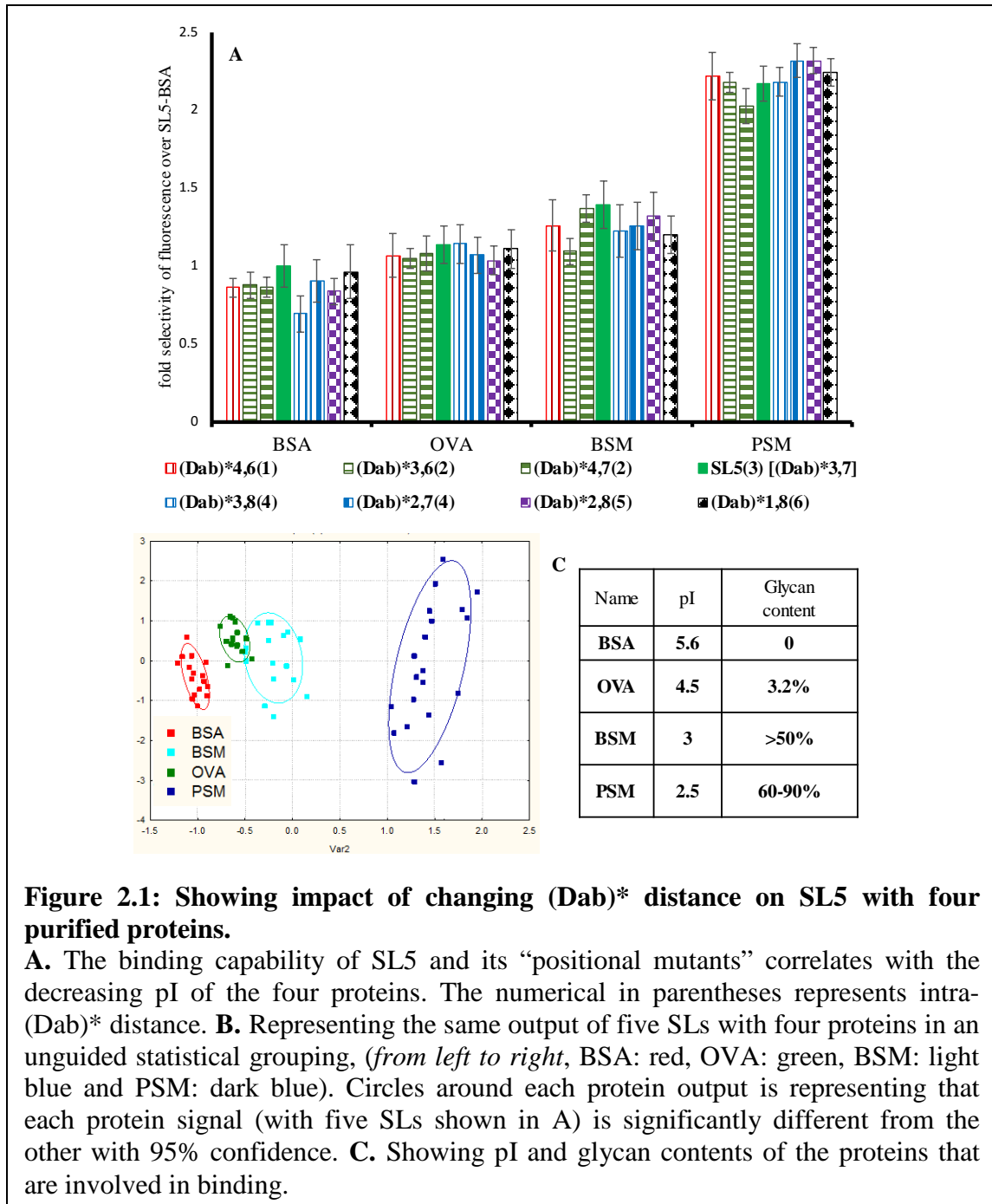


Figure 2.1: Showing impact of changing (Dab)* distance on SL5 with four purified proteins.

A. The binding capability of SL5 and its “positional mutants” correlates with the decreasing pI of the four proteins. The numerical in parentheses represents intra-(Dab)* distance. **B.** Representing the same output of five SLs with four proteins in an unguided statistical grouping, (from left to right, BSA: red, OVA: green, BSM: light blue and PSM: dark blue). Circles around each protein output is representing that each protein signal (with five SLs shown in A) is significantly different from the other with 95% confidence. **C.** Showing pI and glycan contents of the proteins that are involved in binding.

with four proteins, the binding to SLs seems to correlate with the pI of the proteins involved. For example, PSM has the lowest pI and it showed the greatest binding, followed by BSM, then OVA and finally followed by BSA (which is negative control as it has no glycans) as shown in **Figure 2.1C**.

PSM has the lowest pI because out of all proteins used, it has the maximum amount of negatively charged glycans (like sialic acid residues).⁵ It was hypothesized that the positively charged guanidinium group of Arginine (R) would form a favorable salt bridge with negatively charged sialic acid moieties commonly found on cancer related glycoproteins, as well as on two of our proof-of-concept glycoproteins (i.e., BSM and PSM). Moreover, protein with highest pI (e.g. BSA) binds the least to SL5. This is seen in **Table 2.1**, where the fold-selectivity of PSM for SL5 over BSA (negative glycan control protein) is nearly 2.5 times. As expected, the fold-selectivity of PSM for SL5 over its structurally closest mucin (BSM, containing the same type of glycans as PSM, but in a much lower quantity) represents that BSM is the next best binder to SL5 over OVA and BSA. These three results indicate that +3 charge in SL5 sequence might be more important at addressing protein to SL binding, rather than the positional placement of 2-PBA in the SL structure.

In order to comprehend whether the SL outputs with four proteins are significantly different from each other, we employed Principal Component Analysis (PCA).^{6,7} PCA is a statistical analytical method, which provides us a platform to perform unguided groupings of data. As seen in **Figure 2.1B**, the four sets of outputs with four proteins (as seen in **Figure 2.1A**) are represented by squares. The circles around the set of squares mean that the SL binding with each protein is significantly different from the other with

95% confidence. Besides this, the X-axis in **Figure 2.1B** is correlating to the pI of the proteins involved, thus hinting upon the impetus of positively charged AA residues in SL5 sequence towards protein classification.

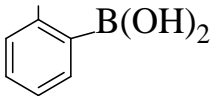
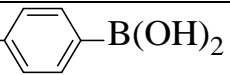
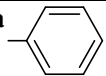
2.2 VARIATION ON SL5 BY ALTERING BORONIC ACID ATTACHMENTS

The previous experiment exemplified the importance of positively charged AA in SL5 in establishment of binding with proteins of different pI. It also indicated no significant impact on change in binding when two 2-phenyl boronic acids (2-PBA) were moved around in the SL5 peptide. This then lead to the question whether 2-PBA is involved in binding proteins to the SL or not. To address this question, several mutations to SL5 structures were made. SL5 normally has two units of 2-PBA in its structure, so will be referred to as 2-PBA (in this section) and is being taken as the control SL. Regio-isomer of 2-PBA was made and named 4-PBA (refer to **Table 2.2** for the structure of boronic acid used in this mutant). Also, a mutant named SL5-Bn was made that only had the phenyl rings and no boronic acid on them. SL5-(Dab)_{3,7A} mutant has two Dabs (which were on AA positions 3 and 7 in SL5) being replaced with alanine (a non-polar, small AA), thus having no 2-PBA in the SL. Finally, mutant SL5-Dab, has no 2-PBA but contains 2 Dabs in the SL. In totality, there are four mutants for testing four situations

1. *Alteration of boronic acid (2-PBA) position on the phenyl ring*
2. *Removal of boronic acid but retention of phenyl ring*
3. *Removal of boronic acid and phenyl ring and both Dabs*
4. *Removal of (2-PBA) but retention of both Dabs*

The structures of the mutants discussed in (**Table 2.2**)

The corresponding output from these SL5 ‘boronic acid mutants’ with 4 proteins is as

Table 2.2: Showing “SL5 Boronic acid mutants’				
#	Mutant name	Boronic acid structural mutation	Fold selectivity PSM/BSA	Fold selectivity PSM/BSM
1	SL5 ^a or (2PBA)	peptide ^a 	2.0	1.5
2	SL5 ^a -4PBA	peptide ^a - 	2.1	1.7
3	SL5 ^a -3,7Bn	peptide ^a - 	2.7	1.5
#	Mutant name	Peptide sequence mutation & no 2PBA	PSM/BSA	PSM/BSM
4	SL5-(Dab) _{3,7A} ^b	Ac-RAATRVA V -BBRM-resin	2.9	1.7
5	SL5-(Dab) ^c	Ac-RA(Dab)TRV(Dab)V-BBRM-resin	3.0	2.2

a: mutants (1, 2 & 3) contain a peptide sequence of Ac-RA(Dab)*TRV(Dab)*V-BBRM-resin. (Dab)* indicates Dab with some boronic acid or phenyl ring attached to it; **b:** SL5-(Dab)_{3,7A} mutant, has diaminobutanoic acids replaced by alanines on 3 & 7 position of peptide while counting from N-terminus represented by Bold Ala (A); **c:** (Dab) stands for diaminobutanoic acid without boronic acid attached to it.

illustrated (**Figure 2.2**) The experiment is conducted multiple times and the following discussion is the result of them together. From **Figure 2.2A** there are several things that come to the fore. Firstly, the binding paradigm of all four proteins correlated with the decreasing trend of pI of these proteins, showing highest binding with PSM and lowest with the negative glycan control (BSA).

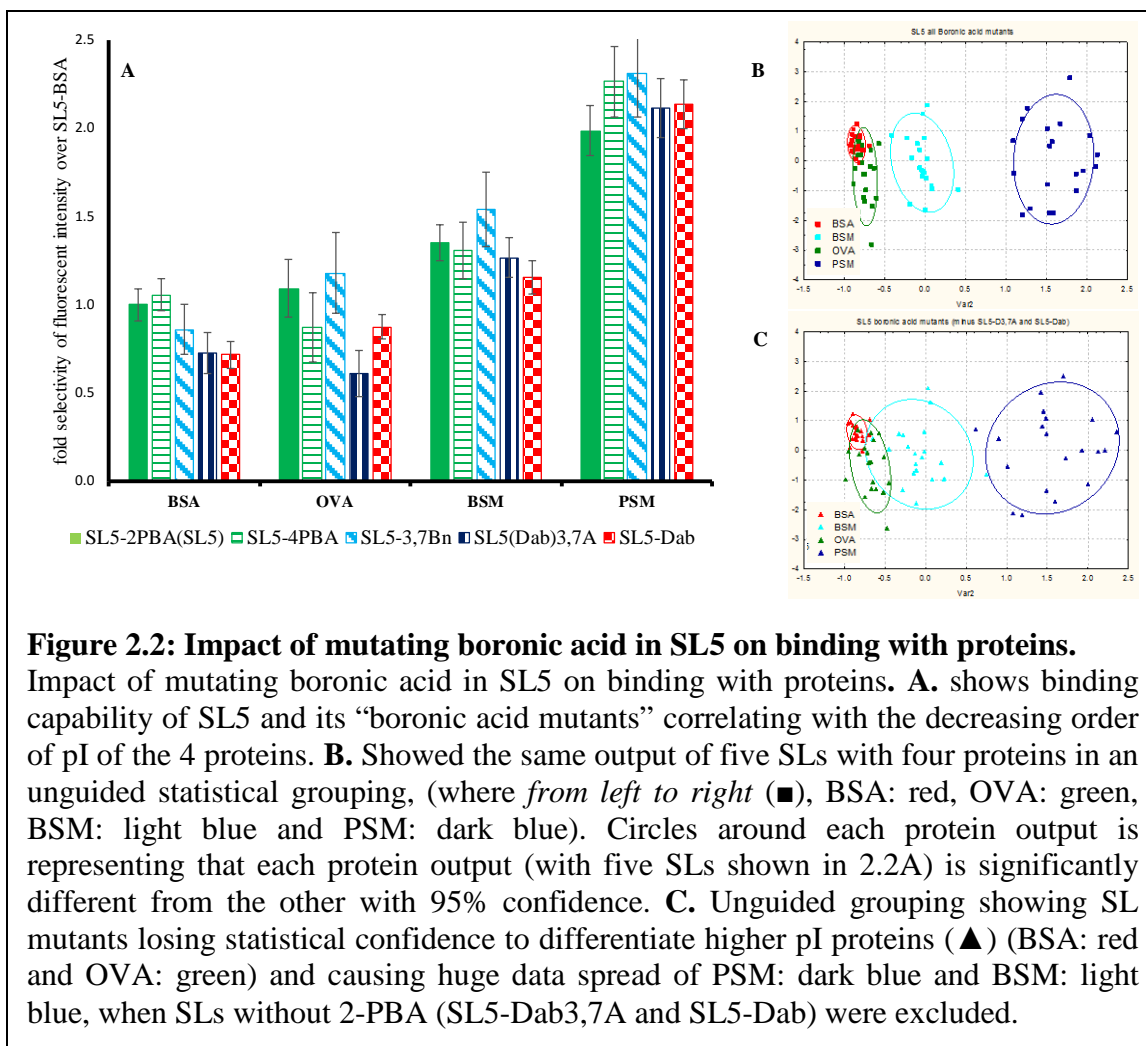


Figure 2.2: Impact of mutating boronic acid in SL5 on binding with proteins.

Impact of mutating boronic acid in SL5 on binding with proteins. **A.** shows binding capability of SL5 and its “boronic acid mutants” correlating with the decreasing order of pI of the 4 proteins. **B.** Showed the same output of five SLs with four proteins in an unguided statistical grouping, (where *from left to right* (■), BSA: red, OVA: green, BSM: light blue and PSM: dark blue). Circles around each protein output is representing that each protein output (with five SLs shown in 2.2A) is significantly different from the other with 95% confidence. **C.** Unguided grouping showing SL mutants losing statistical confidence to differentiate higher pI proteins (▲) (BSA: red and OVA: green) and causing huge data spread of PSM: dark blue and BSM: light blue, when SLs without 2-PBA (SL5-Dab3,7A and SL5-Dab) were excluded.

This correlation between binding and the pI of proteins reflect that interaction via charges of protein is the paramount driving force. PSM showed similar binding affinity with all the mutants, which indicated that presence of *neither* of the three

1. Boronic acid (2-PBA),
2. Phenyl ring of boronic acid or
3. Dab to which 2-PBA has direct attachment

impacted binding of PSM to SL5 significantly. This also indicated the importance of positively charged AA in SL5 structure towards stronger interactions with PSM over any other protein. In higher pI proteins (e.g. BSA and OVA) their binding is seen to be

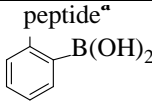
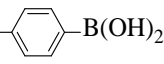
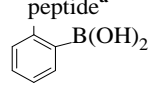
reduced with SL5-Dab (red checkered bars under BSA and OVA) and SL5-(Dab)3,7A (dark blue vertical bars under BSA and OVA) which can be seen by their fold selectivity. The reason for this decrease in binding with higher pI proteins, may be attributed to the absence of phenyl rings in SL5-Dab and SL5-(Dab)3,7A. The phenyl ring of the 2-PBA may be involved in CH- π interactions that are more clearly observed in high pI proteins, since ionic interactions are less pronounced. To further this argument, the binding of SL5-3,7Bn (having phenyl ring, but no boronic acid) mutant with BSA and OVA was very similar to the corresponding SL5-2PBA (SL5) binding to these proteins; thus indicating that CH- π interactions are at play. When SL5-2PBA signal was compared with that of SL5-4PBA, both of them appeared very similar indicating that boronic acid moiety in SL5 was not proving very effective in binding to proteins.

Figure 2.2B shows the unguided grouping of the output obtained using the five SL5s with the four proteins. Each circle is a different protein and indicates 95% confidence interval. The X-axis correlates to pI of the proteins (again indicating the important role of positively charged AA in SL5 sequence at directing greater binding towards lower pI proteins). OVA (green ■) and BSA (red ■) are overlapping, indicating that their output might not be significantly different from each other. This could be attributed to similar looking signals of SL5 mutants (SL5-Dab and SL5-(Dab)3,7A) with BSA and OVA. There signal was significantly different from the other mutants with same proteins because while plotting of the same data in **Figure 2.2A** excluding those two mutants (as shown in **Figure 2.2C**), loss of ability to separate higher pI proteins (95% confidence intervals are no longer tight) is observed. So, one could do without 2-PBA and Dab in the

SL structure. In short, simpler the design of an SL, the better is its capability in discriminating between proteins of different pI.

2.3 COMPARING DISSOCIATION CONSTANTS OF SL5 AND ITS MUTANTS

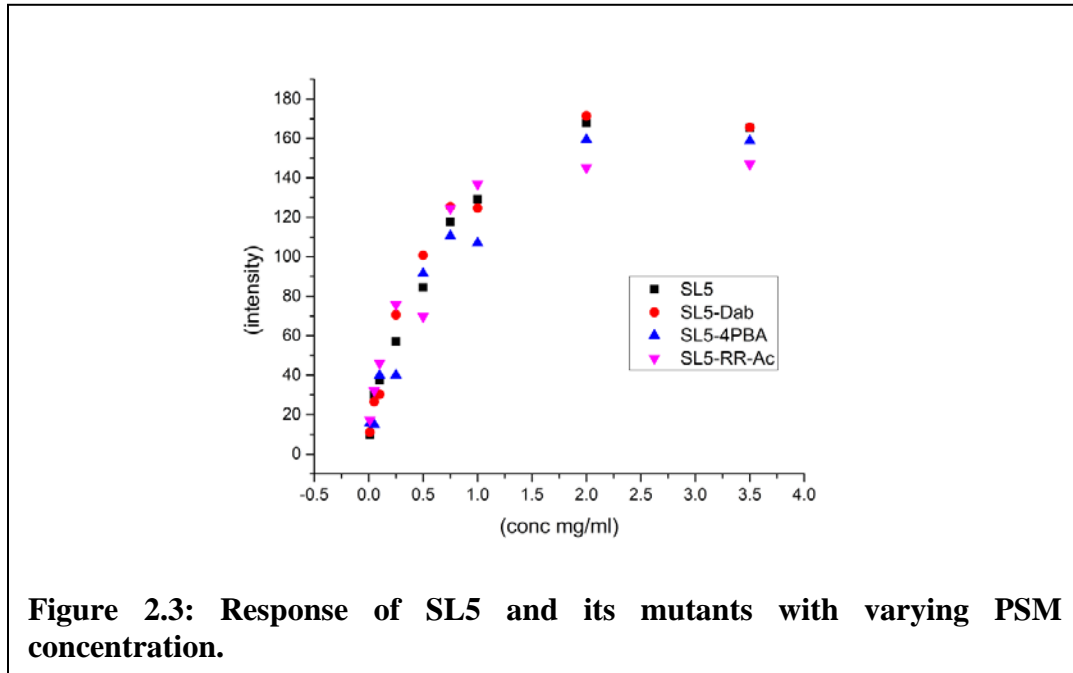
In order to verify that the 2-PBA may not be playing a role in binding of proteins to SL5, K_d values of its structural mutants (as listed in the **Table 2.3**) were obtained.

Table 2.3: showing few SL5 and its mutants involved in K_d investigation using PSM.					
Name	Peptide sequence	BA structure	# R	# +	K_d (μ M)
SL5 (2-PBA)	Ac-RA(Dab)*TRV(Dab)*V-BBRM-resin	peptide ^a 	3	3	0.325
SL5-Dab	Ac-RA(Dab)TRV(Dab)V-BBRM-resin	-	3	5	0.278
SL5-4PBA	Ac-RA(Dab)*TRV(Dab)*V-BBRM-resin	peptide ^a 	3	3	0.324
SL5-RRAc	Ac-RRRA(Dab)*TRV(Dab)*V-BBRM-resin	peptide ^a 	5	5	0.203

a: peptide sequence is of SL5: Ac-RA(Dab)*TRV(Dab)*V-BBRM-resin, where (Dab)* represented Dab with boronic acid attached to it.

These SLs are on the polymeric beads when each of them are put to incubate with different mM concentrations of fluorescently labelled PSM, ranging from 0.01 mM to 3.5 mM. The beads are imaged under a fluorescent microscope to obtain intensities that were used to run a Michalis-Menten iteration. Imaging of these incubated beads are reported at the same exposure time on the microscope camera (in order to avoid correction to

fluorescence intensity due to different exposure times). Concentrations of proteins higher than 3.5 mM, caused the fluorescence signal of the beads to saturate, thus rendering the signal detection to show plateau effect.



Previously, soluble SLs were designed and tested^{8, 28, 29} against the four proteins, their K_d values were significant ($\gg 10\mu\text{M}$), essentially pointing out the necessity of SLs to be on the bead platform to mimic some form of ‘multivalency’ while binding to proteins (which are multivalent themselves).⁹ The fluorescent intensity responses at different PSM concentrations are observed in **Figure 2.3**, where they look very similar to each other. The corresponding K_d values have been reported in **Table 2.3** and all fall in low μM range (which is usually the K_d range of natural lectins as well, i.e. high nM to low μM).¹⁰ The similarity in K_d values is unequivocally denying any particular advantage provided by 2-PBA in binding assistance with PSM. These results also provide strong evidence

that the polyvalent nature of the beads is critical and multiple SLs on a single bead are interacting with PSM to produce high affinity binding.

2.4 PEPTIDE SEQUENCE MUTATIONS ON POLAR AND NON-POLAR AMINO ACIDS

Sequence of SL5 contains polar charged AA like diamino butanoic acid (Dab), polar, uncharged threonine (T) and non-polar branched AA like valine (V) residues. In order to study their plausible impact on SL and protein binding, Alanine mutants were generated. The AA under study was replaced with a small nonpolar AA like Alanine (A) and then incubated as new 'SL mutant' with four proteins as previously described. The SL5 mutants for this study have been listed below in **Table 2.4**.

Table 2.4: Showing alanine mutations on polar and non-polar AA in SL5.

#	Mutant name	Peptide structure after BBRM-resin	#Ph or #2-PBA	+	Fold selectivity	
					PSM/BSA	PSM/BSM
1	SL5	Ac-RA(Dab)*TRV(Dab)*V-	2	3	2.3	1.8
2	SL5-Dab3,7A	Ac-RAATRVAV-	0	3	2.9	1.7
3	SL5-(Dab)*7A	Ac-RA(Dab)*TRVAV-	1	3	2.6	1.8
4	SL5-(Dab)*3A	Ac-RAATRVDab)*V-	1	3	2.9	1.9
5	SL5-V8A	Ac-RA(Dab)*TRV(Dab)*A-	2	3	3.0	1.7
6	SL5-V6A	Ac-RA(Dab)*TRA(Dab)*V-	2	3	2.8	2.0
7	SL5-T4A	Ac-RA(Dab)*ARV(Dab)*V-	2	3	2.5	1.9

(Dab)* indicates diaminobutanoic acid functionalized with boronic acid; Dab indicates diaminobutanoic acid. The naming of mutant is SL5-, AA to be replaced followed by its position in peptide sequence and then followed by AA which replaces it (i.e. A in this instance). Bold A's in the mutants are the places where A has replaced some AA in original sequence of SL5. Ac stood for acetylation of N-terminus of AA. B stands for β-Alanine. Highlighted in grey is varied # of Ph rings.

Since there are two Dabs in SL5, three mutants were designed, first one with both Dabs replaced with A, and the other two where one of the two Dabs were replaced with A. This experiment thus gave three accounts on protein binding: Firstly, the impact of absence of two Dabs, secondly the impact of position of each Dab and hence its corresponding 2-PBAs and thirdly impact of polar T and non-polar V on SL5 binding.

Replacing the number of Dabs with A, consequently led to fewer 2-PBA attachments, which led to formation of mutants having different number of phenyl rings in SL5 and was deliberately done to investigate impact of varying Ph rings on binding episode with four proteins. The corresponding results with the four proteins is as follows in **Figure 2.4**. In **Figure 2.4A**, SL5 and its mutants bound PSM the strongest, followed by BSM, OVA and BSA; showing that the binding of 'SL5 peptide sequence mutants' also correlated with decreasing order of pI of four proteins. All these mutants have +3 charge (from R AA residues) hence signifying the importance of charges in protein binding.

Figure 2.4B represents output with all four proteins to be significantly different from each other. Abscissa in **Figure 2.4B** correlated to pI of proteins, again pointing out importance of +3 AA in SL sequence for selective binding of PSM over other proteins. Comparison of PSM/BSA and PSM/BSM in **Table 2.4** represented the consistency between selectivity of SL5 mutants for PSM over other proteins.

While comparing **Figure 2.4B and C**, there was a slight decline observed in binding of higher pI proteins (BSA and OVA as shown by their tighter clustering in **Figure 2.4C**) with SL5 mutants containing no phenyl rings from 2-PBA (e.g. SL5-(Dab)*3,7A). It reflected an interplay of CH- π interactions of phenyl rings with proteins especially the

ones at higher pI. This might mean that getting rid of phenyl rings could help clean-up background binding.

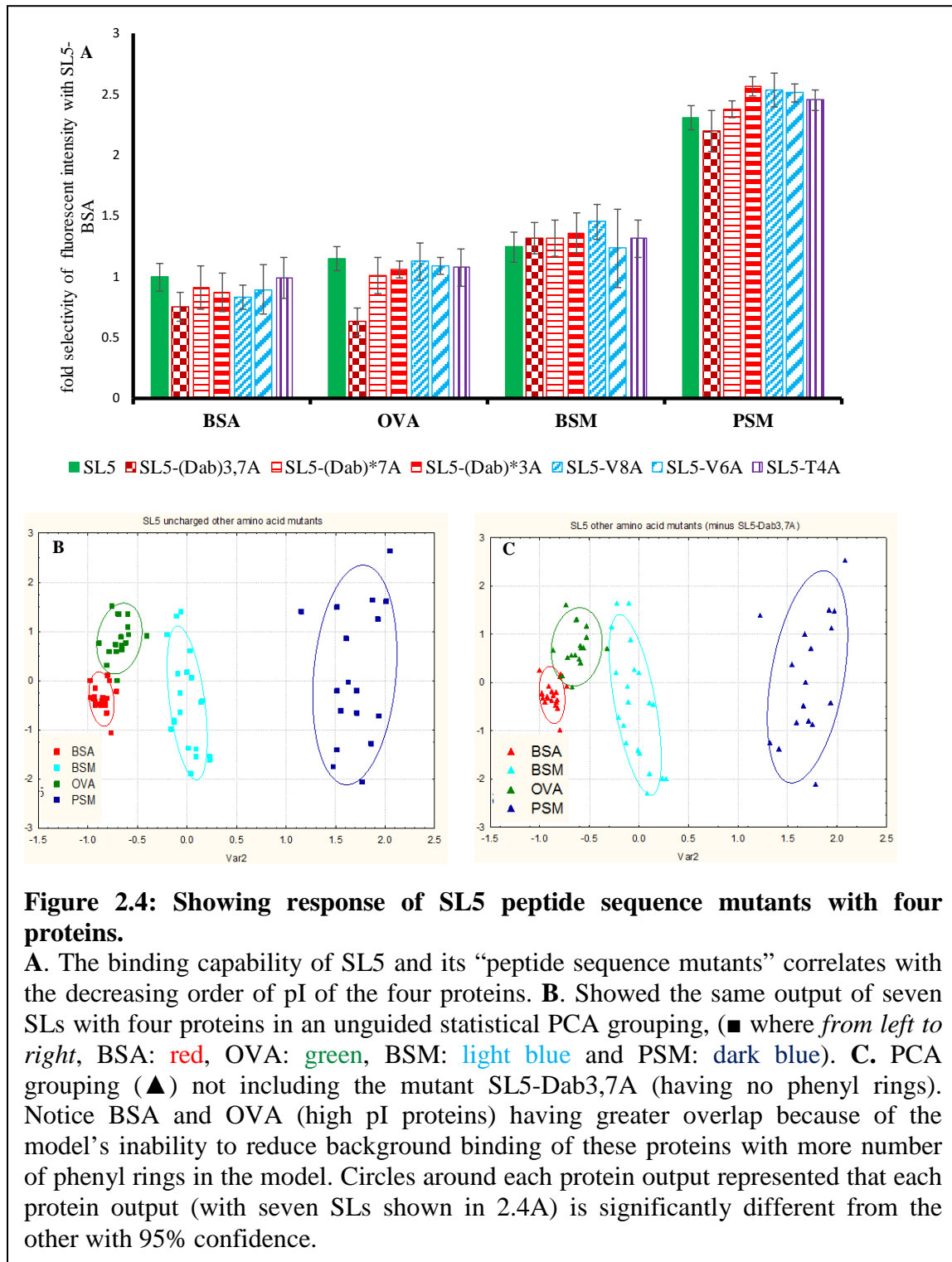


Figure 2.4: Showing response of SL5 peptide sequence mutants with four proteins.

A. The binding capability of SL5 and its “peptide sequence mutants” correlates with the decreasing order of pI of the four proteins. **B.** Showed the same output of seven SLs with four proteins in an unguided statistical PCA grouping, (■ where from left to right, BSA: red, OVA: green, BSM: light blue and PSM: dark blue). **C.** PCA grouping (▲) not including the mutant SL5-Dab3,7A (having no phenyl rings). Notice BSA and OVA (high pI proteins) having greater overlap because of the model’s inability to reduce background binding of these proteins with more number of phenyl rings in the model. Circles around each protein output represented that each protein output (with seven SLs shown in 2.4A) is significantly different from the other with 95% confidence.

In **Figure 2.4A**, SL5 and its mutants bound PSM the strongest, followed by BSM, OVA and BSA; showing that the binding of 'SL5 peptide sequence mutants' also correlated with decreasing order of pI of four proteins. All these mutants have +3 charge (from R AA residues) hence signifying the importance of charges in protein binding. **Figure 2.4B** represents output with all four proteins to be significantly different from each other. Abscissa in **Figure 2.4B** correlated to pI of proteins, again pointing out importance of +3 AA in SL sequence for selective binding of PSM over other proteins. Comparison of PSM/BSA and PSM/BSM in **Table 2.4** represented the consistency between selectivity of SL5 mutants for PSM over other proteins. While comparing **Figure 2.4B and C**, there was a slight decline observed in binding of higher pI proteins (BSA and OVA as shown by their tighter clustering in **Figure 2.4C**) with SL5 mutants containing no phenyl rings from 2-PBA (e.g. SL5-(Dab)*3,7A). It reflected an interplay of CH- π interactions of phenyl rings with proteins especially the ones at higher pI. This might mean that getting rid of phenyl rings could help clean-up background binding.

The absence of one (Dab)* in each of the two mutants SL5-(Dab)*7A and SL5-(Dab)*3A did not impact the PSM/BSM fold selectivity, hence stating that maybe boronic acids are not interacting with the diols of the glycans on proteins. In fact, the position at which (Dab)* is replaced by A does not seem to matter since the output of the two mutants with four proteins looks very similar to each other.

When the two valines, were replaced with Alanine (A) their replacement did not impact binding since Valine is non-polar and binding is expected to be dependent on positively charged AA). Similarly, presence of threonine (being polar uncharged AA) showed no impact on differential protein binding.

2.5 PEPTIDE SEQUENCE MUTATIONS ON POSITIVELY CHARGED AMINO ACIDS

SL5 has three arginine (R) AA residues in its peptide sequence. At neutral pH R has a positively charged guanidinium group, thus providing the +3 overall charge to the SL5. From previous structure and activity relationships, it is well evident that +3 charge of SL5 is strongly assisting in differential binding of purified proteins by harnessing their pI differences. This study is done to investigate

1. *The minimum amount of positive charge that begins to create an impact to the binding of proteins.*
2. *The maximum limit and hence the optimization of positive charge that may be included while keeping the design of future SLs in mind.*

Initially, we tried to investigate the binding of four proteins to acetylated tentagel resin (polymeric beads having no AA on them). The signals from all four proteins were too low and were registered under the limit of detection of fluorescent microscope (hence not included in the following figures). This meant that tentagel beads on their own had no fluorescence associated with them and they have no favourable binding preference with any of the four proteins with fluorescent tags on them. To answer the two queries posed above, we made seven new SL5 mutants that had different number of (R) AA residues in them (the charges on the mutants varied from 0 to +5). The study also included two SL5s that only had A and R AA residues on them to address ‘simplistic SL sequence design’ and its applicability. The sequences of seven ‘SL5 R mutants’ are represented in **Table 2.5**.

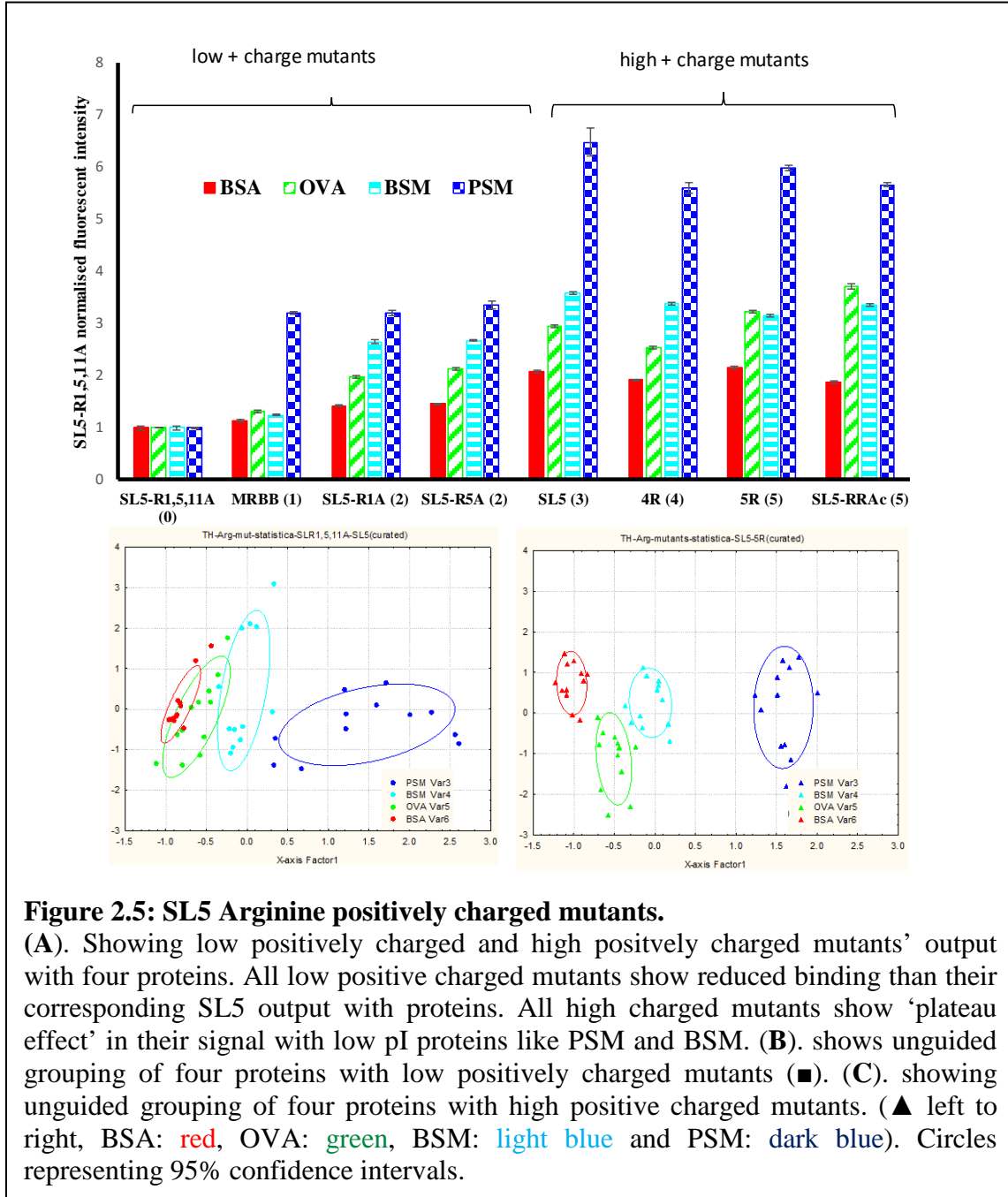
Table 2.5: Showing SL5 Arginine (R) mutants.

The naming of mutant is SL5-(X), where X is AA to be replaced followed by its position in peptide sequence, which is then followed by Ala (A).

Mutant sequences in bold are unique sequences containing different # of A and R. Highlighted in gray are the positions where A has replaced R in original sequence of SL5. Ac stood for acetylation of N-terminus of AA. B stands for β -Alanine.

Type	Name	Peptide structure	+	#Ph rings	PSM/BSA	PSM/OVA	PSM/BSM
Low positive charged mutant	SL5-R,1,5,11A	Ac-AA(Dab)*TAV(Dab)*V-BBAM-resin	0	2	0.9	0.8	0.8
	MRBB	Ac-BBRM-resin	1	0	2.8	2.4	2.6
	SL5-R1A	Ac-AA(Dab)*TRV(Dab)*V-BBRM-resin	2	2	2.2	1.6	1.2
	SL5-R5A	Ac-RA(Dab)*TAV(Dab)*V-BBRM-resin	2	2	2.3	1.6	1.3
	SL5	Ac-RA(Dab)*TRV(Dab)*V-BBRM-resin	3	2	3.1	2.2	1.8
High positive charged mutant	4R	Ac-RAARAARA-BBRM-resin	4	0	2.9	2.2	1.7
	5R	Ac-RARARARA-BBRM-resin	5	0	3.0	1.7	1.5
	SL5-RRAc	Ac-RRRA(Dab)*TRV(Dab)*V-BBRM-resin	5	2	2.8	1.9	1.9

The signal output of given eight SLs with four proteins is discussed below in **Figure 2.5**.



Low positively charged mutants: The binding of SL5-R1,5,11A (having no R amino acid in sequence) was investigated, and the data is listed in **Table 2.5**. The data shows that the mutant has the least binding with all proteins and hence its fold-selectivity for one protein over the other reported to be the smallest out of all the SLs. This showed

that loss of all positively charged AA residues led to loss in protein binding especially with low pI ones, again signifying charges being important in SL-protein binding. Hence, in **Figure 2.5A**, SL5-R1,5,11A was used as a built-in negative control for positive charged AA residues against each protein. It was used as a baseline, and output of other mutants was normalized using this SL. BBRM-resin, is a four AA long conserved residue in every SL that we designed, which essentially acts as a 'spacer' moiety between the resin bead and the 'active' portion of SL sequence. Therefore, we tested acetylated the N-terminus of BBRM-resin and incubated with the proteins; and it reported an increase in binding to PSM due to one R residue in its structure.

SL5-R1A and SL5-R5A mutants, both had an overall +2 charge (from two R AA residues) and they reported a further increase in binding with PSM over other proteins. They both have very similar outputs with four proteins and binding followed the pI trend. Having one less R than SL5, these two mutants do lose some selectivity of PSM over BSM, signifying that for optimal binding to a high pI protein mixture, one might require more than two R AA residues in SL sequence. With low positively charge mutants up to SL5, an overall trend of increased binding is observed in with PSM, BSM and OVA. To conclude, every protein bound less than SL5 with low positive charged pieces.

High positively charged mutants: When positive charge was increased more than what was originally present in SL5, peaking/ plateauing of binding signal with low pI proteins like PSM and BSM was observed. This resulted in fold selectivity of PSM/BSM to become similar to one another. The reason may be attributed to high positive charge rendering the SL 'too sticky', making it unable to differentiate between PSM and BSM. It was also interesting to note that BSA (negative glycan control) reported as the weakest

binding protein with all the mutants irrespective of the amount of positive charge or number of phenyl rings (from 2-PBA). Implying that SLs might be interacting to protein via negatively charged glycans present on low pI proteins. More on this is included as a part of discussion in Chapter 4.

2.6 CONCLUSIONS

In the experiments above, we discussed the impact of positioning of boronic acids both in the peptide and also on the phenyl ring of the phenyl boronic acid (PBA). It was observed that boronic acids e.g. 2- or 4-formyl phenyl boronic acids may not be useful in enhancing the lectin-glycan selectivity. It is hypothesized that they are too generic in design and that they do not effectively bind to hexapyranosides (natural form in which glycans exist in puckered ring). This might be the reason they do not mimic the nature's carbohydrate binding proteins i.e. lectins.¹¹

The impact of amino acids (charged, polar uncharged and non-polar) in SL-protein binding was also discussed. Our model SL (SL5) showed high specificity to PSM and the major contributor to this affinity was the negative charge of this mucin protein itself. PSM is composed of 60-90% glycans, most of those glycans are sialic acid residues (negatively charged glycan) and it might be these that are probably being bound to the positively charged arginine (R) residues. Other mucinous protein (BSM) also has sialic acids, but in a lot less quantity than in PSM and hence came second at responding to SL5. High pI proteins like OVA, have glycans, but uncharged polar ones, thus coming third to binding with SL5, ultimately followed by BSA.

If the positive charge of SL5 mutants was kept constant (as +3) and the number of phenyl rings were reduced from the peptide sequence (by removing 2-PBA), we observed

a decrease binding with high pI proteins like OVA and BSA, thus confirming a possible presence of CH- π type binding between phenyl rings of SL and polar carbons of glycoproteins. These weak interactions are more prominent only in high pI proteins.

It is of interest to see if we could engineer SLs that bound more to low pI proteins, therefore a few ‘simplified’ SLs were synthesized in order to pin-point some direct correlations rather than keeping the discussion speculation-based. SLs like 4R and 5R were mutants made entirely out of alanine and four and five arginine residues respectively (sequence information in **Table 2.5**). These were made in order to test if simplification of the SL design made any impact on binding to purified proteins.

In **Figure 2.5B**, when low positively charged mutants (SL5-R1,5,11A; MRBB; SL5-R1A; SL5-R5A and SL5) were used the PCA was shown to be quite scattered. The low positive charge model did not perform so well at classifying high pI proteins (red, green and light blue circles in **Figure 2.5B** show greater overlap). The most probable reason for this is the absence of adequate amounts of R in the SLs. The other reason may be that four out of five SLs in this model, have phenyl rings in them (coming from 2-PBA), thus increasing the CH- π background interactions with OVA and BSA, making the grouping poor. Another reason can be that low positively charged SLs can be understood as weak acids and high pI proteins can be considered as weak biological bases. Being weaker acids, low positively charged mutants are not doing a good job at classifying weaker bases. Low positively charged mutants, thus being weaker acids can only bind to strong bases like low pI proteins and group them better.

Whereas, in **Figure 2.5C**, when high positively charged mutants (SL5, 4R, 5R and SL5-RRAc) were used a tighter grouping was observed. The tighter statistical grouping is attributed to

- a) greater number of R residues;
- b) two out of four SLs being used in this model were devoid of any phenyl rings, causing reduction in background binding; and
- c) high positively charged SLs behave as strong biological acids and hence are able to classify strong (PSM) as well as weak bases (BSM, OVA and BSA).

This led to formation of a hypothesis for **DESIGNING SLs** in future, which is as follows:

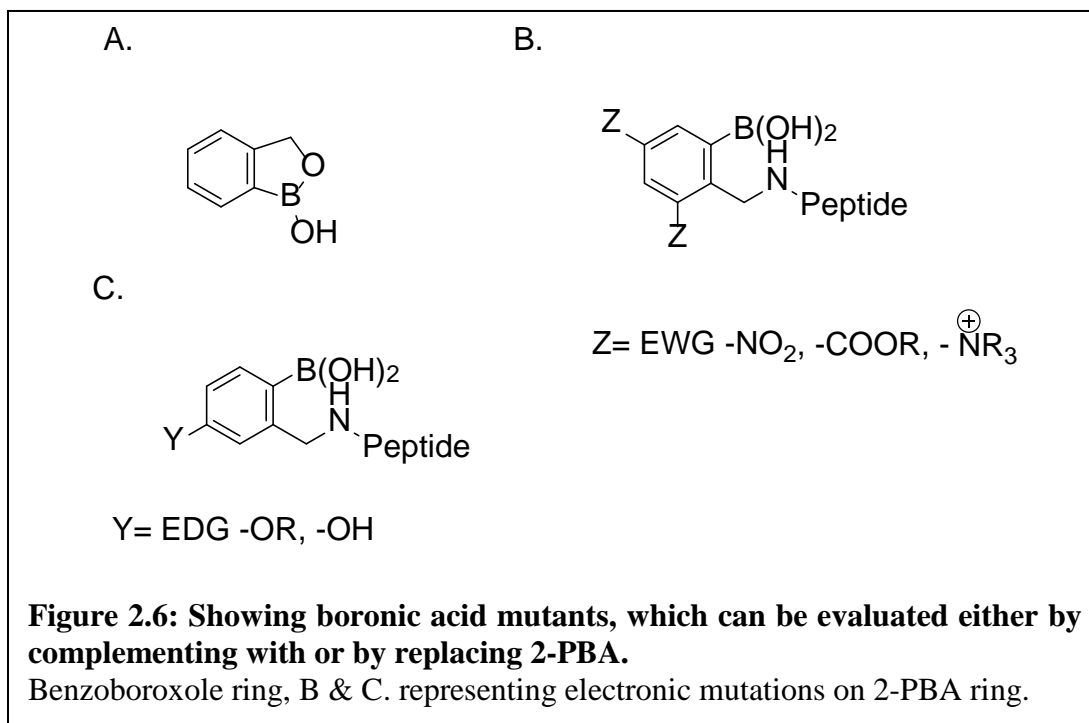
To increase the binding affinity with low pI glycoproteins, use a greater number of arginine residues (3-5) and reduced the number of phenyl rings (0-2) in SL sequences. Conversely, to increase the binding affinity with high pI glycoproteins, use fewer arginine residues (0-2) and increase the number of phenyl rings (3-6).

Using this hypothesis, several SLs were made (DESIGNED SLS) and tested against human proteins extracted from colon, prostate and breast cell lines. These new SLs were evaluated for their ability to discriminate the proteins with varied metastatic potential, to further our hypotheses into a working model with applications to investigate cancers. The discussion shall be covered in the next chapter.

2.7 FUTURE DIRECTIONS

Currently, testing for a different PBA complimentary to 2-formyl phenyl boronic acid is required to establish further structure-activity relationship. The Hall group, have designed peptides with bis(benzoboroxoles) as their glycan binding site.¹⁰ We can also

use these systems (**Figure 2.6A**) and incorporate them in our lectins because benzoboroxozoles target hexopyranosides/ oligosaccharides (e.g. TF Antigen) on naturally occurring glycans in the body which simple phenyl ring substituted boronic acids do not. Phenyl boronic acids reported greater binding affinity with furanose sugars and not so much with glycopyranosides in glycoproteins, which have puckered rings.¹⁰



We can investigate the impact of binding by changing the substituents on the phenyl boronic acid moiety. Since the boron has an empty p-orbital, having an electron-withdrawing group (Z) on the phenyl ring of the boronic acids may enhance the binding of the SLs to glycans by increasing the Lewis acidity of the boronic acid. Hence an investigation into the electronics of phenyl boronic acids could be investigated by adding multiple electron rich (EWG) or electron poor (EDG) functional groups on to phenyl boronic acid (**Figure 2.6 B and C**).

Extensive investigation of Phenyl-Ring scanning mutagenesis can be undertaken during which non-polar (A, V) and polar (R, Dab, T) amino acids on SL5 can be point mutated with Phenylalanine (F). These mutants can also be directly investigated by their binding with cancer-associated glycans, e.g. with fucose. These mutants can provide insight into the possible increase in background binding with high pI proteins. The size of the aromatic surface phenyl versus naphthyl or anthryl, can be investigated as well (e.g. F versus W).

It is still unclear whether binding affinity of the platform (in both possible cases of R-glycan salt bridge or CH- π type interaction) is univalent or multivalent. Using NMR to investigate specific SLs for their propensity to bind pyranosides like fucose shall provide direct insight into plausible CH- π type interactions. The challenge again is to establish a proof-of-concept stoichiometric platform.

2.8 EXPERIMENTAL PROCEDURES

Required Chemicals for synthesis of SLs

2-Formyl phenyl boronic acid (2-PBA) and its regioisomer (4-PBA) were purchased from Combi-blocks. All amino acids (AA) and O-Benzotriazole-*N,N,N',N'*-tetramethyl-uronium-hexafluorophosphate (HBTU) were purchased from Novabiochem Corp and VWR. Glycoproteins (porcine stomach mucin, bovine submaxillary mucin and ovalalbumin), bovine serum albumin and Fluorescein isothiocyanate (FITC) were purchased from Sigma-Aldrich. TentaGel resin (Cat. No. MB-300-002; loading capacity 0.25- 0.3mmol/g) was purchased from Rapp Polymere. All other chemicals (e.g. solvents for coupling amino acids) were purchased from Sigma-Aldrich and Acros Organics. These chemicals were used without any further purification.

General Method for Synthesis of SL5 and its Mutants

1. Peptide synthesis on Polymeric Tentagel beads

To synthesize SLs, we used an automated protein synthesizer machine PS3, which operated under nitrogen at RT to create vacuum for a moisture-free peptide chain growth using Fmoc chemistry. All of the SLs started with the conserved sequence of first four amino acids, BBRM, before introducing randomized amino acids into library. B stood for β -alanine and was used instead of alanine because this region-isomer had a longer main chain, which offered longer spacer link between bead and the AA that followed B. Acetylated arginine was used to cap the $-\text{NH}_2$ ends of SLs. The ivDde protected diamino acids were used for introducing boronic acids to the peptide backbone by reductive amination. TentaGel resin beads (300 μm) were used in synthesizing all peptide based boronic acid containing synthetic lectins. Standard Fmoc/HBTU chemistry employed to extend the peptide backbone. The amino acids used had Fmoc protected amines and the reactive side chains of the amino acids remained protected by acid-labile protecting groups. About 20mg (0.05mmol) of resin was weighed and soaked in DMF for 10 minutes. 8 equivalents of methionine (0.4mmol, 150.6 mg) and HBTU (0.4mmol, 151.72mg) were dissolved in 10mL of 5% *N*-methyl morpholine in DMF for 10 min and added to the beads. After tumbling 45min, the beads were washed with DMF, methanol and DMF. To remove the Fmoc group, 10mL 20% piperidine in DMF was used. Using the same procedure, arginine, β -alanine followed by another β -alanine was coupled. The amino acids utilized in the later steps were dependent on the lectin being resynthesized. After addition of all the amino acids by the desired sequence, acetylated arginine was

used to cap the peptide. The acylation of N-terminal was done using 10mL of CH₂Cl₂ containing 5% pyridine and 5% acetic anhydride.

2. Removal of protecting group on side chain –NH₂ of Dab

Diaminobutanoic acid (Dab) has a 1-(4,4-Dimethyl-2,6-dioxocyclo-hexylidene)-3-methylbutyl (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 5% (v/v) hydrazine monohydrate in DMF and tumbling for 1 hour.

Incorporation of Phenyl Boronic Acid onto Dab side chain

2-Formylphenyl boronic acid (8eq, 0.8mmol, 119.95mg) was dissolved in a mixture of 0.2mL methanol and 9.8mL DMF (10% MeOH in DMF solution). The resin beads, along with activated 3Å molecular sieves, were added into the solution and tumbled at 37°C overnight. NaBH₄ (8eq, 0.8mmol, 30.26mg) was added, H₂ gas evolved and was allowed to escape for the next 45-minutes followed by tumbling the resin beads at 37°C for 4-hours. This was followed by washing the beads with copious amounts of DMF, methanol, and DMF again.

To make boronic acid mutants (SL5-4PBA and SL5-Dab3,7A resp.), 8 equivalents of 4-formylphenylboronic acid and benzaldehyde were used in place of the usual 2-formylphenylboronic acid.

3. Removal of Acid Labile groups on side chains of other amino acids

Molecular sieves were removed and the resin beads were transferred to another reaction vessel. After the boronic acid moiety was coupled, 10mL of 95% TFA, 2.5% water, and 2.5% triisopropylsilane (9.5ml, 250µL, 250µL resp.) was added and the resin was tumbled for 1 h to deprotect the acid-labile protecting groups. The TFA wash was

repeated one more time. The resin was then washed with DMF, methanol, and DMF (these beads could be stored in freezer for longer period -30°C). In order to store the SLs in fridge under 4°C temperature (to reduce washing steps before incubation), we tumbled the beads with PBS without glycerol (pH 7.25) twice for 10 minutes.

Buffer Solution Preparation

1. Phosphate Buffered Saline (PBS) with and without glycerol:

Sodium phosphate monobasic (monohydrate) (27.6g, 200mmol), sodium phosphate dibasic (anhydrous) (28.4g, 200mmol) and sodium chloride (35.06g, 600mmol) were dissolved in 3.5L of distilled water. 400mL of glycerol was added and the pH was adjusted to 7.3 using 3M NaOH. Glycerol was skipped to make regular PBS. All these steps were done at the RT. The volume of NaOH was recorded and the remaining volume of deionised water was added to make the total volume 4L. In general, PBS (without glycerol) was used for all washings, before and after the protein incubation, but PBS with 10% glycerol was used to make screening/pre-blocking buffer using by adding 1% BSA. This screening buffer was also used to dilute labelled purified protein solutions to 0.1mM concentrations before incubation.

2. Carbonate buffer

To prepare carbonate buffer, 64g sodium carbonate (604mmol) and 117.2g sodium bicarbonate (1390mmol) were dissolved in 3.9L of de-ionized water, and the pH was adjusted to 9.8 by the addition of 3M NaOH. The volume of NaOH was recorded, and the solution was then diluted to 4L.

MALDI Sample Preparation

Phenyl Boronic Acid (PBA) Group Removal from SL

This step needs to be done if you have already attached boronic acids to all the resin beads. If PBA group was not cleaved off the resin beads, the MALDI spectra is very dirty with no peaks. PBA group removal step may be avoided, if a small amount of beads were taken out before the boronic acids are added. Once removed, the rest of the MALDI steps were followed on this small amount of bead sample. However, if the boronic acids were still attached and a MALDI was needed to be run these groups will need to be removed. Remove PBA groups by adding 100mmol solution of H₂O₂ to the 1.5mL Eppendorf tube containing the beads. Tumble this tube at 50°C for an hour. After the hour, wash the beads with dH₂O three times. Added 50µL of a 40mg/mL solution of Cyanogen Bromide (CNBr) made in 0.1M HCl. All work with CNBr should be performed under the hood. Placed this solution in the dark (usually your drawer) for 15-18 hours. Evaporated this solution using temperature controlled speed vacuum centrifuge. There should be a small amount of white substance at the bottom of the centrifuge tube.

MS Sample preparation and spotting

Dissolved the sample obtained from the previous section in 10µL of .1% TFA in H₂O. In order to make the matrix, make a saturated solution of α-cyano-4-hydroxycinnamic acid (CHCA) in 50:50 acetonitrile/0.1% TFA in H₂O. Added 1µL of sample to the desired spot on the plate for MALDI. Added 1µL of matrix to the same spot. Pipette the 2µL of solution up and down a few times to mix matrix and sample uniformly. Let the plate dry and take obtain MALDI spectra.

Fluorescent Labelling of Purified Proteins/Glycoproteins

To prepare the FITC labelled glycoproteins, 50mg of the desired glycoprotein (e.g. OVA, BSM, or PSM) was dissolved in 9mL of 0.5 M carbonate buffer (pH= 9.8) and 2

mg of FITC was dissolved in 1 mL of DMF. The ratio of protein to FITC was kept 50:2 and ratio of carbonate buffer to DMF was kept 10:1. The two solutions were combined together in a 15mL Falcon tube and tumbled at 37°C in dark for an hour.

There were two methods used to remove the excess unbound FITC, one being dialysis and the other being concentrating using centrifuge tubes. Dialysis method was typically used for very high batch concentrations of purified proteins, for example to make 500mg labelled protein solutions and centrifuge concentration was typically used for very small protein, cell membrane/tissue extracts, secreted protein concentrations.

Dialysis method- The solution was then transferred to a 12,000 MW cut-off dialysis tube and dialyzed in 0.1M PBS (pH = 7.3). The dialysis buffer was changed every 4-hours until dialysis was complete. That was ensured by checking the fluorescence of the buffer solution, when the buffer ran clear of any visible fluorescence under UV lamp the dialysis went to a completion.

Centrifuge method- For smaller protein concentrations, we would employ centrifugal spin tubes with a MW cut-off at 10kDa (saved time with small concentration of proteins). This protocol is courtesy of Erin E. Gatrone and Kathleen O. Connell. The membrane proteins/glycoproteins and secreted proteins were labelled separately with fluorescein isothiocyanate (FITC) in DMF and carbonate buffer. Herein we took 2.8mg of FITC and was dissolved in 100µL of DMF. 10µL of the FITC solution was put into 200µL of protein. The total volume of the solution was brought up to 600µL by adding 400µL of carbonate buffer. The protein solution was tumbled for 1-hour at 37°C and later purified using Amicon® 10K centrifuge spin tubes. The solution was later put into a tube and spun for about 8 minutes at 4000RPM. These tubes were devoid of minimum cut-off protein

level and risked into complete drying out of the protein, thus causing protein loss. Hence one had to keep an eye on the time and the scale on the spin tubes. The time spun may change depending on protein concentration, and they were to spin until it read “3” line on the tube. At that point, 1.5mL of PBS was put in the exchange tube and spun for 5 minutes. Again it was checked to ensure the solution is at the “3” line. At which point, the small tube was turned inside over and spun out into a micro centrifuge tube on a benchtop centrifuge. Added 300 μ L of PBS and pipette up and down. Collected all the protein. Repeat the rinse step if you were using secreted proteins since there is a large concentration of protein. If not, and you are using membrane extracts there would be as big of a concentration and thus maybe rinse with 2-100 μ L increments of PBS. The concentration of the FITC-glycoprotein was determined using the BCA assay.

BCA Assay

Thermo Scientific BCA kit was purchased from VWR. Standard 2% BSA protein solution in PBS was purchased from VWR.

Known Protein Standard Preparation

BCA kit also comes with a standard 2mg/mL BSA solution. Added 100 μ L of this solution into a small plastic tube. Put 50 μ L of PBS in several other tubes. Then take 50 μ L of the 2mg/mL solution in the first tube. Vortexed each solution every time. Usually a 0 2mg/ml series were run (2, 1, 0.5, 0.25, 0.125, 0.625, 0.3125 and 0 mg/ml). Then took 50 μ L of that solution and put it in the next tube. Continue this process.

Reagent Preparation

Follow the directions that instruct to make the reagent solution. It would be made in a 50:1 reagent B: reagent A ratio. 200 μ L of reagent is required for each well so the amount of total reagent will need to be calculated (200 μ L x # of wells).

96-Well Plate Preparation

Typically, triplicate sampling of the standards/samples are encouraged. Put 10 μ L of protein in each well of standards and samples. Then added 200 μ L of 50:1 reagent mix in all of the wells. The plate should then be placed in the oven at 37°C for 45-minutes or until color development ceases in the standards (around 1.5-hours). The plate was then read on the plate reader with the well plate attachment under end-point absorbance at 562nm. BCA assay is not an endpoint assay and continues to develop over the time.

General Wash and Incubation Procedure for Purified Labeled Proteins

Beads were weighed (2mg) and rinsed with 1mL of PBS with 10% glycerol in it (this step was avoided if the resin beads were already washed with PBS and kept in 4°C refrigerator. The resin was then pre-blocked with screening buffer i.e. PBS with 10% glycerol and 1% bovine serum albumin (BSA) for 15 min. The beads were then incubated with 1mL of 0.1mg/mL FITC labelled proteins (used screening buffer to dilute the protein solutions) for a minimum of 12 hours and maximum of 20-hours at RT (25°C), and then washed with PBS three times. The beads were re-suspended in some PBS to make it easier to pipette beads on to the microscope glass slides. Images of these beads were then taken under microscope.

Microscope Usage and Data Acquisition

All images were taken with a Leica MZ 16F microscope and a GFP filter set (excitation 450-490 nm; emission filter 500-550nm). The image data is in a 10-bit scale.

For a direct comparison, it is suggested that shutter exposure time, intensity of the bulb and magnification of the lens were kept constant throughout. Images were taken with Q Capture[®] to obtain red, green blue and bright values for each bead. Optimize the slide to have the largest value under the green channel without saturating the pixels on the camera screen (value should not exceed more than 190, 255 is the absolute max). A rough estimate of this number could be checked by drawing a circle around a bead whose intensity you are interested in under Adobe Photoshop[®] and pulling up the mean value of the area under the green channel in histogram information. It was suggested that the resin beads remain wet with PBS at all times, which helped in reduction of 'bead reflectance' on glass and kept the beads spread apart without sticking to one another.

The images obtained were in tiff format, which were converted to jpeg format before for further data acquisition. To perform batch conversions on image files, we used Irfanview[®]. These jpeg files were used to churn out mathematical values for each bead by entering the image inputs into an algorithm written using MATLAB[®] 2013 program (provided by the courtesy of Dr. Andrew Greytak) and edits by Anna A. Veldkamp. The MATLAB[®] code, once run, then gave valid consistent numerical values for fluorescence (under specified microscope intensity, exposure, magnification and background correction), hence the human error was reduced that used to occur (in protocol followed by predecessors Kevin L. Bicker and Jing Sun) because of manually drawing circles around images of beads to obtain a value for fluorescence. This code picked out each bead and provided information about its area, mathematical value under red, green, blue and composite (bright) channel. It also told information about location of the bead, which helped to identify each bead. It also gave a circularity value for each bead, closer it was

to value of one, more spherical was the bead, which was helpful to rule out broken and clumped beads as well as 'floaters'. All the bead imperfections contributed to mathematical outliers. These outliers were removed (1.5 IQD) in excel in order to get the array information for the SLs with each protein type. At this point, the data was entered into varying statistical models the most common seen here in the chapter was principal component analysis (PCA). The average of these values was taken and compared. High average values meant brighter beads, which meant stronger binding affinity to proteins.

2.9 REFERENCES

1. Zou, Y.; Broughton, D. L.; Thompson, P. R.; Lavigne, J. J., Peptide Borono-Lectins (PBLs): A New Tool for Glycomics and Cancer Diagnostics. *ChemBioChem*. **2007**, *8* (17), 2048-2051.
2. Wright, A. T.; Anslyn, E. V., Differential receptor arrays for solution-based molecular recognition. *Chem. Soc. Rev.* **2006**, *35*, 14-28.
3. Wright, A. T.; Griffin, M. J.; Zhenlin, Z.; McCleskey, S. C.; Anslyn, E. V.; McDevitt, J. T., Differential receptors create patterns that distinguish various proteins. *Angew. Chem. Int. Ed.* **2005**, *44* (39), 6375-6378.
4. Edwards, N. Y.; Sager, T. W.; McDevitt, J. T.; Anslyn, E. V., Boronic acid based peptidic receptors for pattern-based saccharide sensing in neutral aqueous media, an application in real-life samples. *J. Am. Chem. Soc.* **2007**, *129*, 13575-13583.
5. Amerongen, A. V. N.; Bolscher, J. G. M.; Veerman, E. C. I., Salivary mucins: Protective functions in relation to their diversity. *Glycobiology* **1995**, *5* (8), 733-740.
6. Anslyn, E. V.; McCleskey, S. C.; Griffin, M. J.; Schneider, S. E.; McDevitt, J. T., Differential receptors create patterns diagnostic for ATP and GTP. *J. Am. Chem. Soc.* **2003**, *125* (5), 1114-1115.
7. Anslyn, E. V.; Wright, A. T.; Griffin, M. J.; Zhong, Z. L.; McCleskey, S. C.; McDevitt, J. T., Differential receptors create patterns that distinguish various proteins. *Angew. Chem. Int. Ed.* **2005**, *44* (39), 6375-6378.
8. Moerke, N. J., Fluorescence polarization (FP) assay for monitoring peptide-protein or nucleic acid-protein binding. *Curr. Protoc. Chem. Biol.* **2009**, *1*, 1-15.
9. Mammen, M.; Choi, S.-K.; Whitesides, G. M., Polyvalent interaction in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew. Chem. Int. Ed.* **1998**, *37*, 2754-2794.
10. Hall, D. G.; Pal, A.; Berube, M., Design, Synthesis, and Screening of a Library of Peptidyl Bis(Boroxoles) as Oligosaccharide Receptors in Water: Identification of a Receptor for the Tumor Marker TF-Antigen Disaccharide. *Angew. Chem. Int. Ed.* **2010**, *49* (8), 1492-1495.
11. Berube, M.; Dowlut, M.; Hall, D. G. Benzoboroxoles as efficient glycopyranoside-binding agents in physiological conditions: structure and selectivity of complex formation. *J. Org. Chem.* **2008**, *73*, 6471-6479.

CHAPTER 3:
EMPLOYING DESIGNED SYNTHETIC LECTINS IN ARRAY TO
EVALUATE METASTATIC POTENTIAL OF HUMAN COLON,
AND PROSTATE CANCER *IN-VITRO*

3.0 OVERVIEW

In the previous chapter, the binding of Synthetic lectin 5 (SL5) with several proteins like Porcine stomach mucin (PSM), Bovine submaxillary mucin (BSM), Ovalalbumin (OVA) and Bovine serum albumin (BSA) was discussed. SL5 had higher selectivity for low pI protein like PSM and this was mainly attributed to the +3 charge from the three Arginine residues in the SL5 peptide sequence that were in likelihood, mediating greater affinity for sialic acid residues present in PSM. The binding of SL5 correlated with the order of pI of the proteins. But, *can this charge based discerning capability of SLs also be used to quantify changes in glycoprotein charges, which accompany oncogenesis and/or progression of cancer?* To answer this query, an array of several SLs that could discriminate between healthy and cancerous colon proteins was designed previously (section 1.10). The SLs that were used in making the array previously, had the same number of charges. To assess which chemical attributes of the SL array (charges, phenyl rings, boronic acids & multi-valency) contribute to the differential binding of healthy and cancerous proteins, an *extended lectin array* was

made. Extended lectin array included the previous SLs as well as new designed SLs with various permutations of charge and 2-PBA (section 2.6). listed in **Table 3.1**.

The aim of this chapter is to investigate:

1. Response of extended SL array to different analytes from human colon cell lines
 - I Cell membrane proteins (ref: Section 3.1)
 - II Cell-secreted proteins (ref: Section 3.2)
 - III Cell membrane and cell secreted proteins together (ref: Section 3.2)
2. Role of 2-formylphenylboronic acid residues on enhancing arrays' ability to classify cell lines with different metastatic potential (ref: Section 3.4)
3. Response of extended lectin array to human prostatic cell lines (ref: Section 3.5)
4. SLs that best classify different colon and prostate cell lines and discuss the statistical basis of classification (ref: Sections 3.2, 3.3 and 3.5)

Table 3.1: Designed SLs after structure and activity relationship evaluation. Showing SLs with no positive charge and containing 2-formyl phenyl boronic acids (2-PBA); or high positive charges with few 2-PBA; or high positive charges but no 2-PBA. LC MS in APPENDIX A

SL Name	Peptide structure	# R	# 2-PBA	+	comments
SL5-R,1,5,11A	Ac-AA(Dab)*TAV(Dab)*V-BBAM-resin	0	2	0	no positive charge (with phenyl boronic acids)
SL5-RRAc	Ac-RRRA(Dab)*TRV(Dab)*V-BBRM-resin	5	2	5	positive charge (with phenyl boronic acids)
4R	Ac-RAARAARA-BBRM-resin	4	0	4	positive charge (no phenyl boronic acids)
5R	Ac-RARARARA-BBRM-resin	5	0	5	
SL1-Dab	Ac-RG(Dab)VTF(Dab)R-BBRM-resin	3	0	5	
SL2-Dab	Ac-RT(Dab)RFL(Dab)V-BBRM-resin	3	0	5	

SL3-Dab	Ac-RS(Dab)VTT(Dab)R- BBRM-resin	3	0	5
SL4-Dab	Ac-RR(Dab)TQT(Dab)Q- BBRM-resin	3	0	5
SL5-Dab	Ac-RA(Dab)TRV(Dab)V- BBRM-resin	3	0	5
#R: indicates number of Arginine residues in peptide of SL. (Dab)*: diamino butanoic acid plus 2-PBA and (Dab): diamino butanoic acid only. B: β -Alanine. Highlighted grey SLs were the part of initial extended lectin array.				

3.1 INVESTIGATING COLON CANCER USING EXTENDED LECTIN ARRAY AND PROTEINS EXTRACTED FROM COLON CANCER CELL MEMBRANES

The initial make-up of the extended lectin array included the original five SLs (SL1-SL5) including the ones highlighted in grey (making 9-unit array) in the **Table 3.1**. Over the course of experiments, we increased the scope of the SL array to investigate a broader array of tissue types and subsequently included additional SLs (making a 22-unit array, some of which were made to separately investigate prostate cancer and have been summarized in Appendix B). The un-highlighted SLs in **Table 3.1** did not contain any 2-PBA and were included later for experiments in section 3.4 that required SLs with negative 2-PBA controls.

Proteins from the membranes of human colon-derived cell lines were extracted, tagged, quantified and then incubated with the SL array beads. These beads were then imaged using fluorescent microscopy and the intensities of individual beads in the images were compared and analyzed in the same way as previously described in chapter 2. The colon cell lines used are shown in **Table 3.2**.

Human colon cell line	Type	Metastatic potential
CCD 841 CoN (CONA)**	healthy/normal ¹	NA
HCT-116	low metastatic ²	<10%
HT-29	low metastatic ¹	<10%
LoVo	high metastatic ³	>50%

**CCD 841 CoN abbreviated as CONA

Previously, the metastatic potential of human colon cell lines was investigated using SL1-SL5 array; and that provided a guided plot where the SLs were able to discriminate colon cell lines into 3 categories (healthy, low metastatic and high metastatic) with 89% classification accuracy. When the extended lectin array was used for the same cell lines, the 3-class guided grouping (after repeated experimentation with different biological replicates) increased to 97% (**Figure 3.1A/B**); where the ‘designed SLs’ proved the most useful increased the classification accuracy (**Table 3.3**).

SL Name	# of positive charges	# of (2-PBA)	# Ph rings	F-score
4R	4	0	0	83.42
SL4	3	2	2	79.56
SL5-Dab	5	0	0	51.03
SL5-RRAc	5	2	2	7.01

The SLs that were used in the extended lectin array to distinguish proteins extracted from human colon cell membranes; were SL1-SL5 plus the designed SLs highlighted **grey** in

Table 3.1. Table 3.3 shows four top ranking SLs arranged in terms of their decreasing order of grouping ability. The grouping ability of an SL is explained by its F-score, where greater the F-score of SL, more significant is its contribution to a model.

The output of extended lectin array with colon proteins is represented in **Figure 3.2.** It showed several SLs specific for a cell line and several cross-reactive. In order to assess the applicability of array, this output was statistically grouped using linear discriminant analysis (LDA) and shown in **Figure 3.1.1.** LDA is a guided statistical plotting, where the data-points are input in a way that they are assigned to some group (e.g. healthy, low metastatic etc.) even before the analysis. The program then tries to cluster the similar

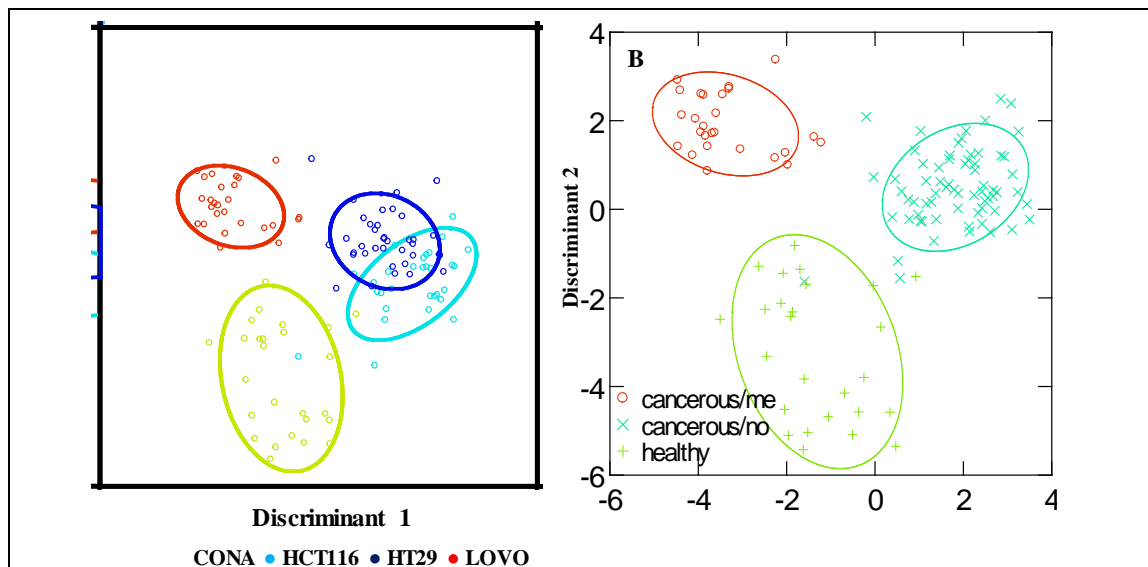
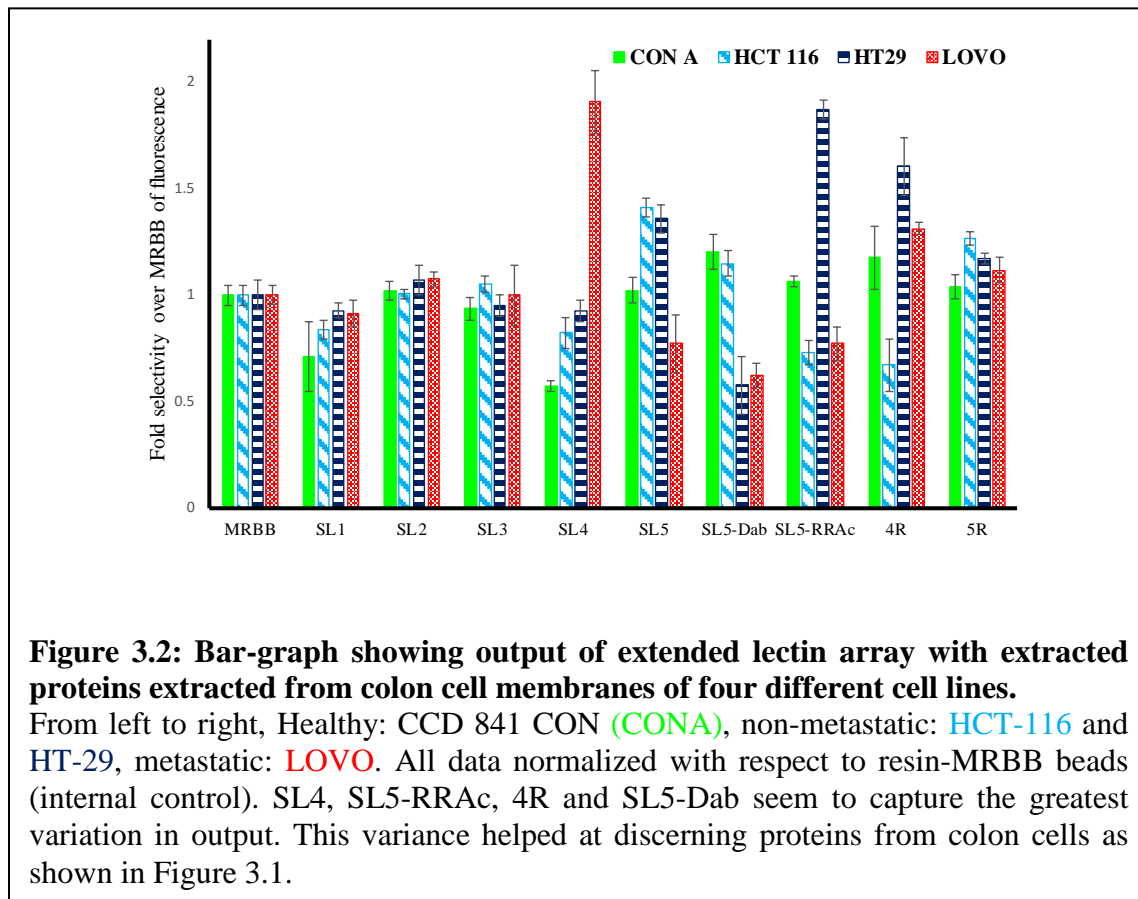


Figure 3.1: Guided statistical plot (LDA) generated using extended lectin array with proteins extracted from cell membranes of four different colon cell lines.

Each data point refers to cumulative output from 10 SLs in extended lectin array. The circles around the data points are 95% confidence intervals. Where number of points N= 119. A: 4-class grouping within different cell lines (HCT-116 and HT-29 low metastatic cell lines overlap each other); healthy (CON A): bottom yellow (○); high metastatic (LOVO): top red (○). B: 3-class grouping with respect to metastatic potential of cell lines; healthy (CON A): green (+); low metastatic (HCT-116 & HT-29): blue (X); high metastatic (LOVO): red circles (○)

data-points closer and separate data-points away from the other cluster.

In **Figure 3.1A**, the array could distinguish between the proteins from 4 different cell lines with 97% accuracy. It was notable that HCT-116 and HT-29 cell lines overlapped each other probably due to their similar metastatic potential. **Figure 3.1B** showed the same data classified according to type of cell line (healthy, non-metastatic or metastatic).



It was also notable that the same statistical parameters were responsible to show the most discrimination in both figures and that *discriminant 2* was in correlation with metastatic potential of the four cell lines. Interestingly, as pointed out in **Table 3.3**, the SLs that statistically contributed the most towards the guided grouping; were highly positive charged SLs (+3 to +5 charge), with fewer phenyl rings (0 or 2, from 2-PBA).

As observed in **Figure 3.2**, SL4 (having +3 charge) selectively bound LOVO (two-fold than the MRBB baseline signal), 4R (having +4 charge) bound well to cancerous cell

lines LOVO and HT-29, SL5-RRAc bound to HT-29 and SL5-Dab bound more to CONA and HCT-116. Together, these four SLs captured the greatest variation in binding, whereas the rest of the SLs (like SL1, SL2, SL3, 5R) their output resembled more closely to MRBB (baseline signal of internal array control).

Similar to the hypothesis made at the end of chapter 2; proteins in metastatic colon cells lines bound well to Arginine-rich SLs like SL4, 4R and SL5-RRAc. This could be due to proteins of colon cancer cells being greater negatively charged than in healthy colon cells. The newly added SLs contributed the most to the array's enhanced classification. Whether the LDA was 3-class (based upon cell line virulence) or 4-class (based on different cell line), the pattern looked similar (**Figure 3.1**). This highlighted that any which way the data was viewed; the statistical model was employing the same factors for separation.

3.2 INVESTIGATING COLON CANCER USING EXTENDED LECTIN ARRAY AND SECRETED PROTEINS FROM COLON CELLS

Aberrant glycosylation during oncogenesis and progression of cancer occurs equally in cell membrane and cell secreted proteins.¹⁰ During an attempt to investigate aberrant glycosylation with SL arrays, it is also useful to design a minimally invasive diagnostic tool. Instead of unnecessary tissue biopsies for extraction of cell membrane proteins, testing secreted proteins directly from blood serum deems more patient friendly. Therefore, we were interested to investigate the performance of extended lectin array on human colon cell secreted proteins. The same four cell lines were used and after subjecting them to 48-hour starvation, the media from the cells was collected and concentrated in spin tubes. The proteins were precipitated using acetone, fluorescently

labeled, quantified and incubated with the extended lectin array containing 20 SLs (old and new ones). This number was increased to include previously found and untested SLs (i.e. SL6, SL7, SL8, SL9, SL11 & SL11 mutants) plus SL5-R1,5,11A (as arginine control for SLs). All the SLs and their related information is as illustrated in **Table 3.4**.

Table 3.4: All 20 SLs in extended lectin array.
(Dab)*: diamino butanoic acid plus boronic acid; Ac: Acetylated N-terminus of peptide; Abu: aminobutyric acid.

SL Name	Amino acid sequence and modifications	#R	#2-PBA	#Ph
SL1	Ac-RG(Dab)*VTF(Dab)*RBBRM	3	2	3
SL2	Ac-RT(Dab)*RFL(Dab)*VBBRM	3	2	3
SL3	Ac-RS(Dab)*VTT(Dab)*RBBRM	3	2	2
SL4	Ac-RR(Dab)*TQT(Dab)*QBBRM	3	2	2
SL5	Ac-RA(Dab)*TRV(Dab)*VBBRM	3	2	2
SL6	Ac-RT(Dab)*NRN(Dab)*FBBRM	3	2	3
SL7	Ac-RS(Dab)*YFT(Dab)*QBBRM	2	2	4
SL8	Ac-RT(Dab)*YGN(Dab)*NBBRM	2	2	3
SL9	Ac-RT(Dab)*YQV(Dab)*ABBRM	2	2	3
SL11	R*L(Dab)*YLT(Dab)*RBBRM	3	3	4
SL11-R8A	R*L(Dab)*YLT(Dab)*ABBRM	2	3	4
SL11-T6A	R*L(Dab)*YLA(Dab)*RBBRM	3	3	4
SL11-Y4A	R*L(Dab)*ALT(Dab)*RBBRM	3	3	3
SL11-Y4F	R*L(Dab)*YLT(Dab)*RBBRM	3	3	4
SL11-T6Abu	R*L(Dab)*YL(Abu)(Dab)*RBBRM	3	3	4
SL5-R1,5,11A	Ac-AA(Dab)*TAV(Dab)*VBBAM	0	2	2
SL5-Dab	Ac-RA(Dab)TRV(Dab)VBBRM	3	0	0
SL5-RRAc	Ac-RRRA(Dab)*TRV(Dab)*VBBRM	5	2	2
4R	Ac-RAARAARA-BBRM	4	0	0
5R	Ac-RARARARA-BBRM	5	0	0

The unguided (PCA) analysis showed the pattern similarities between secreted and cell membrane protein extracts (**Figure 3.3**).

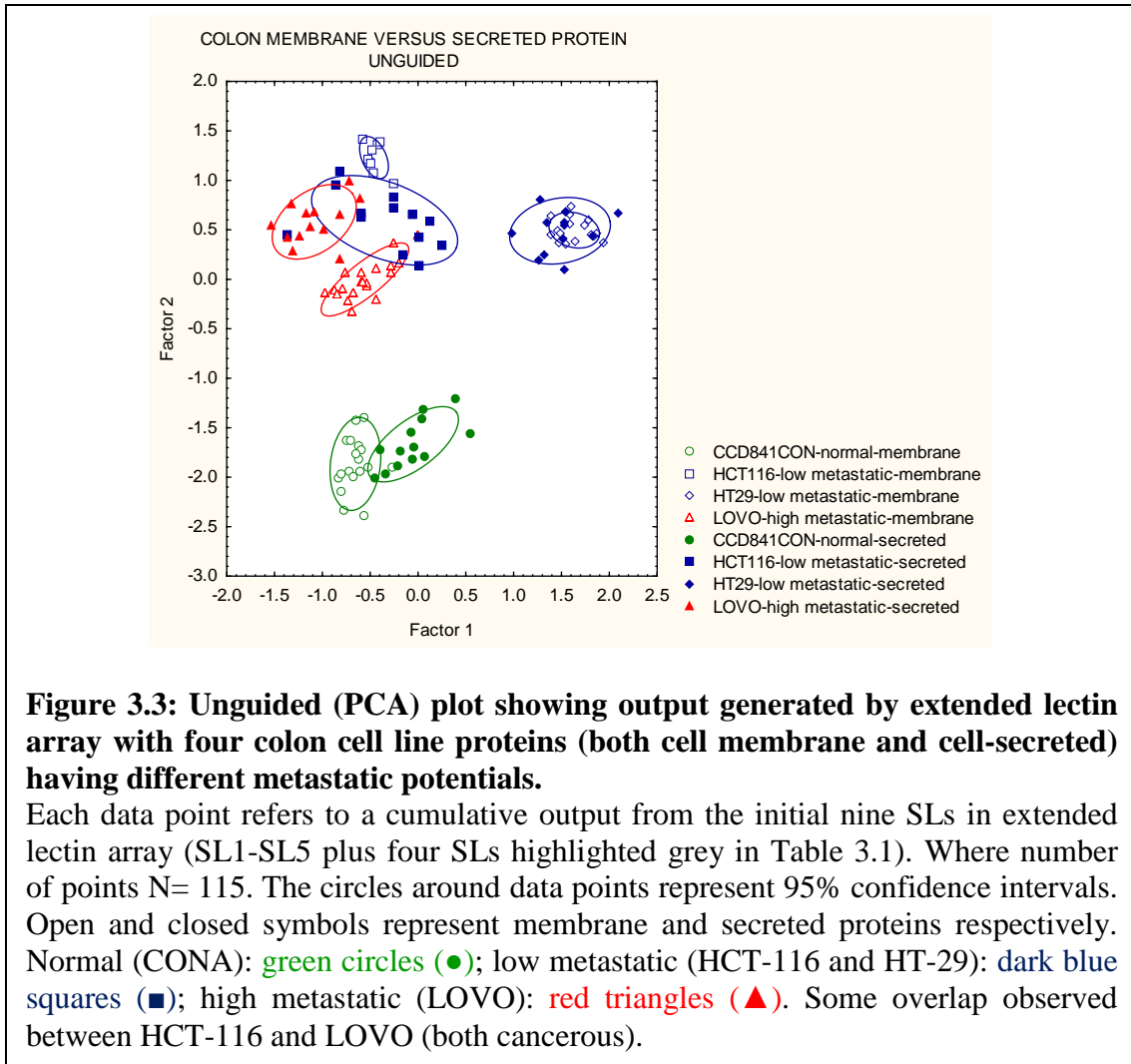


Figure 3.3 suggested that SL array might be targeting binding to the motifs that are present in cell membrane and secreted proteins both. Though the PCA of HCT-116 and LOVO in secreted and membrane proteins does not group very tightly, this might be because the cell lines that were used to extract membrane and secreted proteins belonged to different biological replicates. In the unguided PCA plot in **Figure 3.3**, the statistical parameters that were responsible to show second-most discrimination (y-axes) were in correlation with the metastatic potential of the four cell lines.

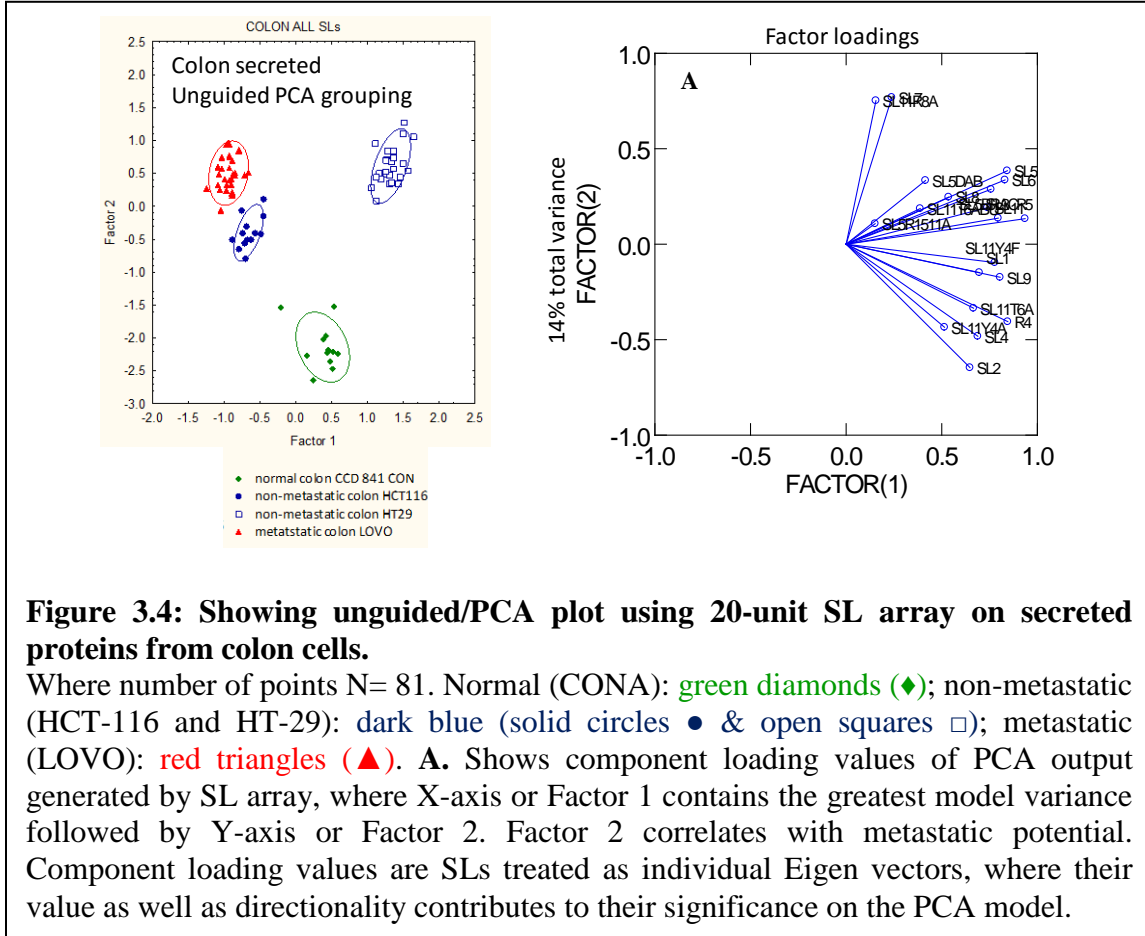
Name	# of positive charges from R	# Ph rings from 2-PBA	# Ph rings from peptide	F-score 3-class
4R	4	0	0	6.32
5R	5	0	0	12.93
SL4	3	2	0	9.35
SL5	3	2	0	6.58
SL7	2	2	2	7.60

Again, as shown in **Table 3.5**, the SLs that statistically contributed the most towards the unguided grouping of secreted colon proteins were highly positively charged SLs (in grey), with fewer phenyl rings (0 to 2).

This trend again verified the hypotheses made at the end of chapter 2; that malignant colon tissue tends to over-sialylate terminal glycans, thus making proteins in metastatic colon cells overall more negatively charged than in healthy colon cells. It could be the chemical rationale linking ability of SLs with greater positive charges (due to greater number of R) at discerning colon proteins from cell lines having different metastatic potentials, with most efficacy. Interestingly, SL7 showed significant binding despite only having +2 charge. One reason for this is the greater number of phenyl rings (four in number, **bold font in Table 3.5**) involved in structure of SL7. The statistical grouping model tries to pick out the SLs that provided the greatest variability between the different cell lines. It hypothesized that SL7 assisted the grouping by showing binding with the proteins through employing its extensive phenyl rings showing CH- π type interactions.

While looking at the F-score in **Table 3.5**, it is observed that the SLs that contributed the highest to 3-class discrimination (according to metastatic potential), seem to follow

two distinct trends. *SLs with higher number of arginines (R) and smaller number of phenyl rings and SLs with smaller number of arginine (R) and higher number of phenyl rings both seem important in classification.*



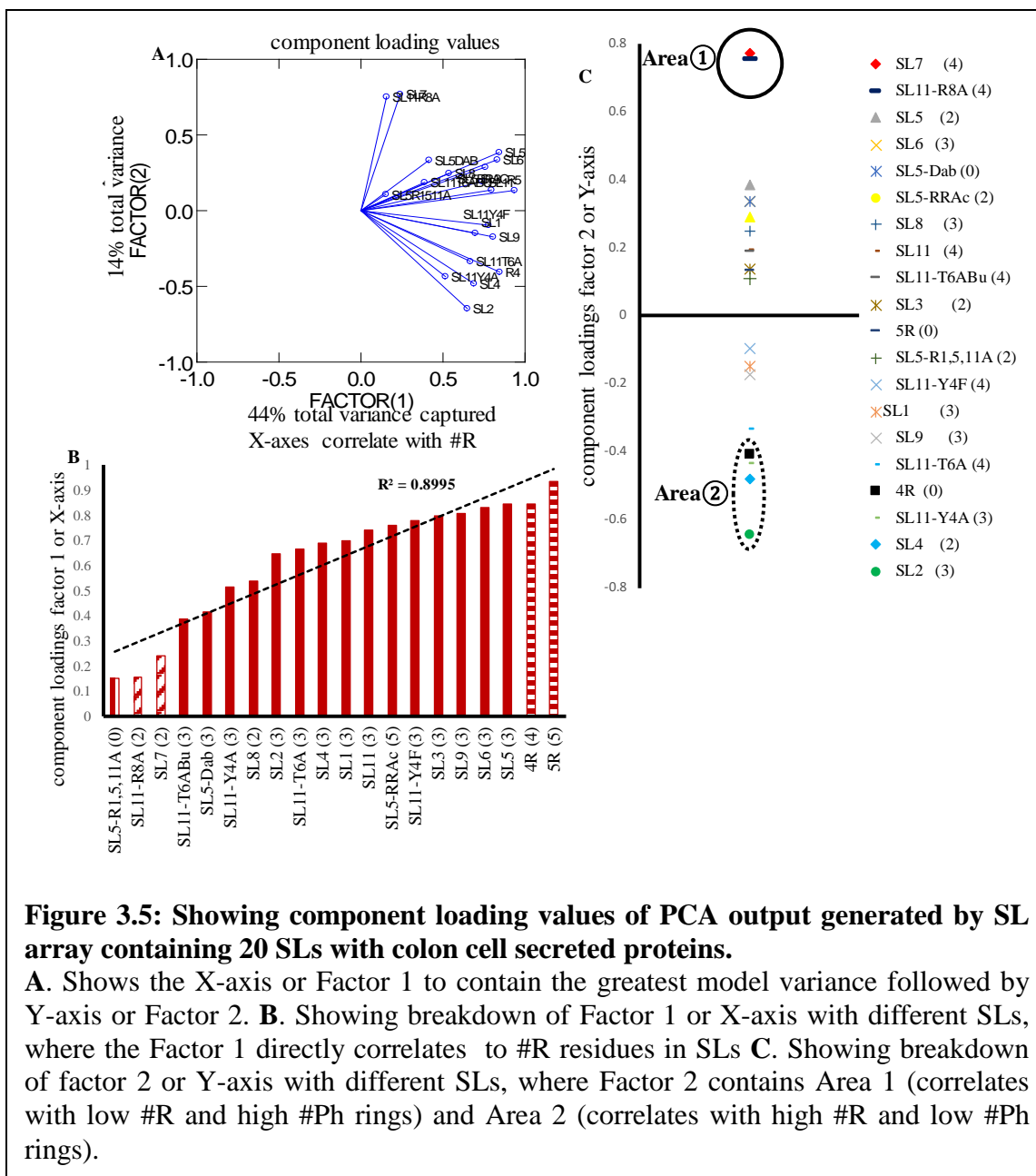
Before agreeing upon which SL is the most important for defining the model's classification, it is important to figure out a constant mathematical parameter for such comparisons. F-scores are from guided methods like LDA, where the program tries to fit the data to reduce the difference "within" and maximize the distance "in-between" the groups. In doing so, one risks data skewing therefore instead of looking at F-scores from LDA, we look at the component loading values from an unguided model like PCA.

Higher F-scores entailing a particular SL in the array contribute highly to LDA, but they have no significance as the grouping varies both with the number and the type of SL included in the model. For example, in **Table 3.5**, the number of SLs used were 20 and it is noticeable that the values of F-score are quite small with respect to **Table 3.3**, where array has 10 SLs. This is because F-score is not a percentage and is dependent on the number of SLs in array therefore varying every time the array size and/or constituency is changed. This serves as a reason to look at component loading values after using an unguided method like PCA.

Unlike F-scores, component loading values are derived from the variance of the model, which remains constant and do not depend upon the number of SLs involved in the statistical model. **Figure 3.4** shows PCA output generated by SL array containing 20 SLs with colon cell secreted proteins, where the Factor 2 (Y-axis) represents the metastatic potential. The component loading values (plotted in **Figure 3.4A**) of each SL are actually Eigenvectors projected on a plane therefore having X and Y coordinates. This means that not only does the magnitude of the Eigenvalue matters but the direction of the vector matters too. The direction of the SLs in **Figure 3.4A** is reflected by the PCA output (given on the left in **Figure 3.4**). The more diverse the relative directionality between the component loading values of the two SLs, the greater is the significance of the two SLs on the variance of model. **Figure 3.4A** shows the two greatest factors across X-axis (Factor 1) and Y-axis (Factor 2) that try to capture the total variance of PCA model. The variance of the model is contributed the most by the SLs that spread apart the furthest.

As seen in **Figure 3.4A**, Factor 2 seems to correlate with metastatic potential of colon cell lines. In **Figure 3.5**, **Figure 3.5B**. shows the breakdown of Factor 1 across X-axis with different SLs. Interestingly SL5-R1,5,11A, an SL with no R residues, shows the smallest value on Factor 1, closely followed by SL11-R8A and SL7 (both having the next lowest, only two R residues). Whereas, 4R and 5R SLs (having four and five R residues respectively) have the greatest projection on Factor 1. This states that the most important cause of model variance to discern colon cell lines is the number of R residues in the SL array. The positive charge on the guanidinium groups are involved in salt bridging with negatively charged proteins/ glycoproteins.

Similarly, **Figure 3.5C**. depicts the breakdown of Factor 2 across Y-axis with different SLs. Factor 2 shows two prominent areas, Area 1 (in bold) and Area 2 (in dotted line) where the most contributing SLs exist. Area 1 correlates with SLs having low number of arginine residues and high number of phenyl rings (SL11-R8A and SL7) and Area 2 correlates with high number of arginine residues and low number of phenyl rings (SL2, SL4, 4R). So both areas are important for discrimination and it is really interesting that basically having less of an interaction matters to the discrimination too. SLs like 5R and SL5-R1,5,11A got placed with the rest of the SLs, which corresponds without hypothesis because SL5-R1,5,11A has no arginine residue for protein binding and 5R has too many R's which make it an indiscriminating binder lacking selectivity.



As observed from **Table 3.5** the higher F-score of SL7 led to a question whether the positive charges of SLs were the only chemical guidance to array's classification ability or was there a significant role played by phenyl rings in the SLs as well?

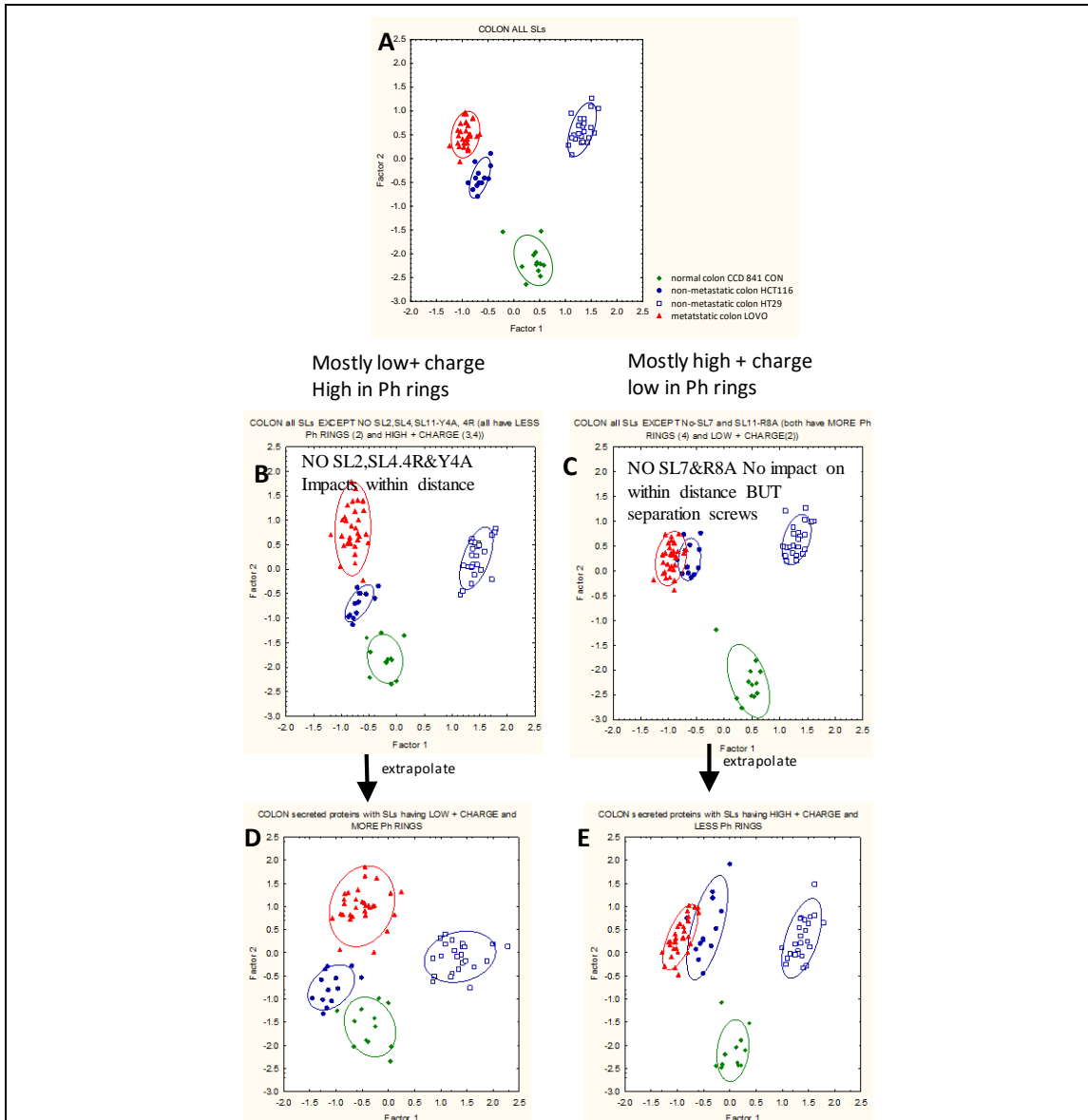


Figure 3.6: Showing A: unguided/PCA plot of 20-unit SL array with secreted proteins from colon cells being impacted by B: Removal of Area ② SLs from PCA model; C: Removal of Area ① SLs from PCA model.

The circles around data points represent 95% confidence intervals. Where number of points N= 81. Normal (CONA): green diamonds (◆); non-metastatic (HCT-116 and HT-29): dark blue (solid circles ● & open squares □); metastatic (LOVO): red triangles (▲). PCA model **D**: containing all SLs with more # of phenyl rings and lesser positive charge(0 to +2), causing loss of precision of grouping. Model **E**: containing all SLs with greater positive charges (+4-+5) and less # of phenyl rings, causing a decrease in accuracy of grouping especially between cancerous cell lines.

For investigation, the secreted colon data was re-modeled in several different ways. In

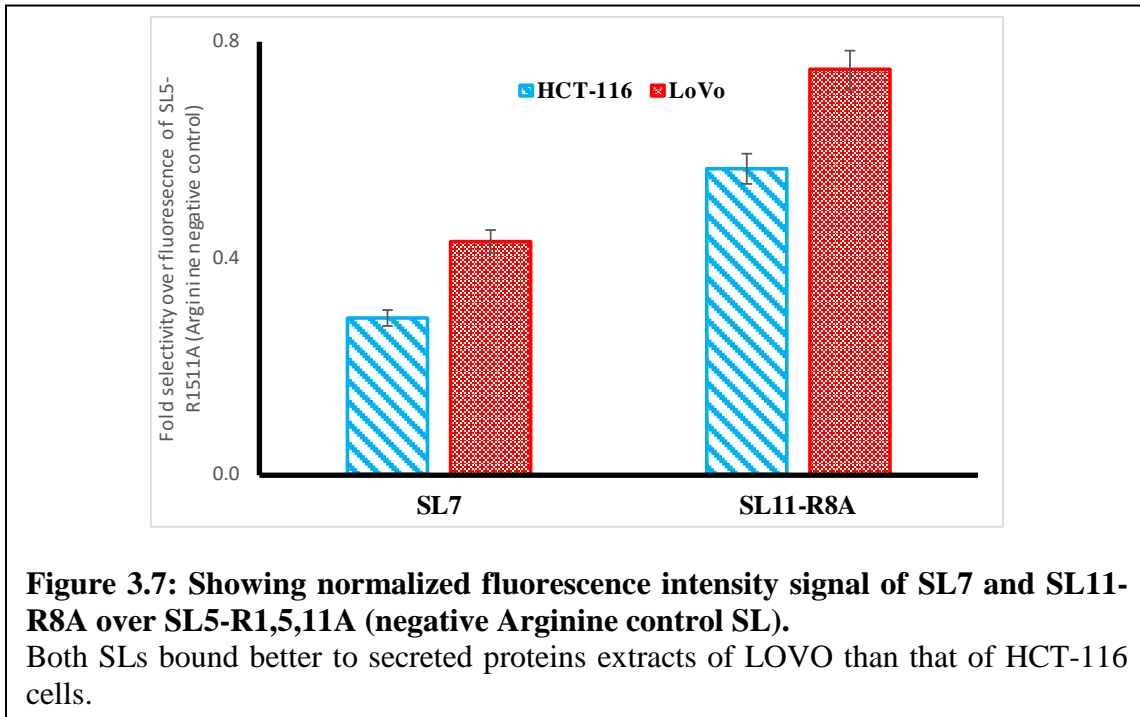
Figure 3.6A, the original PCA output with colon secreted proteins and 20 SLs were broken down into two different grouping experiments (shown in B and C).

In **Figure 3.6B** the PCA model excluded all SLs in Area ②, leading to loss of precision (see widening of 95% confidence interval circle of LOVO) and causing slight loss of discrimination between healthy and cancerous cell lines (see green and blue circles coming closer).

Later, a PCA model was studied by including SLs with greater number of phenyl rings (4) and lower positive charge (0 to +3) and excluding all other SLs. The result of this extrapolation is seen in **Figure 3.6D**. Excluding SLs with high positive charges (+4 and higher), caused the model to lose overall precision between each cell line, making grouping more randomized suggesting the charged SLs overall, played a significant role interacting with proteins from different colon cells. One possible reason is highly positive charged SLs are strong acids and can bind to proteins (bases) of any strength, thus increasing the overall binding capability of the array but at the same time, the model tends to lose its selectivity. Another potential reason can be correlated to proteins in colon cells undergo changes in rate of sialylation during and after the onset of cancer.

In **Figure 3.6C** the PCA model excluded all SLs in Area ①, leading to loss of accuracy (see 95% confidence interval circles of HCT-116 and LOVO merging) but the model still maintained tight confidence ranges overall. From bar graph data, (**Figure 3.7**) it was seen that SL7 and SL11-R8A, bound better to LOVO over HCT-116 cells, hence their loss in model C led to the loss in accuracy between these two cell lines. Tight confidence intervals, as explained earlier are due to the presence of highly positively charged SLs.

Figure 3.6E, likewise is the extrapolated version of C, where the unguided statistical model was made by including SLs with greater positive charge (+3 to +5) and lower number of phenyl rings (0-2). Consequently, SLs that contained greater number of phenyl rings (3 or 4) and lesser positive charges (0 to +2), were excluded. As observed, modeling with highly charged SLs led to tighter clustering between different cell lines, making the grouping more precise than the original array output (compare with **Figure 3.6A**). An overlap between non-metastatic (HCT-116) and metastatic (LOVO) was observed (in **Figure 3.6E**), indicating that loss of SLs with greater number of phenyl rings and less positive charges (SL7) caused a loss in accuracy of the model, making phenyl rings in the SLs useful for “staging” colon cancer. SL7 and SL11-R8A were included in **B** and **D** models and they helped in discriminating between HCT-116 and LOVO.

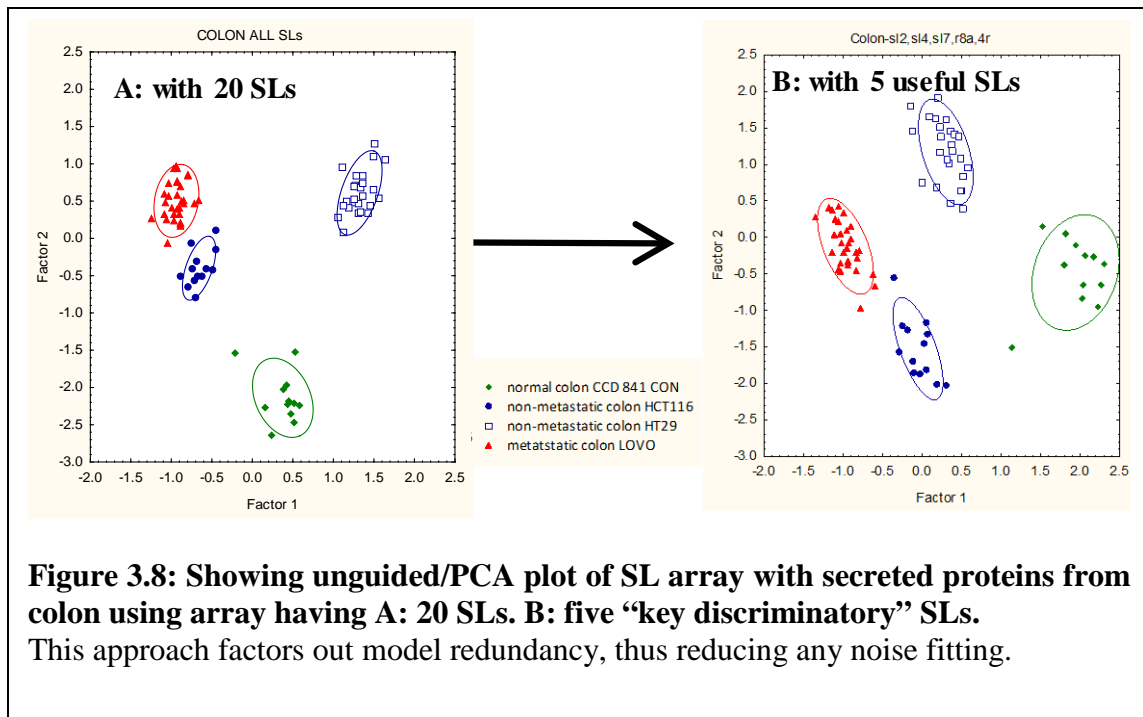


From this section, it was surmised that SLs with higher number of arginines and smaller number of phenyl rings contribute to classification of colon cell lines by

enhancing the ‘tightness’ of PCA clustering and on the other hand, SLs with smaller number of arginines and higher number of phenyl rings seem important in ‘staging’ of colon cancer. PCA method of analysis seems impartial at discerning various cell lines rather than LDA. Regardless of LDA or PCA, both methods seem to use metastatic potential of cell lines as one of the statistical parameters to define colon cancer.

3.3 IMPACT OF MANY LECTINS IN SL ARRAY ON TOTAL VARIANCE OF PCA MODEL

As shown in **Table 3.4**, to investigate the metastatic potential of secreted colon proteins, an array of 20 SLs was used. The resulting PCA output is shown in **Figure 3.4**



and **3.6A**. After investigation of the impact of each SL on the PCA grouping, it was found that only five SLs contributed significantly to discern colon cancer proteins—SL2, SL4, 4R, SL7 and SL11-R8A. A PCA containing only these five SLs was modeled.

(**Figure 3.8**) **Figure 3.8B** shows PCA model with these five SLs, wherein the X-

axis/Factor 1 is the factor that captures the greatest percentage out of the total variance of the scatterplot. Interestingly, using only the few key SLs caused the metastatic potential axis to align with Factor 1. This may suggest, *using 20 different SLs may be introducing redundancy to the model, therefore fitting in the noise and skewing the actual data* (law of diminishing returns). For example, in **Table 3.4**, the F-score of 5R was one of the highest out of the 20 SLs, but the high F-score value had no significance to the correlate PCA grouping to metastatic potential. As observed in **Figure 3.5A**, wherein Factor 2 is shown correlating with metastatic potential, in **Figure 3.5C**, the Eigen vector 5R is neither the part of Area ① nor ②. This is due to 5R binding indiscriminately to all proteins and enhancing noise. Therefore, getting rid of such SLs from the PCA model is useful to classification.

3.4 INVESTIGATING ROLE OF 2-PHENYL BORONIC ACIDS IN CLASSIFYING COLON CELLS ACCORDING TO THEIR METASTATIC POTENTIAL

Previously, while classifying proteins from colon cell membrane, SL5-Dab played a significant role even while it contains no phenyl boronic acids (2-PBA). Besides this, during the classification of colon cell secreted proteins, there were no direct correlations that could be drawn between changes in ability of SL array at classification of cells and loss of boronic acid from the SL.

These observations led to the current investigation as to whether the boronic acid moiety in 2-formyl phenyl boronic acid plays any direct role at classifying colon cells according to their metastatic potential? In order to test this, a specialized SL array was made that included the initial five SLs (SL1 to SL5) and the same five SLs without

adding 2-PBA onto them (namely SL1-Dab to SL5-Dab). The sequence information has been previously shown in **Table 3.1** and **Table 3.4**

Colon cell membrane proteins were extracted, fluorescently tagged and incubated with this SL array. In (**Figure 3.9**) **Figure 3.9A**, shows the unguided LDA plot of normal, low and high metastatic colon cell lines output with SLs not containing 2-PBA, representing clear discrimination between normal and cancerous cell lines. The 3-class guided grouping of this model discriminated the cell lines with 91% accuracy.

Whereas, in **Figure 3.9B**, the PCA of cell lines using SLs that contained 2-PBA was not as discriminating as in **Figure 3.9A** showing greater misclassifications between normal and cancerous cell lines. The SLs without 2-PBA were all highly positively charged SLs (+5) and employing that charge, they were able to pull apart normal from cancerous cells with greater efficacy. The 3-class guided grouping of colon cell line output using SLs containing 2-PBA was 89% (very similar to 91% of the other model). This was indicative that *2-PBA was not directly contributing significantly to the classification of cell lines. This may suggest that glycoproteins or proteins in the cell lines might not be interacting with SLs via boronic acid moieties in phenyl boronic acids.* Further, 2-class guided grouping between non-metastatic and metastatic cell lines were calculated using the same data that were used in **Figure 3.9A** and **B**. Both classified with 98% accuracy, denoting that boronic acids might not be useful in ‘staging’ of colon cancer. Using reporter dyes like Alizarin red (ARS), which form cyclic esters with boronic acid, a change in the color or fluorescence of the dyes is observed. For example, Alizarin red, containing catechol in its unbound form has no fluorescence, but formation

of cyclic esters with boronic acid moieties of SLs, showed change in emission wavelength to 580 nm (red region)

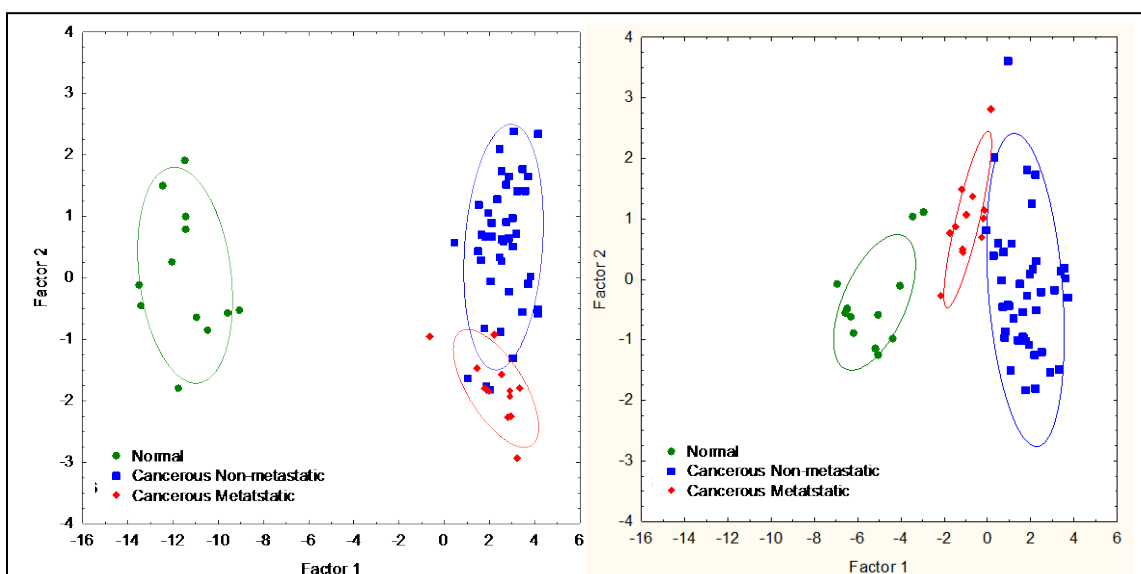


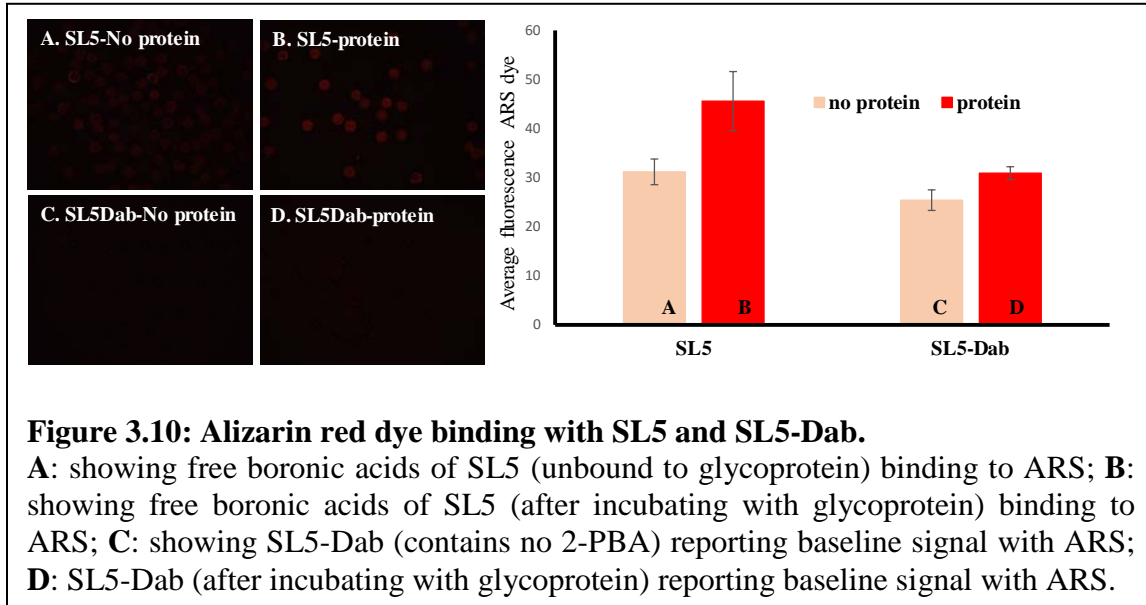
Figure 3.9: LDA plots showing direct impact of boronic acid moiety in SLs included in array at classifying different protein samples from colon cells.

Unguided/PCA plot with SLs **A**: without 2-PBA; **B**: with 2-PBA. Normal (CCD 841 CON): green closed circles (●); (non)/low metastatic (HCT-116 & HT-29): blue closed squares (■); high metastatic (LOVO): red closed diamonds (◆). Circles indicate 95% confidence intervals. SLs without 2-PBA, representing clearer discrimination between normal and cancerous cells. Where number of points N= 66.

That is because protons on catechol moiety in ARS are lost while forming the boronate ester and these protons are the ones that normally quench ARS fluorescence.^{4,5}

To investigate whether glycoproteins were binding to boronic acid moieties in 2-PBA via involvement of cyclic ester formation between boronic acid and cis-diols of glycans in glycoproteins, Alizarin red dye was used as a signaling/reporter molecule (**Figure 3.10**). The premise was if glycans formed cyclic ester to boronic acid of SL, then it would be accompanied by loss in fluorescence intensity signal of Alizarin red (since boronic acid sites would be involved in binding to glycoproteins). SL5 (containing two 2-PBA units) and SL5-Dab (devoid of 2-PBA), were labeled with 100 μ M Alizarin red solution (ARS) in 0.1M PBS and imaged (images shown in **Figure 3.10A** and C).

SL5 and SL5-Dab samples are incubated with unlabeled cell membrane proteins for 12-hours and then stained with ARS; which was followed by imaging (**Figure 3.10B** and **D**). All the images were taken at the same exposure times and microscopic conditions.



When signals of A and C were compared, SL5 bound more to ARS than SL5-Dab; this was attributed to presence of two 2-PBA in SL5 which bound ARS dye by forming cyclic esters with the dye. SL5-Dab on the other hand showed no fluorescence with ARS because it contained no 2-PBA units.

After both the SLs were incubated with proteins/glycoproteins from cancerous cell membrane extracts, and later stained with ARS to determine any free boronic acid sites; SL5 registered an increase in fluorescence. This indicated that boronic acid sites on SL5 were free and that the glycans did not bind to them. It is possible that ARS dye being negatively charged (basic), bound to the proteins and SL5 being +3 charged (possibly through positively charged amino acid residues), ARS may be targeting the positive charge, just as it targets Ca^{2+} deposits in synovial fluids.⁶ This may also explain why SL5-Dab with protein reported slightly greater fluorescence intensity with ARS when

compared to the SL5-Dab sample without proteins. These findings indicate that the boronic acid functional group in 2-formyl phenyl boronic acid might not play a significant role in binding to glycans of glycoproteins.

3.5 INVESTIGATING PROSTATE CANCER USING EXTENDED LECTIN ARRAY AND SECRETED PROTEINS FROM PROSTATE CELLS

In a transformed cell, glycosylation changes most often arise from changes in the expression levels of glycosyltransferases and lead to modifications in the core structure of *N*-linked and *O*-linked glycans.¹⁷ One of the most common changes is an increase in branching of *N*-linked glycans, creating additional sites for terminal sialic acid residues. This, along with an upregulation of sialyltransferases and fucosyltransferases, ultimately leads to an increase in global sialylation and fucosylation in a diseased cell. This results in abundance of certain glycan structures: termed as cancer-associated glycans (CAGs). The following **Table 3.6** shows the overexpressed CAGs in malignant human tissues from breast, colon and prostate (chapter 1 Ref).²³ Breast tissue involves greater number of CAGs reporting over-sialylation, followed by colon tissue and then prostate. Whereas, all three tissue types report an increase in core-fucosylation.⁷⁻¹²

Exclusive presence of poly sialic acid in malignant breast tissue makes it most negatively charged while malignant prostate is the least negatively charged tissue out of the three, according to glycan composition. Hence, prostate cell line model is employed to examine the impact of positively charged SLs and SLs with greater number of phenyl rings in the array upon their ability to comprehend tumorigenicity.

Table 3.6: Cancer associated glycans in malignant tissues. *PSA stands for poly sialic acid.				
Cancer glycan CAG	Glycans present in CAG	Malignant tissue		
		Breast	Colon	Prostate
sLex	sialic acid, fucose, glcNAc, gal	√	√	
sLea	sialic acid, fucose, glcNAc, gal	√	√	
sTn	sialic acid, galNAc	√	√	√
TF	gal, galNAc	√	√	√
Ley	fucose, glcNAc, gal	√	√	√
Globo H	fucose, gal, galNAc, glc	√	√	√
*PSA	sialic acid	√		
GM2	sialic acid, gal, galNAc, glc	√	√	√

Five prostate cell lines were used and after subjecting them to 48-hour starvation the media from the cells was collected and concentrated. The proteins were precipitated using acetone, fluorescently tagged, quantified and then incubated with the extended lectin array containing the same array of 20-SLs used in the colon cell study. All the SLs and their related information is shown in **Table 3.4**.

Table 3.7: Human prostate cell lines used with extended SL array. Arranged according to increasing order of metastatic potential.			
Cell line	Type	Percent Invasiveness	Colony forming efficiency %

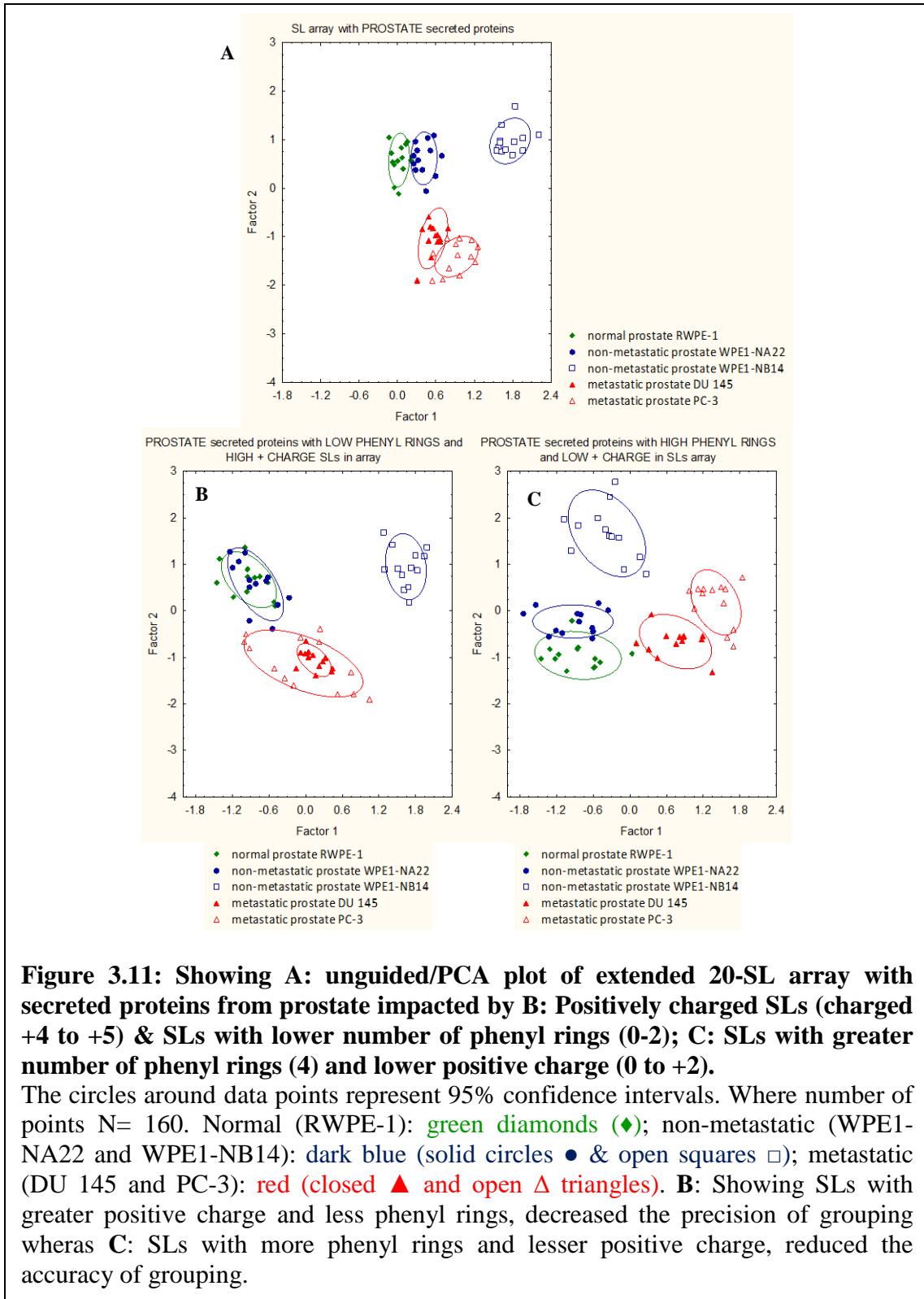
RWPE-1	normal	-	-
WPE1-NA22	low metastatic	9 ¹³	0.06 ¹³
WPE1-NB14	low metastatic	30 ¹³	1.85 ¹³
DU 145	high metastatic (grade II)	23 ¹¹	36 ¹⁶
PC-3	high metastatic (grade IV)	70 ¹⁵	52 ¹⁶

Table 3.7 illustrates five prostate cell lines arranged in an increasing order of metastatic potential. RWPE-1 cell line was taken as normal epithelial control. WPE1-NA22 and WPE1-NB14 cells belong to a family of cell lines, referred to as the MNU cell lines, which all derived from RWPE-1 cells after exposure to N-methyl-N-nitrosourea (MNU). MNU is a direct-acting carcinogen, where WPE1-NA22 were cultured from first-generation tumor after injecting the mice with cells treated with 50 μ g/ml MNU and WPE1-NB14 cells were from second-generation tumor obtained after injecting the mice with cells treated with 100 μ g/ml MNU. These three MNU cell lines are +AR and AR-mediated +PSA expression, indicating prostate epithelial origin.¹³ AR stands for androgen receptor. The larger family of cell lines, including RWPE-1 cells with a common lineage, mimicking the multiple steps in progression from normal epithelium to prostatic intra-epithelial neoplasia, and then to invasive cancer. The metastatic potential of MNU cell lines is defined by their colony forming efficiency percent (CFE %) in soft agar. WPE1-NA22 and WPE1-NB14 had low and moderate CFE, respectively whereas RWPE-1 had none. DU 145 and PC-3 cell lines are from moderately and highly aggressive adenocarcinomas, respectively. The SLs that statistically contributed the most towards the unguided grouping of secreted prostate proteins are illustrated in **Table 3.8**. These were highly positively charged SLs (highlighted in grey), with fewer phenyl rings (0 or 2, from 2-PBA). This trend could be correlated since malignant prostate tissue tends

to over-sialylate terminal glycans, thus making proteins in metastatic prostate cells overall more negatively charged than in healthy prostate cells. Interestingly, SL11-Y4A and SL7, showed significant binding even when it contained fewer positive charged amino acid residues in them. This could be attributed to the greater number of phenyl rings (3 and 4 respectively, **bold** in **Table 3.8**). The statistical grouping model tries to pick out the SLs which provided the greatest variability between the different cell lines. It could be hypothesized that SL11-Y4A and SL7 assisted the grouping by showing binding with the proteins through employing phenyl rings showing CH- π type interactions.

Table 3.8: Showing most significant SLs at classifying proteins secreted from prostate cells according to their metastatic potential.
(SLs highlighted in grey indicate highly positive-charged and in bold indicate SLs with higher number of Phenyl rings).

Name	# of + charges from R	# Ph rings from PBA	# Ph ring from peptides
SL11-Y4A	3	3	0
SL7	2	2	2
SL4	3	2	0
SL5	3	2	0
4R	4	0	0



To further investigate the role played by a greater number of phenyl rings in some SLs

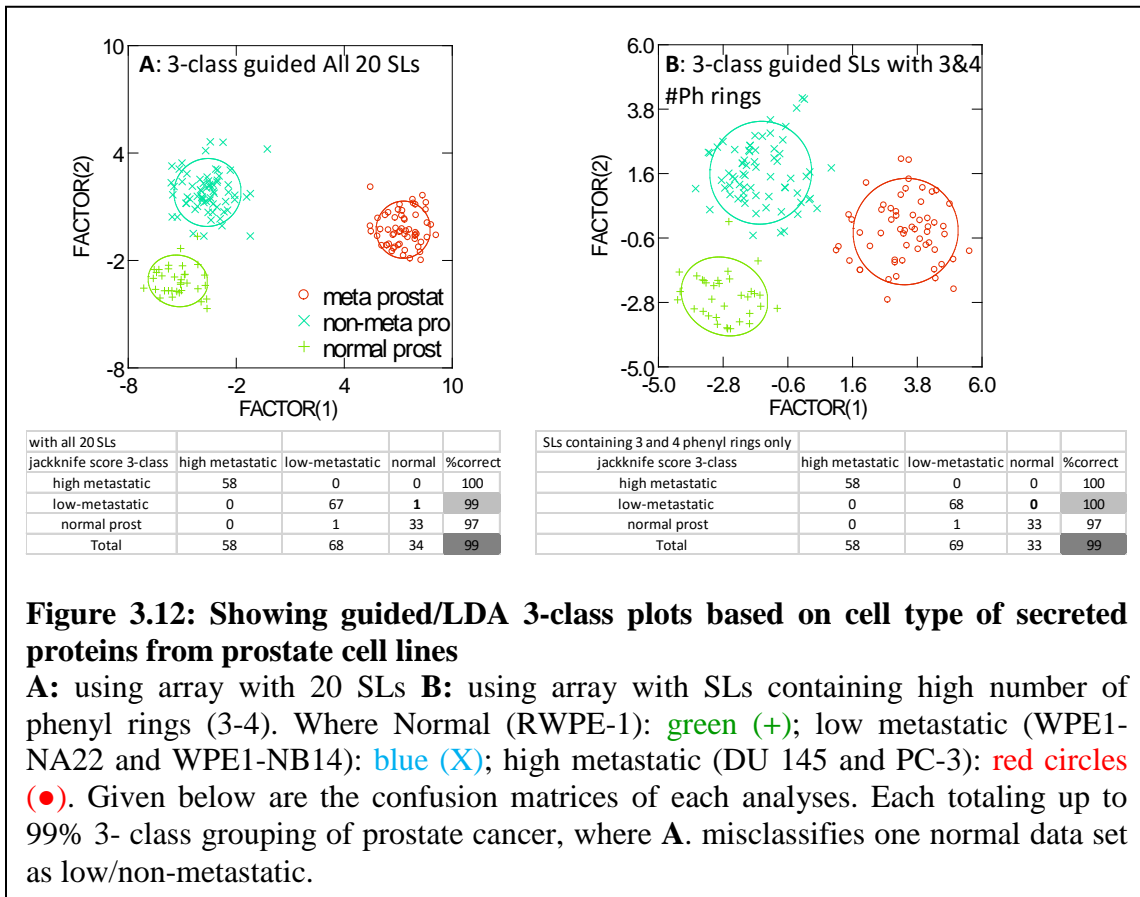
(e.g. SL11-Y4A and SL7) on the SL array's classification ability; the prostate cell line PCA data was re-modeled in several different ways (**Figure 3.11**).

In **Figure 3.11B** the unguided statistical model was made by including SLs with greater positive charge (+4 to +5) and lower number of phenyl rings (0-2). Consequently, SLs that contained greater number of phenyl rings (3 or 4) and lesser positive charges were excluded, (e.g. SL11-Y4A and SL7). As observed, modeling with high positively charged SLs led to a loss in the discrimination between proteins from prostatic normal and cancerous cell lines, hence losing more precision than the original array output (see A). An overlap between non-metastatic (WPE1-NA22) and normal (RWPE-1) was observed indicating that loss of SLs with greater number of phenyl rings and less positive charges (SL11-Y4A and SL7) caused a randomization of the model, making phenyl rings in the SLs useful for grouping.

In **Figure 3.11C**, the unguided statistical model was made by including SLs with greater number of phenyl rings (4) and lower positive charge (0 to +3) and removal of SLs that were highly positively charged (+4 to +5) and low on phenyl rings (0 to 2). SL11-Y4A and SL7 were included in this model (SL4, SL5 and 4R were excluded) and they helped in maintaining discrimination between each cell line.

Thus SLs with greater number of phenyl rings did significantly contribute in prostate cell line classification; although, losing SLs with high positive charges, did cause the data to spread out more within each cell line, which meant that positively charged SLs overall, also played a significant role interacting with proteins from different prostate cells. That could be correlated to proteins in prostate cells undergoing changes in sialylation during and after the onset of cancer. This sialylation expression in prostate cells was less than

that observed in colon cells hence core-fucosylation became a relevant marker towards discriminating prostate cell lines. It was then *hypothesized that probably SLs with greater phenyl rings might be targeting core-fucosylation changes via CH- π type interactions in prostate cells*. Meanwhile, looking at the 3-class LDA of serum proteins based on prostate cell types used according to the metastatic potential, **Figure 3.12A** shows a LDA of an overall 99% using 20 SLs. The same dataset when modeled using only SLs containing higher number of phenyl rings in the array also gave an overall 3-class 99% classification (see **Figure 3.12B**).



In fact, model B grouped low metastatic cell lines better than model A (where one of the low metastatic data point misclassified under normal cell line group). *This elucidates the importance of phenyl ring containing SLs to discern prostate cell lines.*

3.6 CONCLUSIONS

Malignant colon tissue tends to over-sialylate terminal glycans, thus making proteins in metastatic colon cells overall more negatively charged than in healthy colon cells. It could be the reason due to which SLs with greater positive charges (from R) were able to discern colon proteins from cell lines having different metastatic potentials with greatest efficacy.

Extended array containing SLs with higher number of arginines and smaller number of phenyl rings contribute to classification of colon cell lines by enhancing the ‘tightness’ of PCA clustering. Due to the great amount of positive charge, these SLs behave like strong acids that indiscriminately bound to proteins with different pIs. On the other hand, SLs with smaller number of arginines and higher number of phenyl rings seem important in discriminating HCT-116 and LoVo cells. Top 6 SLs that best describes colon cancer are SL2, SL4, 4R, SL11-Y4A, SL7 and SL11-R8A

PCA method of analysis seems impartial at discerning various cell lines rather than LDA. Regardless of LDA or PCA, both methods seem to use metastatic potential of cell lines as one of the statistical parameters to define colon cancer. 2-PBA was not seen to be directly contributing significantly to the classification of colon cell lines. This may suggest that glycoproteins or proteins in the cell lines might not be interacting with SLs via boronic acid moieties in phenyl boronic acids.

In prostate cancer, SLs with greater phenyl rings might be targeting core-fucosylation changes via CH- π type interactions. When prostatic cell lines were investigated by using SLs containing higher number of phenyl rings only (99%), the array classified better than the model containing SLs with highest F-scores (91%, all 3 SLs high in R, namely 4R,

SL4 and SL5). Top 5 SLs that best describes prostate cancer are SL7, SL11-Y4A, SL4, SL5 and 4R

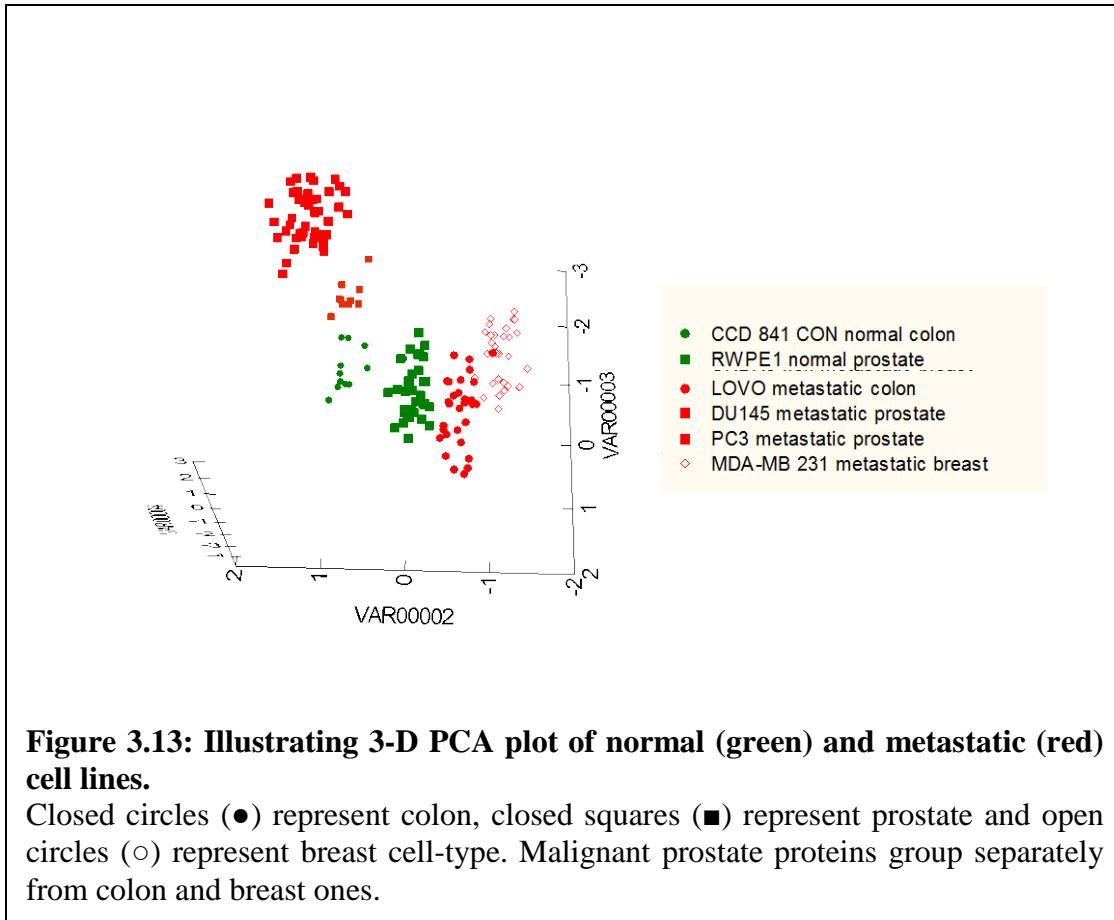
Altered glycosylation is a hallmark of cancer that has helped to shape the management and understanding of cancer.¹⁷ Currently, several glycan-based biomarkers are in use worldwide¹⁸ and glycans have been established as key participants in tumorigenesis and cancer progression.¹⁹ Knowledge of altered O-glycan structures in cancer has led to the development of O-glycan-based biomarkers.²⁰⁻²³ Antibodies targeting glycan or glycoprotein biomarkers include: CA15-3, CA125, CA19-9, and CA72.3, against MUC1, MUC16, sLe^a and sTn respectively; which are up-regulated in breast, ovarian and pancreatic cancers.²⁴⁻²⁶ However, glycans like sTn are expressed in most carcinomas (e.g., stomach, colon, pancreas, biliary tract, ovaries, and cervix), limiting its utility in identifying tissue of origin.²⁷⁻²⁹ Hence all these biomarkers are agreeable prognostic methods as opposed to being used in diagnosis because they lack tissue specificity. In general, the focus has moved from hunting for specific biomarkers towards observation of a change in global sialylation or fucosylation.³⁰ Alterations in protein glycosylation associated with colon cancer have been reported; which among several changes, included increased sialylation, and core-fucosylation.³⁰ Similarly, there is literature precedence supporting an observed increase in sTn and core-fucosylation in malignant prostate tissues.³¹ Several groups have change in different sialyl Lewis antigens in metastatic prostate tissues and found them to be very sporadically expressed.³²

Table 3.6, shows metastatic colon tissue being more negatively charged (due to greater number of sialylated CAGs), followed by metastatic prostate tissue (expressing

sialylation mainly through increased sTn expression). Meanwhile, the core-fucosylation of N-glycans decreased in colon cancer and increased in prostate malignancy; however, due to reduced sialylation in metastatic prostate cancer, the fucosylation becomes important in monitoring prostate cancer.³²⁻³³ In addition to the colon and prostate cell lines, data from a metastatic breast cell line (MDA-MB-231) was also included in this analysis, primarily for the purpose of increasing the number of high metastatic reference cell lines. (Courtesy: Daniel Gordon)

The extended SL array outputs (as discussed in the preceding sections) for colon and prostate cell lines were analyzed together using PCA (unguided) to observe any trends resulting from tissue typing. **Figure 3.13** depicts the results from the 3-D unguided PCA analysis of data from the extended SL array. Notice that normal colon and prostate cells cluster close to each other, whereas highly metastatic prostate and highly metastatic colon cell lines cluster on the two extreme ends. Interestingly, the highly malignant breast cell line, MDA-MB-231, clustered close to the highly metastatic colon cell line, LOVO, possibly due to high sialylation in these two metastatic cell lines.^{3, 34}

As discussed in the preceding sections, the highly positively charged SLs, such as 4R, SL4 and SL5, as well as SLs containing a greater number of phenyl rings, for example SL11-Y4A and SL7, were the most significant contributors in classifying colon and prostate tissues based on cell line tumorigenicity (see APPENDIX C). Positively charged SLs contributed to the tight clustering of PCA plot, they may be targeting the difference in sialylation between colon and prostate tissues. More on the quantification of sialic acid residues as well as correlation of sialic acid amount and metastatic potential of cancer is discussed in the next chapter.



3.7 FUTURE DIRECTIONS

It may be of interest to investigate how the 20-unit extended lectin array behaves against a modified cell line with minimal sialic acid expression. Another set of experimentation is proposed with a smaller array including select SLs and some cell lines like the Chinese Hamster Ovary (CHO) cell line. CHO cells, due to their *reduced* expression of sialic acids (they have no glycoproteins with sialic acids linked via α -2,6 linkage and has only α -2,3 linked ones)³⁵ can serve as baseline control for sialic acid expression. In retrospect, CHO cells might not be an ideal candidate for negative sialic acid expression control, since it still has sialic acid linked to glycoproteins via α -2,3 linkage. In order to make a better negative sialic acid control cell line, a plasmid that halt

the complete sialic acid expression can be used to transfect the CHO cell line. Granted that CHO cells are non-human in origin nevertheless are known to be easily transfected.

3.8 EXPERIMENTAL PROCEDURES

Details of the complete procedure for SL synthesis, incubation with fluorescently tagged analytes, imaging, data acquisition, data analysis and statistical analysis and the programs used are provided in the Experimental section in Chapter 2.

General Cell Culture Protocol

All cell lines were purchased from ATCC[®]. A total of four human colon cell lines and five human prostate cell lines were grown following the ATCC[®] cell culture protocols.³⁶⁻⁴⁵ The names of the colon cell lines followed by prostate cell lines are listed below in accordance with their increasing metastatic potential. Shellab 3502 CO₂ water-jacketed incubator (by Marshall Scientific) was used to grow all cell lines in 37 °C and under an atmosphere of 5% CO₂.

Colon: CCD 841 CON (ATCC[®] CRL-1790[™]) {normal}, HCT-116 (ATCC[®] CCL-247[™]) {lowly metastatic, grade II}, HT-29 (ATCC[®] HTB-38[™]) {lowly metastatic, grade II} and LoVo (ATCC[®] CCL-229[™]) {highly metastatic, grade IV}. HCT-116 and HT-29 were isolated from the colon whereas LoVo is metastatic colorectal cancer isolated from a site located in left supraclavicular region.

Prostate: RWPE-1 (ATCC[®] CRL-11609[™]) {normal}, WPE1-NA22 (ATCC[®] CRL-2849[™]) {lowly metastatic}, WPE1-NB14 (ATCC[®] CRL-2850[™]) {lowly metastatic}. Both these cell lines are isogenic with RWPE-1 but MNU exposed to induce tumorigenicity. DU 145 (ATCC[®] HTB-81[™]) {highly metastatic}, PC-3 (ATCC[®] CRL-

1435TM) {highly metastatic}. DU 145 and PC-3 were metastatic prostate cancer isolated from sites in brain and bone, respectively.

Media Used in Cell Culture

All cell lines grown required 10% v/v Fetal Bovine Serum (FBS) and 1% v/v antibiotic/anti-mycotic along with their specific medium (FBS and antibiotics added into 500ml cell growth media bottles). Antibiotic/ anti-mycotic used typically contains a solution of 100IU/ml penicillin, 50-100µg/ml streptomycin and amphotericin B 2.5µg/ml. A Complete cell growth media for the normal epithelial human colon cell line CCD 841 CoN contained Minimum Essential Media (MEM), (FBS) 10% v/v (50mL) from VWR (45000-734) and 1% (5mL) antibiotics/anti-mycotic from VWR (45000-616). Colon cancer cell lines HCT-116 and LoVo used a base media made up of RPMI-1640, from HyClone (SH30605.01). To this base media the following were added: 10% v/v (FBS) (50mL) purchased from VWR (45000-734) and 1%v/v (5mL) antibiotics/anti-mycotic from VWR (45000-616). The colon cancer cell line HT-29 required Dulbecco's Modified Eagle Medium (DMEM) from VWR (45000-734), along with addition of 10% FBS v/v (50mL) from VWR (45000-734) and 1% v/v (5mL) antibiotics/anti-mycotic from VWR (45000-616).

For the growth of RWPE-1 and both MNU cell lines (WPE1-NA22 and WPE1-NB14), the cell growth media used was by Thermo-Fischer. The Thermo-Fischer-formulated growth kit Keratinocyte Serum Free Medium (K-SFM), (17005042) contained two additives required to grow these cell lines, namely bovine pituitary extract (BPE) and human recombinant epidermal growth factor (EGF). These components were added to the base medium at the ratios of 0.05mg/ml BPE and 5 ng/ml EGF to make the complete

growth media. It is recommended to not filter the complete culture media. The cell growth media for the DU 145 cell line used was ATCC[®]-formulated Eagle's Minimum Essential Medium, (30-2003[™]). To make the complete growth media, added 10% v/v (50mL) of FBS from VWR (45000-734) and 1% (5mL) antibiotics/anti-mycotic from VWR (45000-616). The cell growth medium for the PC-3 cell line used was ATCC[®]-formulated F-12K Medium, (30-2006[™]). To make the complete growth media, added 10% v/v (50mL) of FBS from VWR (45000-734) and 1% (5mL) antibiotics/anti-mycotic from VWR (45000-616).

Before addition of any FBS and antibiotic/anti-mycotic was done, approximately 50 ml of the cell growth media was pipetted out and stored under 4°C to be used later to culture cells during their period of starvation. All media preparation used sterilized pipettes to minimize contamination. The cell culture media was brought to the 37°C before using. All growth conditions required 5% carbon dioxide (CO₂) at a temperature of 37°C.

Thawing of Cell Stocks and General Cell Growth Protocol

This protocol was courtesy of Dr. Kathleen O' Connell. Stored frozen cells were thawed in a Corning[®] T-25 flask or a small petri-dish, the process is explained below. If the cell viability and the growth rate of a cell stock is typically high, for example most of the cancerous cell lines, the cells can be thawed directly into a Corning[®] T-75 flask.

1. The cell culture media needs to be at 37°C before use. The cryopreservation solution usually contains 5% v/v DMSO and is toxic to cells so thawing and suspension of cells in fresh media, needs to be done quickly. To do so, 5mL of media at 37°C, was used to transfer the contents from the cryogenic vial into the appropriately sized flask that

already contained 5ml of media at 37°C. The cells were left to grow in the incubator at 37°C.

2. After 24-hour duration, the flask containing the cells and media was inspected for any broken, particulate-like cell debris and the media was pipetted out and replaced with 10 ml of fresh media at 37°C. It is quite likely to find cell debris at this stage because many cells die due because they could not withstand the temperature and chemical shock while they are frozen.

3. The following day the old media was replaced with 10ml of fresh media added at 37°C. For the cells to maintain a healthy life cycle, continue the growth of cells in the incubator by removing and replacing old media while keeping a check for excessive cell debris and any possible infection until 75% confluence was reached. Typically, the cell growth was monitored three times per week. The cells media (containing phenol indicators) was replaced whenever the cell media in the flask would begin to turn a tinge of orange instead of its original pink color. If the cell media turns yellow, that usually happens when the cells have reached greater than 90% confluence levels and the cell media is completely exhausted. At this point of time, the cells are most likely under stress and are beyond the point of salvage. At any point of time if there was an excessive debris formation, the media was quickly replaced.

4. Before harvesting the cells for protein, split and passage cells 2-3 times to ensure that cells have achieved a normal growth momentum and are no longer in quiescent stage.

Trypsinization and Subsequent Passaging of Cells

Volumes provided are for Corning® T-75 size flasks.

1. After the cells reach 70% confluence levels, pipette off and discard the cell culture medium. Briefly rinse the adherent cell layer with 2ml of 10% Trypsin-1% EDTA solution in PBS and after rinsing, discarded the solution. To make 10% Trypsin-1% EDTA solution, add 50ml of Trypsin 10X to a solution containing 125ml of autoclaved 1% (w/v) EDTA solution in PBS and top off with 325ml PBS to make a total volume of 500ml.
2. Added 2-3ml of 10% Trypsin-1% EDTA solution to culture flask and left in 37°C incubator for 5- 15 minutes. This step allows the trypsin, a proteolytic enzyme to act on cellular surface proteins, which help the cell to adhere on to the flask. Trypsinization of cell surface proteins cause the cells to dislodge. The time taken for trypsinization is dependent on how firmly the cells attach to the flask. For example, CCD 841 CON cell line takes the longest time out of all the other cell lines grown, due to their epithelial morphology the cells firmly adhere on to the flask and take longer time to trypsinize. Continue to observe the cells under the microscope to ensure that cells do not remain in trypsin for too long, else causing an irreversible cellular damage. Cells are trypsinized until layers/clumps of cells start to dislodge from the flask. Keeping the cells in the incubator helps in trypsinization because the temperature facilitates the cell dispersal by enhancing action of trypsin.
3. Added some fresh warm media at 37°C to the flask containing trypsinized cells. This addition helps to suspend the action of trypsin. Pipette the cells along with the media and trypsin-EDTA mix into a 15 mL centrifuge tube and centrifuged for 5 minutes at 500 RPM. It is suggested not to increase the speed of centrifuge higher than 500, otherwise cells will die.

4. After centrifugation, discarded the supernatant and re-suspended the cell pellet using 1-2 ml of fresh cell culture media at 37°C.
5. After re-suspension, pass the cells through p1000 pipette several times to break the larger cell clumps. This ensures formation of smaller bunches of cells, so when these cells are introduced to a new culture flask, the cells will spread out uniformly and grow in several well spread out colonies thus causing a more regular reattachment of cells to the flask.
6. Added 6- 8mL of complete growth medium to the pipetted cells and aspirated the cells by gentle pipetting. Added appropriate aliquots of the cell suspensions to new culture vessels. The information on the ratios required for passaging of cells can be obtained from ATCC® website. Extra cells were discarded. Added the media to the flask to make it to the total volume for each flask type (5ml for T-25, 1 ml for T-75, and 20ml for T-125). Checked the following day for cell attachment to the flasks. If there is too much cell debris detected, then accordingly remove and replace fresh media 37°C.

Procedure of Cell Starvation

This step is required for cells whose media is to be collected to obtain cell-secreted proteins. After passaging the cells until they achieve an optimal growing momentum, the cells were starved after they reached a confluence level not less than 50% but not more than 70%. Cell starvation involved feeding the cells an incomplete cell growth media without FBS and supplements. The starvation period lasts for 48-hours (exact). Typically, following the starvation protocol designed by Dr. Erin Gatrone, the cell media from one T-75 flask is pipetted out into a 15ml falcon tube. This volume of media provides quite

concentrated pellet of serum protein, which is more than enough for a triplicate experiment.

Concentration of Cell Media after Starvation for Harvesting Cell-Secreted Proteins

After the cells (whose media is required for cell-secreted proteins) were starved for 48-hours, the pipetted cell media in 15 ml falcon tube was centrifuged (@1000 RPM) to remove cell debris. The media was concentrated down using Amicon® 3-KDa centrifuge tubes purchased from EMD Millipore. These centrifuge tubes have a cut-off volume that does not allow the media to be concentrated below a certain threshold. This is done to ensure that the 3-KDa membrane never dries out, else risking loss of protein.

Acetone Assisted Precipitation of Secreted Proteins from Concentrated Cell Media to form a Protein Pellet

1. After pipetting the cut-off volume from 3-KDa centrifuge tube, used approximately 4X v/v of ice-cold acetone for each ml of concentrated cell media to precipitate proteins. Using ice-cold acetone, causes the proteins to crash out of solution the precipitated proteins were stored in the -20°C freezer overnight.

2. After 24-hours, the protein sample was spun down and acetone was discarded. The protein aggregate was air dried under the fume hood by lightly covering the centrifuge tube. Protein concentration was ascertained using BCA assay, where proteins were re-suspended into carbonate buffer before (Procedure of BCA assay and method of making different buffers are listed under the experimental section after Chapter 2).

Extraction of Cellular Membrane Proteins

For extraction of proteins from cell membranes, larger number of cells are required than those required for collection of cell-secreted protein from media. Hence the

procedure involved usage of four T-75[®] flasks (70-75% confluent) to create one big cell pellet.

1. To do so, the cells were scraped off the bottom of flask using a Corning[®] plastic cell scraper. Small 1-2ml aliquots of ice-cold PBS was used to assist cell transfer and subsequent centrifugation of cells to generate a pellet from four flasks.

2. The ice-cold PBS in supernatant was discarded and the resulting cell pellet was stored in the -80 °C freezer for extended periods (not more than 3 months).

3. The plasma glycoproteins and proteins were extracted from this cell pellet using Qiagen[®] Plasma Membrane Protein Kit and the protocol provided with it. (Complete detail of this protocol is listed under experimental section in Chapter 4).

4. The protein extracted from the membrane protein kit was suspended in Qiagen[®] supplied buffer and was stored in -20°C until later use. Whenever required, the protein aliquot was thawed and its concentration was assessed using BCA assay (Chapter 2 experimental section).

5. Assessment of protein concentration was followed by labeling the protein using FITC involving Amicon Pro[®] 10-KDa centrifuge and exchange columns (Chapter 2 experimental section) and subsequent assessment of labeled protein by employing one more BCA assay.

Labeling of SL and SL bound with protein to Alizarin-Red (ARS)

1. Approximately 2mg of SL5 and SL5-Dab (stored in -20°C freezer) were weighed in separate 1.5mL Eppendorf tubes and tumbled using PBS buffer for 10-minutes at RT. PBS buffer was later discarded by pipetting.

2. SLs were then tumbled in 1ml PBS Screening Buffer (100mM NaH₂PO₄, 100mM Na₂HPO₄.7H₂O, 150mM NaCl and 10% glycerol, pH 7.2) for 15-minutes at RT.
3. The Screening Buffer was then removed and beads were tumbled in 10μl of 100μM solution of ARS in water at 37°C by wrapping in aluminum foil. The optimization of the quantity of ARS required was cited from Dr. Kevin L. Bicker's thesis (pg. 147). The incubation with ARS dye was done for 10 minutes at 37°C. SLs were subsequently washed three times with PBS (without glycerol) at RT before being analyzed under fluorescent microscope. Similarly, the SLs that were first bound to protein for 12-hours and then tumbled in 10μl of 100μM solution of ARS in water at 37°C also followed the same washing and imaging steps. (Bead imaging under microscope and image analysis is detailed in experimental section of Chapter 2).

3.9 REFERENCES

1. Zhang, Z.; Lang, J.; Cao, Z.; Li, R.; Wang, X.; Wang, W. Radiation-induced SOD2 overexpression sensitizes colorectal cancer to radiation while protecting normal tissue. *Oncotarget*. **2017**, *8* (5), 7791–7800.
2. Brattain, M. D.; Fine, W. D.; Khaled, F. M.; Thompson, J.; Brattain, D. E. Heterogeneity of malignant cells from a human colonic carcinoma. *Cancer Res.* **1981**, *41* (5), 1751-6.
3. Sul, J. Y.; Song, I. S.; Bae, C. S.; Chang, E. S.; Yoon, W. H. Metastatic Potential of Human Colon Cancer Cell Lines, LoVo and SW480. *J. Korean Cancer Assoc.* **1995**, *27* (2), 209-223.
4. Springsteen, G.; Wang, B., A detailed examination of boronic acid-diol complexation. *Tetrahedron* **2002**, *58*, 5291-5300.
5. Palit, D. K.; Pal, H.; Mukherjee, T.; Mittal, J. P., Photodynamics of the S1 state of some hydroxy- and amino-substituted naphthoquinones and anthraquinones. *J. Chem. Soc. Faraday Trans.* **1990**, *86* (23), 3861-3869.
6. Paul, H.; Reginato, A. J.; Schumacher, H. R., Alizarin Red S Staining as a Screening Test to Detect Calcium Compounds in Synovial Fluid. *Arthritis Rheum.* **1983**, *26* (2), 191-200.
7. Peracaula, R.; Tabares, G.; Royle, L.; Harvey, D.J., Dwek, R.A.; Rudd, P.M.; de Llorens, R. Altered glycosylation pattern allows the distinction between prostate specific antigen from normal and tumour origins. *Glycobiology* **2003**, *13*, 457-470.
8. Barrabes, S. Glycosylation of serum ribonuclease I indicates a major endothelial origin and reveals an increase in core-fucosylation in pancreatic cancer, *Cancer Res.* **2007**, *17*, 388-400.
9. Kyselova, Z.; Mechref, Y.; Kang, P.; Goetz, J. A.; Dobrolecki, L. E.; Sledge, G. W.; Schnaper, L.; Hickey, R. J.; Malkas, L. H.; Novotny, M.V. Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. *Clin Chem.* **2008**, *54*, 1166–1175.
10. Peracaula, R.; Barrabes, S.; Sarrats, A.; Rudd, P. M.; de Llorens R. Altered glycosylation in tumours focused to cancer diagnosis. *Dis Markers.* **2008**, *25*, 207–218.
11. Osumi, D.; Takahashi, M.; Miyoshi, E.; Yokoe, S.; Lee, S. H.; Noda, K. Core-fucosylation of E-cadherin enhances cell-cell adhesion in human colon carcinoma WiDr cells. *Cancer Science*, **2009**, *100*, 888–895.
12. Listinsky, J.J. Siegal, G.P. Listinsky, C.M. The emerging importance of α -L-fucose in human breast cancer: a review. *Am. J. Transl. Res.*, **2011**, *3* (4), 292–322.
13. Webber, M. M.; Quader, S. T. A.; Kleinman, H.K.; Bello-DeOcampo, D.; Storto, P. D.; Bice, G.; DeMendonca-Calaca, W.; Williams, D. E. Human Cell Lines as an InVitro/InVivo Model for Prostate Carcinogenesis and Progression. *The Prostate*, **2001**, *47*, 1-13.
14. Pfeiffer, M. J.; Schalken, J. A. Stem Cell Characteristics in Prostate Cancer Cell Lines. *European Urology*, **2009**.
15. Zhang, H.; Zhang, Y.; Duan, H. O.; Kirley, S. D.; Lin, S. X.; McDougal, W. S.; Xiao, H.; Chin-Lee Wu. TIP30 is associated with progression and metastasis of prostate cancer. *Int. J. Cancer* **2008**, *123*, 810–816.

16. Klarmann, G. J.; Hurt, E. M.; Mathews, L. A.; Zhang, X.; Duhagon, M. A.; Mistree, T.; Thomas, S. B.; Farrar, W. L. Invasive Prostate Cancer Cells Are Tumor Initiating Cells That Have A Stem Cell-Like Genomic Signature. *Clin. Exp. Metastasis* **2009**, *26* (5), 433–446.
17. Munkley, J.; Elliott, D. J. Hallmarks of glycosylation in cancer. *Oncotarget* **2016**, *7* (23), 35478–89.
18. Ruhaak, L. R.; Miyamoto, S.; Lebrilla, C. B. Developments in the Identification of Glycan Biomarkers for the Detection of Cancer. *Mol. Cell Proteomics*. **2013**, *12* (4), 846–855.
19. Pinho, S. S.; Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer* **2015**, *15*, 540–555.
20. Kirmiz, C.; Li, B.; An, H. J.; Clowers, B. H.; Chew, H. K.; Lam, K. S.; Ferrige, A.; Alecio, R.; Borowsky, A. D.; Sulaimon, S. A serum glycomics approach to breast cancer biomarkers. *Mol. Cell Proteomics* **2007**, *6*, 43–55.
21. Mechref, Y.; Hu, Y.; Garcia, A.; Zhou, S.; Desantos-Garcia, J. L.; Hussein, A. Defining putative glycan cancer biomarkers by MS. *Bioanalysis* **2012**, *4*, 2457–2469.
22. An, H. J.; Miyamoto, S.; Lancaster, K. S.; Kirmiz, C.; Li, B.; Lam, K. S.; Leiserowitz, G. S.; Lebrilla, C. B. Profiling of glycans in serum for the discovery of potential biomarkers for ovarian cancer. *J. Proteome Res.* **2006**, *5*, 1626–1635.
23. Storr, S. J.; Royle, L.; Chapman, C. J.; Hamid, U. M.; Robertson, J. F.; Murray, A.; Dwek, R. A.; Rudd, P. M. The O-linked glycosylation of secretory/shed MUC1 from an advanced breast cancer patient's serum. *Glycobiology*, **2008**, *18*, 456–462.
24. Barry, M. J., Prostate-specific-antigen testing for the early diagnosis of prostate cancer. *N. Engl. J. Med.* **2001**, *344*, 1373–1377.
25. Koprowski, H.; Herlyn, M.; Steplewski, Z.; Sears, H. F., Specific antigen in serum of patients with colon carcinoma. *Science* **1981**, *212* (4490), 53–55.
26. Magnani, J. L.; Steplewski, Z.; Koprowski, H.; Ginsburg, V., Identification of the gastrointestinal and pancreatic cancer-associated antigen detected by monoclonal antibody 19-9 in the sera of patients as a mucin. *Cancer Res.* **1983**, *43* (11), 5489–92.
27. Nuti, M.; Teramoto, Y. A.; Mariani-Costantini, R.; Hand, P. H.; Colcher, D.; Schlom, J. A monoclonal antibody (B72.3) defines patterns of distribution of a novel tumor associated antigen in human mammary carcinoma cell populations. *Int. J. Cancer* **1982**, *29* (5), 539–545.
28. Cao, Y.; Karsten, U. R.; Liebrich, W.; Haensch, W.; Springer, G. F.; Schlag, P. M. Expression of Thomsen-Friedenreich-related antigens in primary and metastatic colorectal carcinomas. A reevaluation. *Cancer* **1995**, *76* (10), 1700–1708.
29. Thor, A.; Ohuchi, N.; Szpak, C. A.; Johnston, W. W.; Schlom, J. Distribution of oncofetal antigen tumor-associated glycoprotein-72 defined by monoclonal antibody B72.3. *Cancer Res.* **1986**, *46* (6), 3118–3124.
30. Durrant, L. G.; Noble, P.; Spendlove, I. Immunology in the clinic review series; focus on cancer: Glycolipids as targets for tumour immunotherapy. *Clin. and Exp. Immunology* **2012**, *167* (2), 206–215.
31. Hua, S.; An, H. J.; Ozcan, S.; Ro, G. S.; Soares, S.; DeVere-White, R.; Lebrilla, C. B. Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers. *Analyst*, **2011**, *136*, 3663–3671.

32. Saldova R.; Fan, Y.; Fitzpatrick J. M. ; William R. ;Watson, G. ; Rudd, P. M. Core fucosylation and α 2-3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia. *Glycobiology* **2011**, *21* (2), 195–205.
33. Zhao, Y. P., Ruan, C. P., Wang, H., Hu, Z. Q., Fang, M., Gu, X. Identification and assessment of new biomarkers for colorectal cancer with serum N-glycan profiling. *Cancer* **2012**, *118* (3), 639–650.
34. Shin, S. Y.; Kim, C. G.; Jung, Y. J.; Lim, Y. L.; Lee, Y. H. The UPR inducer DPP23 inhibits the metastatic potential of MDA-MB-231 human breast cancer cells by targeting the Akt–IKK–NF- κ B–MMP-9 axis. *Nature Scientific Reports* **2016**, *6*, 1-12.
35. Lin, N.; Mascarenhas, J.; Sealover, N.R.; George, H.J.; Brooks, J.; Kayser, K.J.; Gau, B.; Yasa, I.; Azadi, P.; and Archer-Hartmann, S. Chinese hamster ovary (cho) host cell engineering to increase sialylation of recombinant therapeutic proteins by modulating sialyltransferase expression. *Biotechnol. Prog.* **2015**, *31*, 334–346.
36. ATCC CCD-841-CON (ATCC® CRL-1790™). <https://www.atcc.org/products/all/CRL-1790.aspx> .
37. ATCC HCT-116 (ATCC® CCL-247™). <https://www.atcc.org/products/all/CCL-247.aspx> .
38. ATCC HT29 (HTB-38™). <https://www.atcc.org/products/all/HTB-38.aspx> .
39. ATCC LoVo (ATCC® CCL-229™). <https://www.atcc.org/Products/All/CCL-229.aspx> .
40. ATCC MDA-MB-231 (ATCC® HTB-26™). <https://www.atcc.org/products/all/HTB-26.aspx> .
41. RWPE-1 (ATCC® CRL-11609™). <https://www.atcc.org/Products/All/CRL-11609.aspx> .
42. WPE1-NA22 (ATCC® CRL-2849™). <https://www.atcc.org/Products/All/CRL-2849.aspx> .
43. WPE1-NB14 (ATCC® CRL-2850™). <https://www.atcc.org/Products/All/CRL-2850.aspx> .
44. DU 145 (ATCC® HTB-81™). <https://www.atcc.org/products/all/HTB-81.aspx> .
45. PC-3 (ATCC® CRL-1435™). <https://www.atcc.org/Products/All/CRL-1435.aspx> .

CHAPTER 4:

IN-VITRO CHEMICAL TARGETS OF EXTENDED SL ARRAY AND ITS STRUCTURAL CORRELATIONS WITH CANCER TARGETS

4.0 OVERVIEW

Using statistical methods of data analysis, the previous chapter highlighted that certain SLs, such as SL2, SL4, 4R and SL11-Y4A, due to their charges, contributed to the precision/tightness in cancer classification, whereas SL7 and SL11-R8A – owing to their greater number of phenyl rings- contributed to the overall accuracy of classification of cancer. This classification aligned well with the increasing metastatic potential of the cancer cell lines used.

Moreover, the aforementioned SLs were also useful at tissue typing the cancers of colon and prostate, which shows that chemical basis of SLs' interaction with the proteins, coincides with the statistical parameters (variance) which contribute to the ability of proteins to be classified by the array.

However, the questions which arise are:

1. What are the chemical targets of the SL array?
2. Do cancer-associated glycans (CAGs) respond to SLs? If yes, then what is the chemical basis for it?
3. Can SL array be useful in classification of tissue targets?

4. Can the abovementioned chemical interactions become the guiding benchmarks at designing future SL targets specific for colon and prostate cancer?

The current chapter intends to address these questions, beginning by explaining the chemical basis of SL interaction with colon and prostate cell lines in sections 4.1 and 4.2. The chemical basis of CAG and SL interaction is discussed in section 4.3. The SL array based classification of patient tissues is discussed in section 4.4. Finally, section 4.5 presents a working model for generating tissue-targeted SL development.

4.1 SIALIC ACID QUANTIFICATION OF COLON AND PROSTATE CELLS

Previous two chapters indicate that SLs target through positively charged amino acid residues (i.e. Arginine R). The chemical target could be a negatively charged glycan or proteins. Based on literature precedence, it is reasonable to believe that sialic acid modifications are important in defining SL-glycan interactions.¹ Previous published work (Bicker, K. L. *ACS Chem. Sci.*) (Chapter1: ref 29) in the group also reports a preferred interaction of SL5 to methylated-sialic acid over other common monosaccharides like (methyl-fucopyranosides, methyl galactopyranoside and methyl 2-acetamido-2-deoxyglucopyranoside). Hence before delving into non-specific SL peptide-protein interactions, it was decided to investigate for any specific glycan-SL interaction. It was likely to begin investigation by choosing sialic acid. There were two reasons to quantify sialic acid residues *in-vitro*:

1. Sialic acid (N-Acetylneuraminic acid) is a negatively charged, non-reducing sugar (refer **Figure 1.1**), and since SLs have been shown to interact via their positive charges, investigation of any possible interaction through sialic acids on protein residues seems quite likely.

2. Over-sialylation is commonly reported, in terminal glycans on protein structures both during oncogenesis and cancer progression. Thus cleaving sialic acid off to quantify, as well as to study the rest of the integral protein in their absence, seemed to be a reasonable beginning.¹ Taking this approach may cause the least intended perturbation to the integral protein structure.

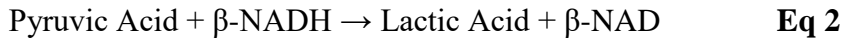
The objective of this section is to investigate the amount of sialic acid residue isolated from glycoprotein secreted from both prostate and colon derived cell lines. Later to investigate if the quantified sialic acid has any correlation to the virulence of both colon and prostate cancers.

Quantification of sialic acid is broadly accomplished using two methods, chemical as well as enzymatic. In the past, chemical cleavage of sialic acid residues from glycoproteins has been performed in our lab using hydrazine. This chemical method induces non-specific denaturation of purified proteins (e.g. BSM) because of higher temperatures and usage of hydrazine. The chemical leaching method risks the protein core to be denatured and may potentially disturb any possible protein-binding epitopes. Therefore, enzymatic cleavage is employed in the following experiment, which offers a softer method of removing sialic acid residues and leaves the protein core as intact as possible. In this enzymatic cleavage, bound sialic acid residues (all types of sialic acid linkages) from complex glycoconjugates (such as cell surface glycoproteins, polysialic acids, and capsular polysaccharides) are released through the action of a non-specific neuraminidase/sialidase enzyme. The free sialic acid residue is broken down into pyruvic acid using N-Acetylneuraminic acid aldolase. The aldolase enzyme catalyzes the following reversible reaction:

N-Acetylneuraminic acid (sialic acid) → N-Acetylmannosamine + Pyruvic acid

Eq 1

Pyruvic acid, in turn is reduced to lactic acid using β-NADH and lactic acid dehydrogenase enzyme as follows:



Under proper temperature conditions (37 °C), the first forward reaction predominates, and when coupled with β-NADH reduction of pyruvic acid, the reaction goes to completion. β-NADH oxidation can be accurately measured spectrophotometrically. Hence, the total amount of sialic acid residue obtained from each cellular extract correlates to the nanomoles of β-NADH oxidized. Total sialic acid counts from secreted proteins, isolated from the media of four different colon and five different prostate cell lines, are listed below in **Table 4.1**.

Table 4.1: Illustrating total sialic acid residue count of secreted proteins extracted from media of cell lines.					
Human colon cell lines			Human prostate cell lines		
Cell line name	metastatic potential	# sialic acid residues (nmol)	Cell line name	metastatic potential	# sialic acid residues (nmol)
CCD 841 CoN	-	56	RWPE-1	-	102
HCT-116	low	108	WPE1-NA22	low	102
HT-29	low	74	WPE1-NB14	low	80
LoVo	high	77	DU 145	high	270
			PC-3	high	240

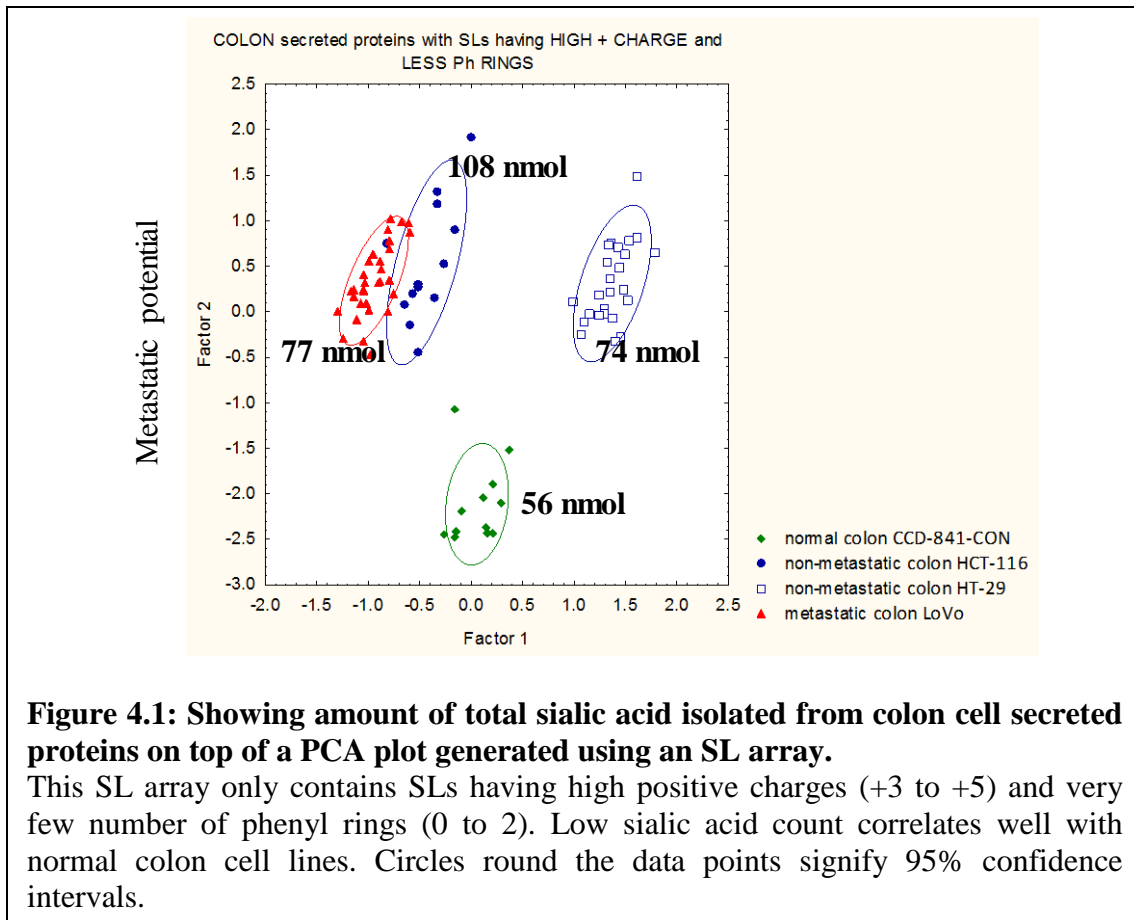
While other groups also have successfully quantified sialic acid from cell membrane proteins,^{2,3} our effort to determine levels secreted into the media is unique and innovative providing an insight into the development of serum-based diagnostics. Metastatic

potential for each cell line from the literature is confirmed using standard cell invasion assays (**Table 3.2** and **Table 3.7**).⁴

Based on the literature precedence on the varied sialic acid expression during oncogenesis of proteins extracted from colon cell membranes,⁵ it is anticipated that sialic acid expression correlates with the metastatic potential of colon cell lines. The total sialic acid content values (**Table 4.1**) show that the content increases going from normal to cancer cells.

To further elucidate any correlation between sialic acid content and metastatic potential, PCA outputs of colon and prostate were evaluated. In secreted colon proteins however, this happens when we compare the PCA output of SL array containing *only highly positively charged SLs* (SL3, SL4, SL5, SL5-R1511A, SL5-Dab, SL5-RRAc, 4R and 5R) with very few (0-2) number of phenyl rings.

Figure 4.1 (same as **Figure 3.6E**) shows that by only using highly positively charged SLs, different colon cells can be discerned. Factor 1 and 2 are the top two statistical parameters that describe most of the variance of the data set. Factor 2 seems to show correlation with metastatic potential. **Figure 4.1** also attempts to correlate the sialic acid concentration and metastatic potential of colon cell lines. It shows that sialic acid quantification especially works well at classifying normal from cancer. This means that one of the statistical basis of discrimination, through which the SL array is classifying colon cell lines, is through varied sialic acid concentration, while the positive charges derived from arginine residues of SLs assist in targeting these sialic acid residues.

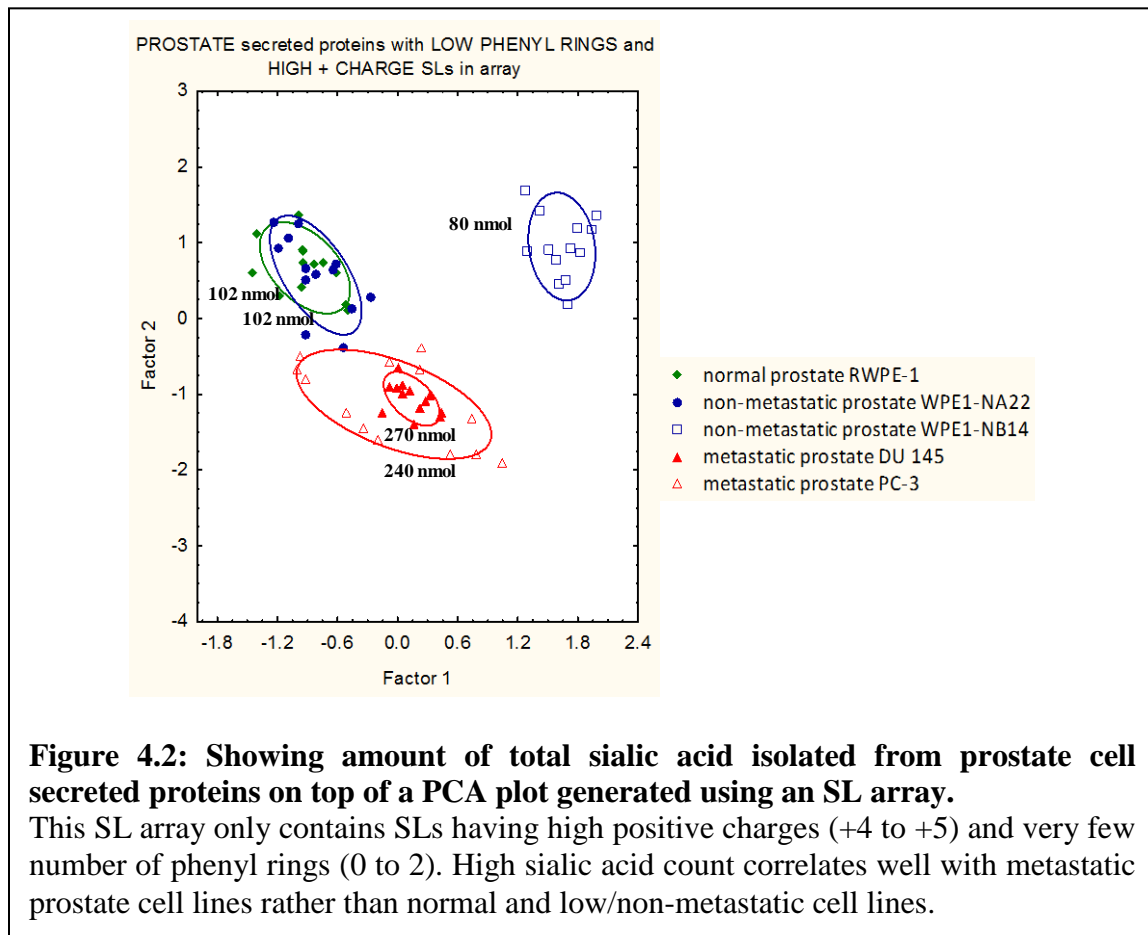


It is noted that the HCT-116 and LoVo cell lines group together, which is because the SLs that help in tightening their clustering (SL7 and SL11-R8A) are not included in the high positive charge-low phenyl ring model (refer **Figure 3.7**).

Similarly, sialic acid concentration and PCA output of prostate cell lines is illustrated in **Figure 4.2**. The PCA output again makes use of the SL array that contains highly positively charged SLs (SL5-R1511A, SL5-Dab, SL5-RRAc, 4R and 5R) with very few (0-2) number of phenyl rings. By only incorporating highly positive SLs, the model did not discern prostate cell lines that well, especially normal and low/non-metastatic ones. This was expected, since to describe prostate cancer model, the SLs with greater number

of phenyl rings were paramount and those SLs are missing from the PCA model depicted in **Figure 4.2**.

It is noted that the amount of sialic acid obtained from normal prostate cell lines exceeds even that of metastatic colon cell lines. Although, the total sialic acid content of DU 145 and PC-3 cell lines exceeds that of RWPE-1 as seen in **Figure 4.2** (same figure as **Figure 3.11B**), it is also noted that RWPE-1 and WPE1-NA22 cell lines have the same amount of sialic acid content. This may mean that using sialic acid quantification to diagnose prostate cancer at early stage may not be very helpful. Literature also disputes the use of sialic acid quantification for prostate cancer diagnosis.⁶



As mentioned in the last chapter in **Figure 3.11A**, for discerning various prostate cell lines, high positively charged SLs contribute to model precision, by increasing the tightness of the grouping. This can be explained by positive charge of arginine that enhances binding through sialic acid residues of prostate cells. In **Figure 3.11A**, factor 1 roughly aligns with metastatic potential, pinpointing that the ionic interactions being one of the predominating factors towards disease classification by the array.

But to accurately discern each prostate cell line, the SLs containing *greater number of phenyl rings and lesser positive charges* play an important role, too. Which is why **Figure 4.2** model (that does not contain SLs with high number of phenyl rings) is performing badly at classification.

RWPE-1 and WPE1-NA22 always seem to cluster nearby, which may be due to the latter being an MNU-derived cell line from the former and is quite low in its metastatic potential (**Table 3.7**). This supports the need for better SLs that specifically target prostate cancer for future investigations.

From this section, comparing the amount of sialic acid between CCD 841 CON and HT-29 cell lines, it is reasonable to conclude that change in sialylation correlates well with early diagnosis of colon cancer. Whereas comparing the amount of sialic acid between RWPE-1 and WPE1-NA22 cell lines, it can be concluded that change in sialylation can only predict prostate cancer at its later stages and is not a predictive marker for earlier detection of prostate cancer.

4.2 INTERACTION OF SL-ARRAY WITH SECRETED COLON PROTEINS THAT CONTAIN OR ARE DEVOID OF SIALIC ACIDS

The purpose of doing this experiment is to determine the source of negative charges that SL array is targeting for its discrimination. Is the observed SL array classification primarily from sialic acid or from lower pI protein isolated from cell culture media? It is important to investigate whether SL array is specific to a glycan moiety or is it simply involved in some transient non-specific interaction. This might help us ultimately to understand and enhance the utility of the array towards diagnosing and staging colon cancer. Since there is an evidence that global fucosylation⁷ and sialylation⁸ correlate with cell metastasis, it will be useful to compare array outputs that derive from samples that are rich in sialic acid residues to the outputs that are not.

For this experiment, normal colon (CCD 841 CON) and highly metastatic colon (LoVo) cell lines were used. Both the protein samples, with and without sialic acid (after the action of sialidase enzyme) were incubated with a 20-unit SL array. **Figure 4.3** below, illustrates SL array assisted, 4-class guided LDA grouping of colon cell isolated proteins, with and without sialic acids. In the **Figure 4.3** going from left to right is CCD 841 CON with sialic acid residues, followed by CCD 841 CON and LoVo after removal of all sialic acid residues and the last cluster indicates LoVo with sialic acid residues attached to the proteins. Close clustering of proteins devoid of sialic acid residues shows a loss of discriminative ability of SL array. This indicated that the SLs' primarily target are sialic acid residues of the cell lines. Role of negative charge from protein in binding to SL could not be evaluated because sialidase enzyme may have caused some protein

denaturation. Using circular dichroism spectroscopy to evaluate protein stability as a function of added denaturants is suggested in future experimentation.

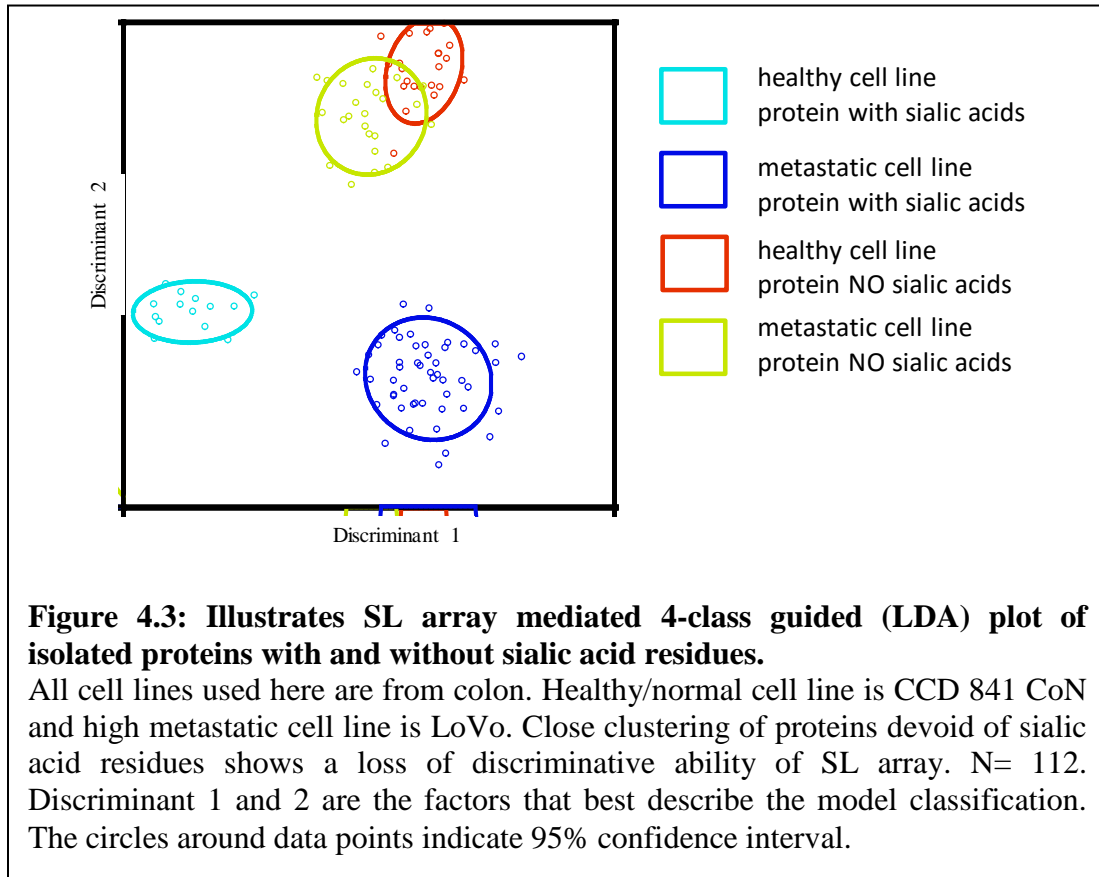
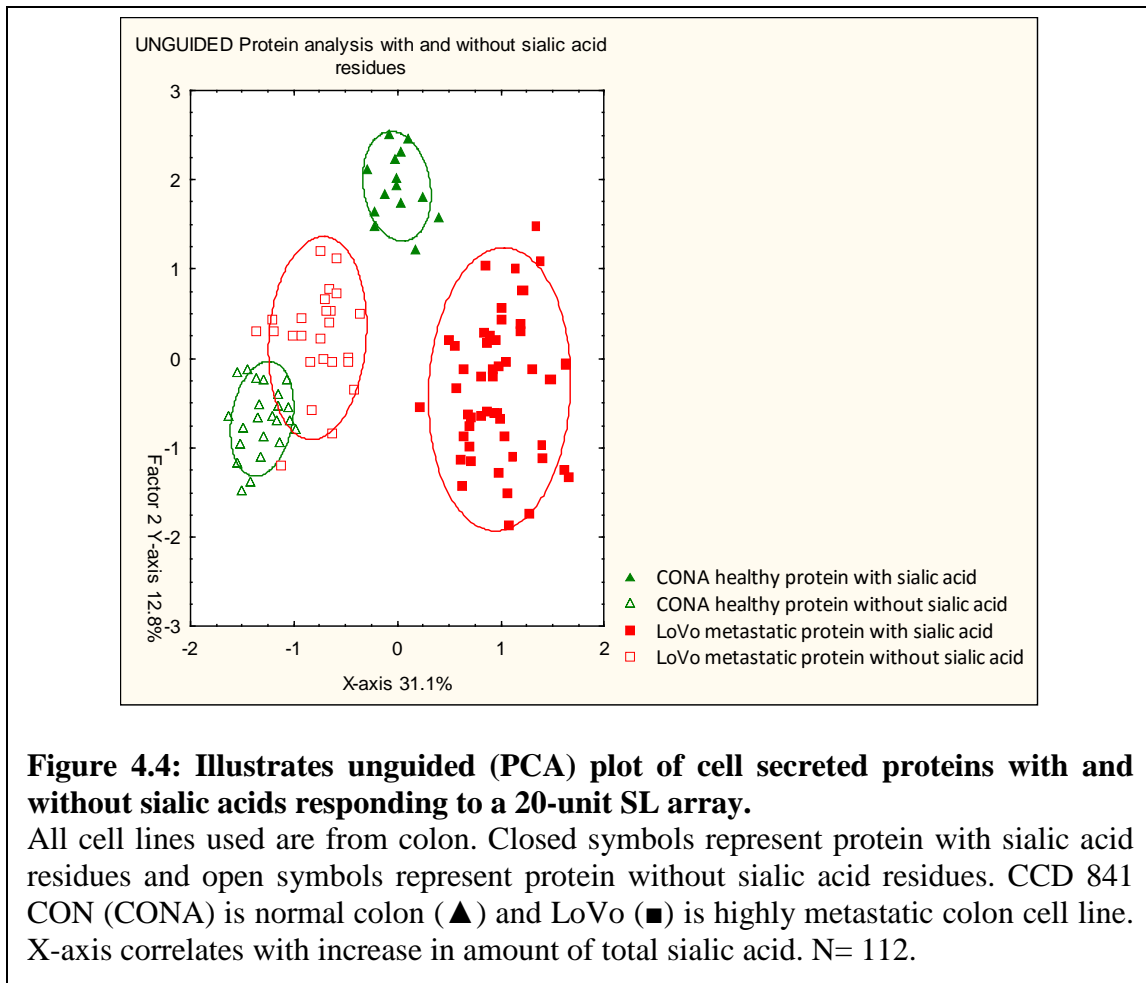


Figure 4.3 showed LDA, a guided method to observe different protein groupings and did not indicate the actual spread of the individual data points according to their variance. That is because guided grouping method tries to minimize the distance within each circle and maximize the distance between two different circles. In order to remove the statistical bias in data, unguided grouping (PCA) with respect to variance of each data point was conducted and is illustrated in **Figure 4.4**.



The greater spread out of LoVo (■ and □) may be indicating a greater amount of protein changes associated with increased metastatic potential. In **Figure 4.4**, Factor 1 represents the largest factor that described the greatest variance in the model and interestingly also correlated the amount of sialic acid residues in the different protein samples (open symbols indicate protein samples devoid of sialic acid).

From this section, it can be concluded that the primary chemical cause for SL array assisted classification of various colon cancer cell lines is the presence of sialic acid residues associated with the proteins extracted from these cell lines. Future work would involve similar investigation with prostate cancer.

4.3 RESPONSE OF CANCER ASSOCIATED GLYCANS TO SL ARRAY

As discussed earlier in sections 1.3 and 3.5, Cancer-Associated Glycans (CAGs) are the cause of aberrant glycosylation in oncogenic cells. The following section attempts to evaluate the ability of the SL array to discriminate several different CAGs. Some of these CAGs are complex glycans i.e., TF antigen, Le^a, Le^x, sLe^a, sLe^x and some are simpler monosaccharides like fucose. Fucose, (refer Figure 1.1) is not a CAG, rather is a part of several CAGs. Besides that, core-fucosylation of N-linked glycans that is also known to correlate with metastasis.⁷ These glycans were chosen because they represent some of the more common saccharide motifs over-expressed by cancerous cells. These CAGs are also composed of many of the same monosaccharides that were also found in glycoproteins that were used to establish proof-of-concept in chapter 1 and 2 (OVA, PSM and BSM).

For evaluation of these six CAGs (**Figure 1.1**), a 19-unit SL array containing SL1-SL9, SL11, SL1Dab-SL5Dab (no phenyl-boronic acids), SL5-R1,5,11A, SL5-RRAc, 4R and 5R was used. This different SL array along with some of the SLs from extended lectin array also contained SLs without 2-PBA in them, to investigate the direct impact loss of PBA has during interaction of CAGs with SLs.

Firstly, the binding affinity of each SL was investigated with six different glycans for which expression is commonly altered during prostatic and colorectal cancers. This was done primarily to identify SLs selective towards certain glycans and then to further comprehend the chemical basis of the selectivity. To ensure that SLs bind to glycans and not the fluorescent dye and also to maintain the dye:glycan ratio, we made use of biotin tagged CAGs. After incubation with the CAGs, fluorescently tagged streptavidin was

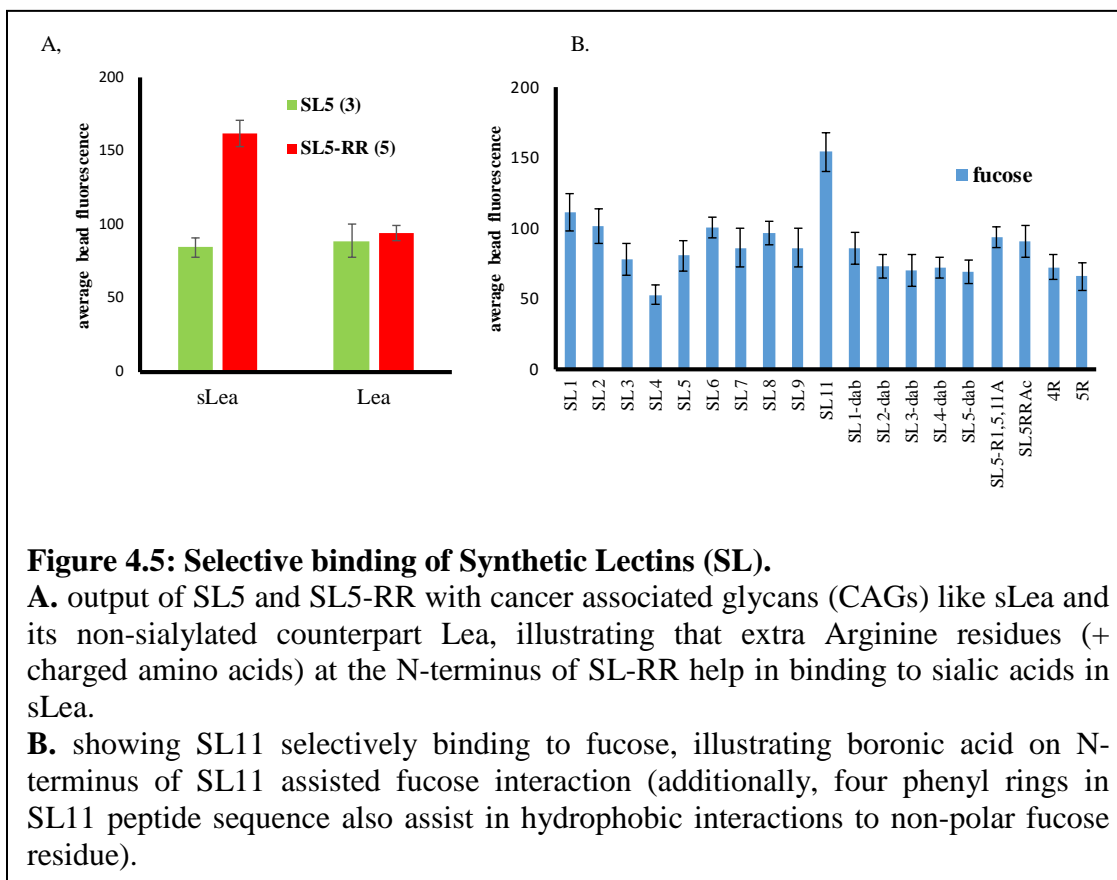


Figure 4.5: Selective binding of Synthetic Lectins (SL).

A. output of SL5 and SL5-RR with cancer associated glycans (CAGs) like sLea and its non-sialylated counterpart Lea, illustrating that extra Arginine residues (+ charged amino acids) at the N-terminus of SL-RR help in binding to sialic acids in sLea.

B. showing SL11 selectively binding to fucose, illustrating boronic acid on N-terminus of SL11 assisted fucose interaction (additionally, four phenyl rings in SL11 peptide sequence also assist in hydrophobic interactions to non-polar fucose residue).

incorporated to transduce an optical signal upon conjugation of biotin to streptavidin. The complete output of six CAGs with 19-SLs is covered in APPENDIX D, but some noteworthy points are discussed below.

Several chemical correlations could be drawn after investigating all the CAGs .

Figure 4.5A shows increased binding of SL5-RR over SL5 with sialylated Le^a. This is an example of the way the SL-glycan interaction is enhanced upon the introduction of additional positively charged arginine residues (R) in the sequence of SL5-RR, thereby leading to a stronger charge-pairing interaction with sLe^a over non-sialylated Le^a. SL5-RRAc has an overall +5 charge i.e. two extra positive charges than SL5.

Figure 4.5B displays selective binding of SL11 over 18 other SLs to non-reducing fucose. This could be attributed to extra boronic acid present on the free N-terminus of

SL11, thus assisting in fucose binding. SL11 also contains four phenyl rings (three from phenyl boronic acids and one from tyrosine) that could contribute to CH- π type interactions with fucose. SL11 was identified (courtesy Dr. Anna Veldkamp) from a library of SL-polymeric beads when the library was incubated with extracted membrane-bound proteins from RWPE-1 and PC-3. PC-3 protein extract competed against RWPE-1 to bind to SL11, thus making SL11 a prostate cancer 'hit'. The theoretical size of the SL library was 11^5 peptide sequence permutations (ref: Section 4.5).

To investigate the cross-reactivity of SLs and their applicability to distinguish all six CAGs, an array was constructed fostering the hypothesis of SL-glycan interactions and boronic acid-diol binding. Depicted in **Figure 4.6A** is the output when a guided statistical approach LDA was employed to accurately discern these six CAGs with >99% classification accuracy. It is noteworthy that sialylated CAGs (sLe^a and sLe^x) and their non-sialylated counterparts (Le^a and Le^x) are close to each other in these models where the only structural difference is the regio-chemistry of the linkage to the core GlcNAc moiety. Additionally, Le^a also closely clustered near sLe^a, where the difference is merely that of presence of sialic acid residue in the latter. This signifies the capability of the array to discern sialylated from non-sialylated and sLe^a from sLe^x, the latter have only small structural differences. TF-antigen (TFA) is a disaccharide and does not possess the glycan motif shared by any other CAGs involved in this study (Gal β 1-3GalNAc α 1), hence it is uniquely classified.

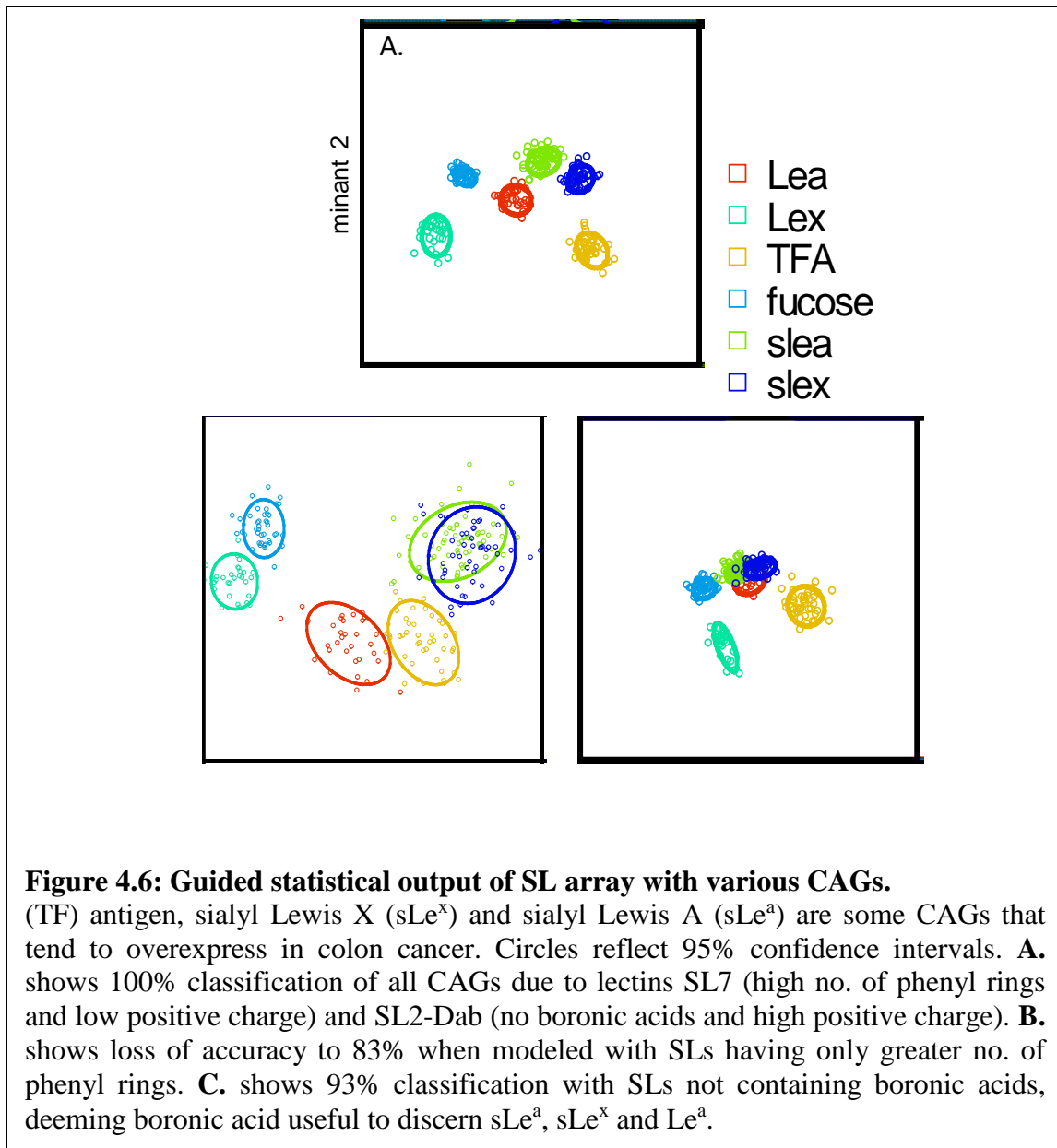
Based on the comparison of F-scores of all 19 SLs, the two SLs, which contributed the most to the classification of all six CAGs are SL7 and SL2-Dab. It is hypothesized that these two SLs dominate, primarily because of greater number of aromatic rings in

SL7 (contributing via CH- π interactions) and the greater amount of positive charges in SL2-Dab (contributing via ionic interactions). The F-score data from LDA can be relied on because the LDA and PCA models picked out the same SLs that contributed to the discernment of CAGs.

An LDA model with SLs containing a greater number of aromatic rings (and low positive charges), similar to SL7, resulted in a decrease in the tightness of model (**Figure 4.6B**), indicating a loss in precision and compromising the classification accuracy, reducing it to 83%. Interestingly, the two sialylated CAGs, sLe^a and sLe^x overlapped with each other the moment SLs with great number of Arginine residues were removed from the model.

Similarly, the LDA model, with SLs having a greater number of positive charges and no phenyl-boronic acids (similar to SL2-Dab) shows a reduction in classification accuracy to 93% (**Figure 4.6C**). These results suggest that boronic acid residues are important to discriminate sLe^a, sLe^x and Le^a. In addition, the model derived from positively charged SLs retains excellent ‘tightness/precision’, indicating that these positively charged SLs provide a microenvironment necessary for accurate binding.

The unguided (PCA) statistical models offer similar insight, indicating that *positively charged amino acids (like Arginine) in SLs are important for precision and phenyl boronic acids are important for accuracy.*



In **Figure 4.6B**, it was also noted that with SLs containing high number of phenyl rings and low positively charged sequences, sialylated CAGs like sLe^a and sLe^x did not classify well and grouped together. These two sialylated CAGs lost the grouping accuracy in the model, further indicating sialic acid- arginine interactions.

One of the concerns in these experimentations is that it is possible for the SLs to interact, not only with the glycans, but also with the protein portion of glycoproteins. To

counteract this, in this analysis since the protein component, FITC- streptavidin, is the same for each glycan that is being analyzed, therefore any observed difference in the response from the array must be attributed to the glycan constituent and not to the protein (streptavidin) component. Given the structural similarities between these glycans, it is remarkable that there were not many misclassifications.

In summary, these results validate our ability to differentiate structurally similar cancer associated glycans with high accuracy using a small, cross-reactive SL Array. Sialylated CAGs are the direct targets of SLs with greater number of Arginine residues. Loss of the SLs with greater number of Arginine residues limits the arrays' ability to discern sialylated CAGs. On the other hand, fucose and SL11 show an exceptional binding interaction, that may be due to CH- π type interaction of pyranose ring with greater number of phenyl rings in SL11.

4.4 APPLICATIONS OF SL ARRAY TO DISCERN HUMAN PROSTATE TISSUE SAMPLES

While the current array discriminates cell lines with classification accuracies that may be improved by refining the SLs in the array and better statistical modeling, this is not our ultimate goal. Cell lines do serve as acceptable *in vitro* models for investigating cancer, however they do not always represent the complexity of the tumor microenvironment. To examine whether our SL array could work with clinical specimens, patient-matched tissue from four prostate cancer patients were analyzed using the extended SL Array. Prostate tissue was obtained from the Prostate Tissue Core Facility at USC (courtesy: Ms. Ella Weinkle). Each patient tissue sample's virulence was confirmed by tissue histology (courtesy: Dr. Loulia Chatzistamou) and consisted of one

tumor sample and one normal or healthy sample taken from an adjacent site. Out of the four samples, two were of patients from African American (AA) and two from Euro-American (EA) descent. Pathologists grade prostate cancers using numbers from 1 to 5 based on how much the cells in the cancerous tissue look like normal prostate tissue under the microscope. This is known as the Gleason system. Grades 1 and 2 are not used for biopsies, because most biopsy samples are grade 3 or higher. If the cancerous tissue looks much like normal prostate tissue, a grade of 1 is assigned and if cells and their growth patterns look very abnormal, it is assigned 5. Grades 2 through 4 have features in between these two extremes. Since prostate cancers often have areas with different grades, a grade is assigned to two areas that make up most of the cancer. These 2 grades are added to yield the Gleason score (also called the Gleason sum). The highest Gleason score can be 10. The first number assigned is the grade that is most common in the tumor. For example, the Gleason score written as 3+4=7, is understood as the majority of the tumor being grade 3 and some of it being grade 4, when they are added together, they make up for a Gleason sum of 7.⁹ Histology of the acquired tissues confirmed each metastatic tissue sample to contain adenocarcinoma (Gleason sum (3+4=7)).

For the protein extraction, the tissues were pulverized using mortar and pestle in liquid nitrogen and the resulting powder added to lysis buffer. The resulting tissue slurry was centrifuged and the supernatant containing cells were collected. Membrane proteins were extracted using the Qiagen membrane extraction kit, labeled with FITC and incubated with extended SL array. Fluorescence intensity data were collected for each sample using fluorescence microscopy. Outliers were rejected at 1.8 interquartile distances (IQDs) and intensity readings were normalized to one using the brightest

reading for each patient sample. Using de-identified patient disease data, LDA analysis was carried out to determine the ability of the array to differentiate patient samples based on a number of factors.

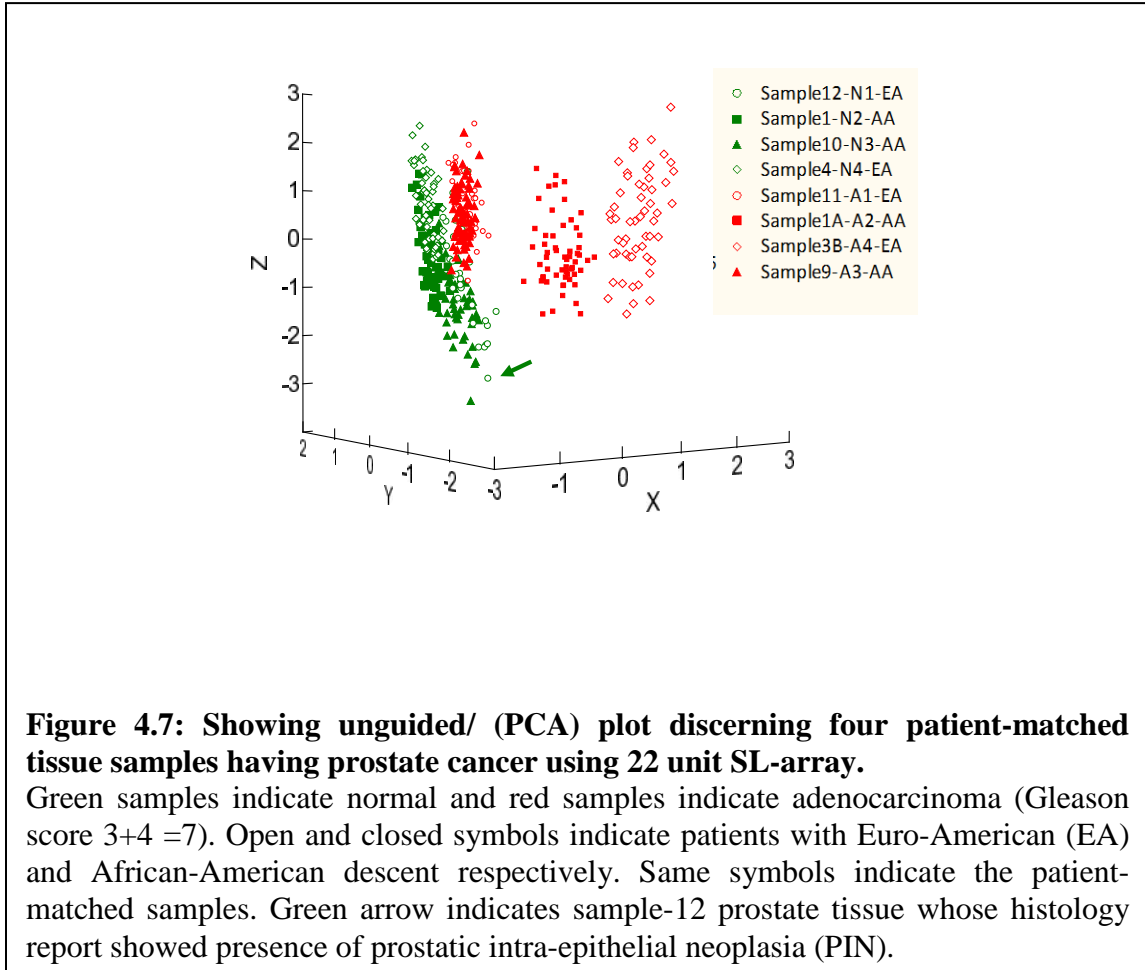


Figure 4.7: Showing unguided/ (PCA) plot discerning four patient-matched tissue samples having prostate cancer using 22 unit SL-array.

Green samples indicate normal and red samples indicate adenocarcinoma (Gleason score 3+4 =7). Open and closed symbols indicate patients with Euro-American (EA) and African-American descent respectively. Same symbols indicate the patient-matched samples. Green arrow indicates sample-12 prostate tissue whose histology report showed presence of prostatic intra-epithelial neoplasia (PIN).

The previously used SLs (20 in extended lectin array) and two new SLs, found from newer library screenings (i.e. SL15, SL16), were used together to make a 22-SL array that was employed to distinguish between two African American (AA, denoted by solid color) and two Euro American (EA denoted by open color) prostate tissue samples. (Ref: APPENDIX B).

Each tissue sample was patient-matched, making i.e. eight samples, whose histology was first confirmed, were pulverized and proteins were extracted from them. These

protein samples were then tested using the SL array. **Figure 4.7** illustrates an unguided 3-D classification. The guided (LDA) 8-class model was >99% accurate at predicting cancer (red) from normal tissue (green). The most useful chemical tool which correlated with statistical separation of healthy and adenocarcinoma samples, was phenyl ring gradation. This was true since SL9, SL7 and SL11-Y4F were the most useful SLs and all are having large number of phenyl rings in SL sequence.

In **Figure 4.7**, all the four red highlighted samples were histologically confirmed to have moderately differentiated adenocarcinoma and were positive for elevated PSA after surgery. Patient sample 9 and 11 (red▲ and ○) had leftover microscopic primary tumors at the surgical margins at the primary tumor site, whereas patient samples 1A and 3B (red■ and ◇) had clear surgical margins at the primary tumor site after surgery. This indicated the utility of SL array as prognostic tool by indicating cancer even when it became invisible to the surgeon. Green samples were patient-matched normal prostate tissue, except for patient sample 12 (green ○, indicated by green arrow) which was low-grade prostatic intraepithelial neoplasia (PIN) with no adenocarcinoma. As observed, it grouped a bit closer to the healthy region, needing further investigation where more tissue samples containing PIN can be investigated and checked how they classify.

In the future work, the proteins being expressed by these tissue samples, can also be investigated. For example, fucosyltransferase (FUT8) catalyzes glycan transfer of α 1.6-fucose (core fucose) to glycoproteins. It is known in the literature that both core-fucosylation and FUT8 expression are positively correlated to prostate-tissue malignancy.¹⁰⁻¹¹ The donor and acceptor substrates for FUT8 are GDP-fucose and a biantennary N-glycan, respectively. One of the ways to test for FUT8 activity is by

monitoring enzymatic activity by radiolabeling the oligosaccharide as the acceptor substrate. Literature suggests a convenient and sensitive assay method using 2-aminopyridine (PA)-labeled asialo-, agalacto-biantennary sugar chain as an acceptor substrate. N-Linked sugar chains can be prepared in relatively large amounts from egg yolk, and after the removal of sialic acid and galactose residues by glycosidases, the sugar chain can be labeled with PA and analyzed by HPLC.¹²

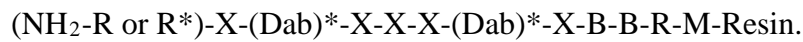
Thus making use of a 22-unit SL array, healthy and cancerous samples were distinguished with >99% accuracy. These initial results demonstrate that SL-array not only discriminates between cell types effectively in *in vitro* cell line models but also in tissue samples. This promising data suggests that the development of the array for clinical utility is possible.

4.5 DESIGNING FUTURE SL TARGETS FOR COLON AND PROSTATE CANCER

As of now the ability of SL array, not only to discriminate between healthy and cancerous cell lines but also between different cancer types has been discussed. In the beginning of this chapter, a possible chemical basis for SLs to distinguish between colon and prostate malignancies has also been discussed. The SLs that could bind to one type of carcinoma over the other, are termed as SL ‘hits’. In the light of these events, it is required to investigate the structural basis of the SL ‘hits’ which enables them towards this ability of “tissue-typing”.

To do so, some new SLs sequences were invented by interaction of different peptide-boronic acid based SL libraries to tissue specific targets. For e.g. two libraries (2-BA and 3-BA) were designed where the numbers indicate the total number of the phenyl-boronic acid (BA) attachments per SL sequence.

Each SL polymeric bead consisted of multiple copies of the same peptide-BA sequence. The difference between the two is that 2-BA library, instead of having a third phenyl-boronic acid has a free $-NH_2$ at the N-terminus. In order to keep the N-terminal free, it was protected using $(BOC)_2O$ anhydride (TFA sensitive). Using $(BOC)_2O$ protection, enabled orthogonal cleavage of ivDde (basic de-protection), The ivDde was the protecting group on the $-NH_2$ side chains of (Dab), which were removed for the subsequent attachment of phenyl boronic acid (BA) on the diamino butanoic acid. Unprotected N-terminal $-NH_2$, causes attachment of a third phenyl boronic acid residue on it. The rest of the peptide sequence for both libraries is generic:



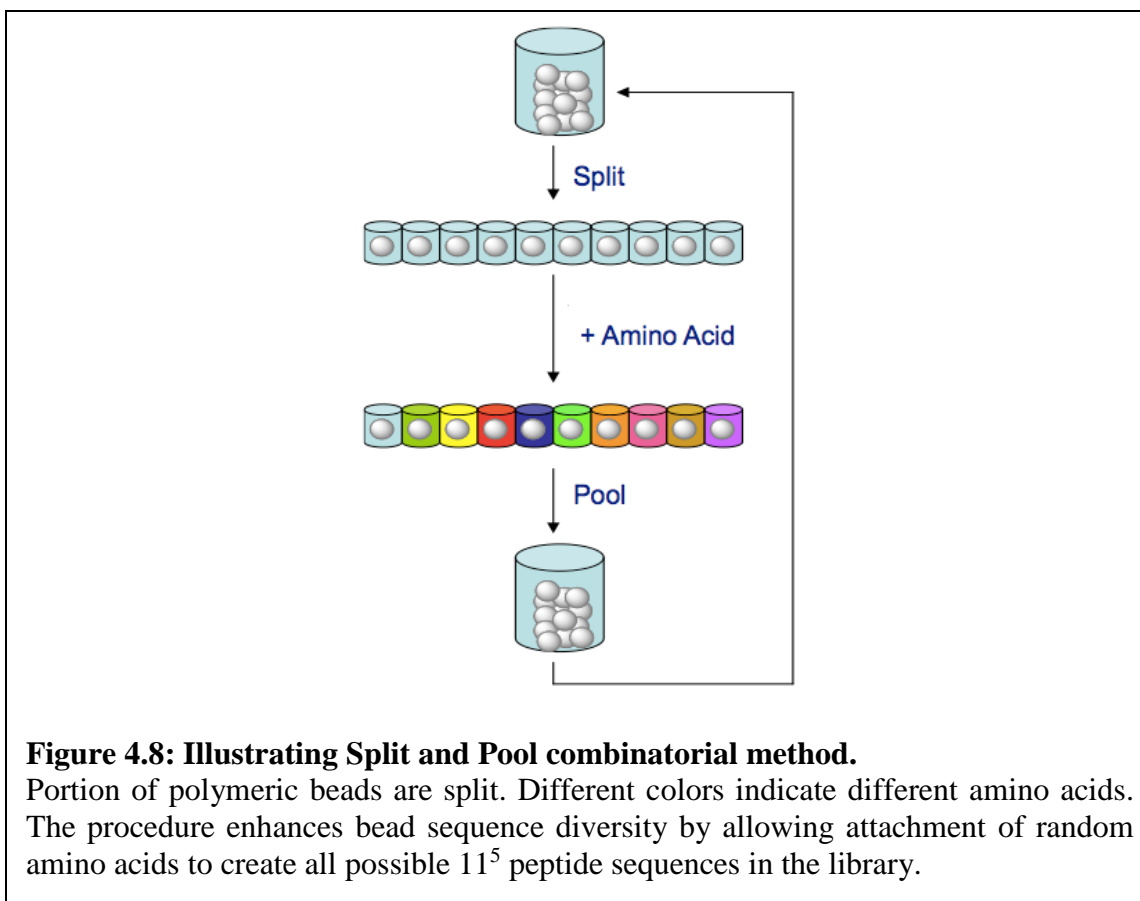
Here, phenyl-boronic acid (BA) moieties are fixed and denoted by * while R^* and $(Dab)^*$ denoted 2-phenyl-boronic acid functionalized R and Dab, respectively. B represented β -alanine and acts as spacer since it is insensitive to Edman degradation. Methionine (M) is added on to C-terminus to avail orthogonal cleavage of the whole peptide using CNBr while X is any one of the 11 random amino acids (A, N, Q, G, L, F, S, T, Y, V and R). R is the only positively charged amino acid, which is incorporated because its guanidinium charge assisted in MS/MS sequencing and creation of salt bridges to sialic acid as well. The standard Fmoc strategy was used for peptide synthesis.

Previous lab members established this generic library design to contain the greatest bead sequence diversity at the same time also maintaining the ease of sequencing while using MS/MS or Edman techniques. Sequence diversity of each SL bead is introduced by five randomized amino acid spot (X) in the peptide sequence. Since there are 11 different amino acid choices, it would mean 11^5 sequence diversity of each SL bead.

Both the libraries were devoid of C-terminal acylation because that hindered with sequencing. Dab was used in place of Lysine because the shorter side chain of the former assists to reduce background binding and increase interactions between the glycoproteins and the peptide backbone. (Dab)* moieties also were fixed to further aid in MS/MS sequencing. The randomized amino acids were coupled using a split-and-pool method as described in **Figure 4.8**. To do so, the polymeric resin was split into 11 tubes and a different amino acid is added to each tube. After the amino acid was coupled, all of the resin is combined again for a collective de-protection, before splitting again.

The remaining nine amino acids (out of 20), i.e. M, C, I, P, H, D, E, K and W, were excluded from the randomized positions. Methionine was not repeated in the rest of the peptide sequence and cysteine was not used in the randomized amino acid positions to eliminate the chance of cutting the peptide sequence into pieces due to CNBr cleavage. Isoleucine was omitted to eliminate the chance of disambiguation with leucine when sequencing for the unknown SLs. Proline was excluded to avoid having 'kinks'. Histidine, glutamic acid, aspartic acid, and lysine were avoided to reduce interference of charge during MS/MS. Tryptophan was not included because of its big and bulky side chain. Glutamate and Aspartate were not used to remove molecular weight disambiguation with their corresponding amine counter parts.

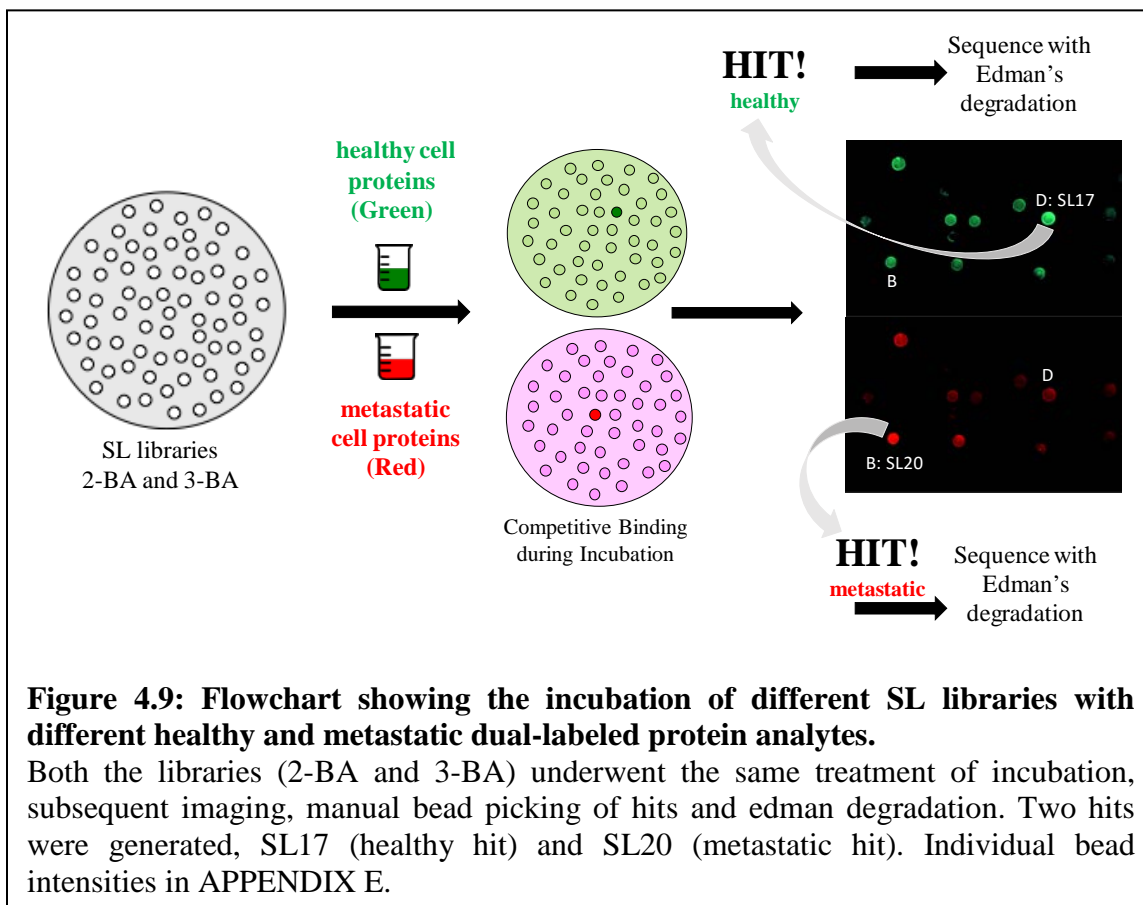
In order to identify SLs that could bind differently to non-cancerous versus cancerous samples, the protein extracts (from healthy and cancerous samples) were fluorescently labeled with two different colored tags, then combined together and interacted with the SL libraries.



This procedure of competing non-cancerous and cancerous analytes to identify SL beads that are selective hits for cancer: is termed as a DUAL DYE LIBRARY SCREENING (see **Figure 4.9**). The SL that bound specifically to healthy and cancerous proteins was termed a ‘healthy hit’ and ‘cancerous hit’ respectively. After these SL ‘hits’ were manually isolated under the microscope, the individual SL beads were prepared for peptide sequencing. In order to look out for colon as well as prostate specific hits, different healthy and metastatic protein targets that were used against the two libraries, and are below:

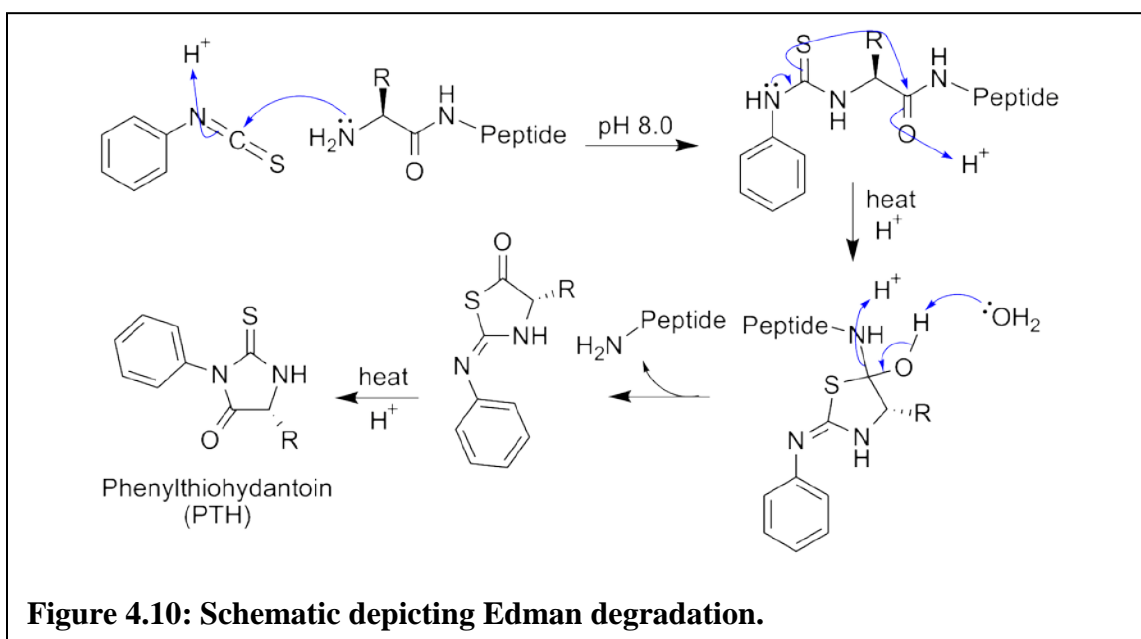
1. Colon cell secreted proteins (CCD 841 CoN versus LoVo)
2. Prostate cell secreted proteins (RWPE-1 versus DU-145) and
3. Prostate patient-matched tissue proteins (including race related diversity)

The healthy protein analyte was labeled with FITC (green) and metastatic analyte was labeled with rhodamine (red). All the three different targets were separately incubated with 3-BA and 2-BA SL libraries. **Figure 4.9** shows a typical workflow of incubation. This process generated tissue-specific, healthy and metastatic hits.



To ensure that there was no specific dye-based discrepancy while finding SL hits, reverse dual-dye screening was carried out as well, where healthy protein analyte was labeled with rhodamine (red) and metastatic analyte was labeled with FITC (green). It was noted that there were no impartialities related to dye charge towards affinity for a specific analyte. There are advantages to using secreted proteins rather than proteins extracted from cell membranes because once a membrane-bound protein is extracted, the native protein tends to lose its integral structure. Using cell-secreted proteins rather than

cell membrane isolated proteins also tests the ‘sensitivity’ of SLs (because the actual protein of interest is low in concentration in a secreted protein mixture).



The bound protein was washed away from the SL (by alternatively washing the resin using 70% TFA in H₂O and 0.1 M NaOH). To sequence hits from library screenings against different protein analytes, Edman degradation was used. Edman peptide sequencing, involves immobilization of resin bound peptide followed by removal of one amino acid at a time and subsequent analysis using high-performance liquid chromatography (HPLC).

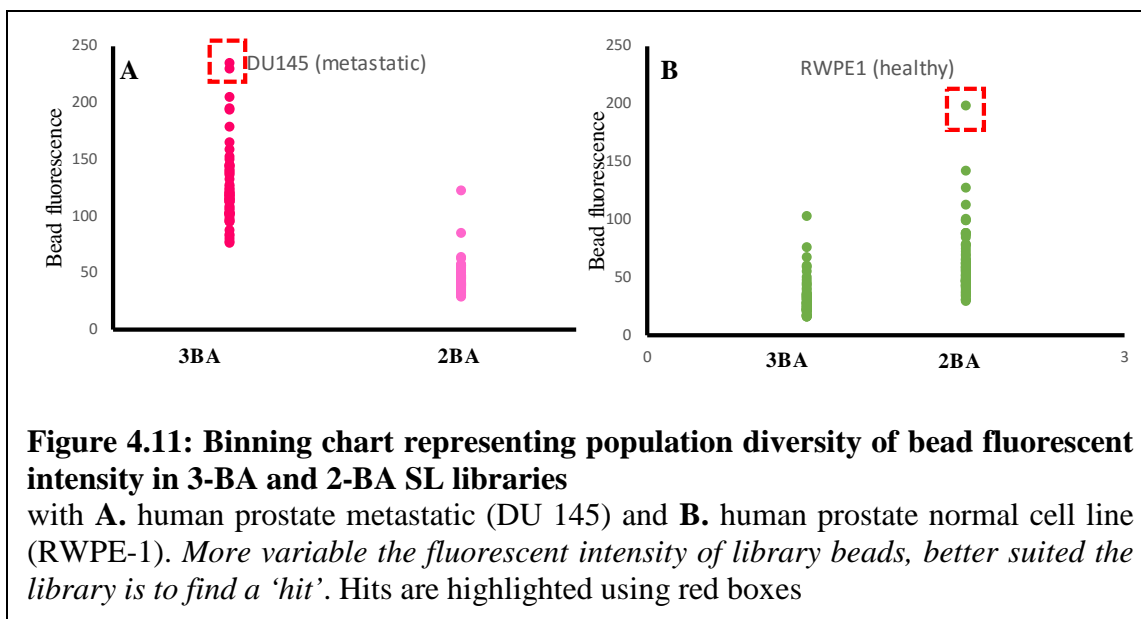
In the first step of Edman Degradation, the N-terminal –NH₂ reacted with phenyl isothiocyanate to form a phenyl thiocarbonyl-peptide derivative (**Figure 4.10**). If the N-terminus is acylated, this reaction cannot take place, making sequencing obsolete.

The presence of phenyl-boronic acid causes difficulty at obtaining peptide sequences using MALDI due to H-bonding with water causing messy signal with great signal-to-noise ratio. To avoid this situation several techniques for removing the PBA moiety from

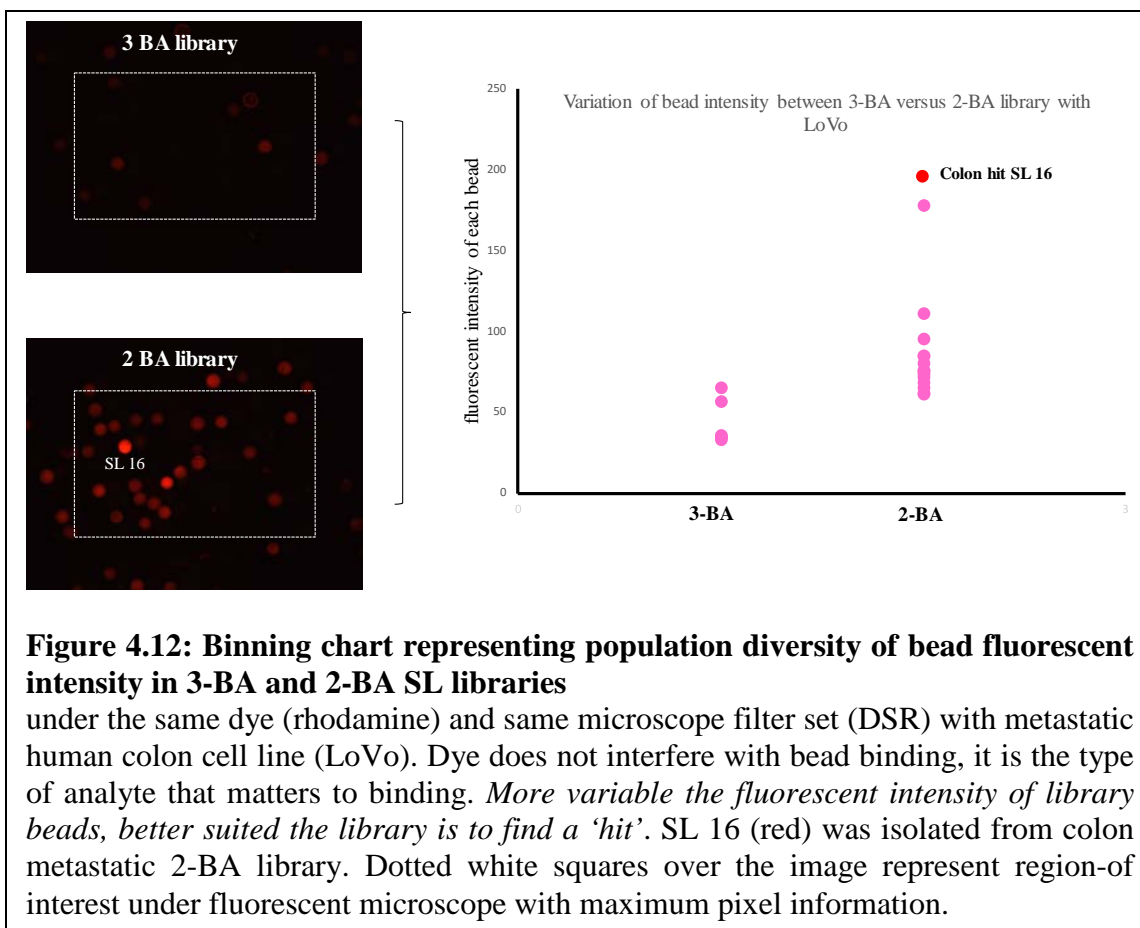
the SL prior to MS-based sequencing, were used. One of them was treating resin bound SL with hydrogen peroxide in water to enable oxidation of BA to phenol.¹³ Since Edman method was used and the location of boronic acid was already known, there was no need to cleave off boronic acids prior to sending them for sequencing.

The 2-BA library peptide hits were easier to sequence than 3-BA hits. In all of the hits procured from 3-BA library, the N-terminal R, was inconclusive from the chromatogram, probably due to signal interference from the attached third boronic acid. Nevertheless, the position of the N-terminal R was fixed in the library design thus confirming the attachment of third BA there. To verify the presence of R on the N-terminus, while making the SL library, R was manually coupled and its complete attachment was verified through a subsequent ninhydrin test. It is also hard to distinguish the Dab positions when referring to the sequencing chromatograms but it is similar to the last AA position. It is most important to look for changes from cycle to cycle within the SL sequence. The absolute quantities of the AAs in a single cycle is not as important due to the fact that different AAs have different background levels.

Proteins concentrated from normal cell lines of prostate (RWPE-1) were labeled green using FITC and proteins from metastatic non-androgenic, non-PSA expressive (DU 145) were labeled red with rhodamine. Competitive binding dual dye tests were done on two libraries one involving an extra boronic acid at the N-terminus, thus having three boronic acid residues on SLs(3-BA) and the other having free N-terminus (2-BA). 3-BA libraries consequently have one extra phenyl ring in SL sequences and both the libraries were tested on prostate and colon cell proteins.

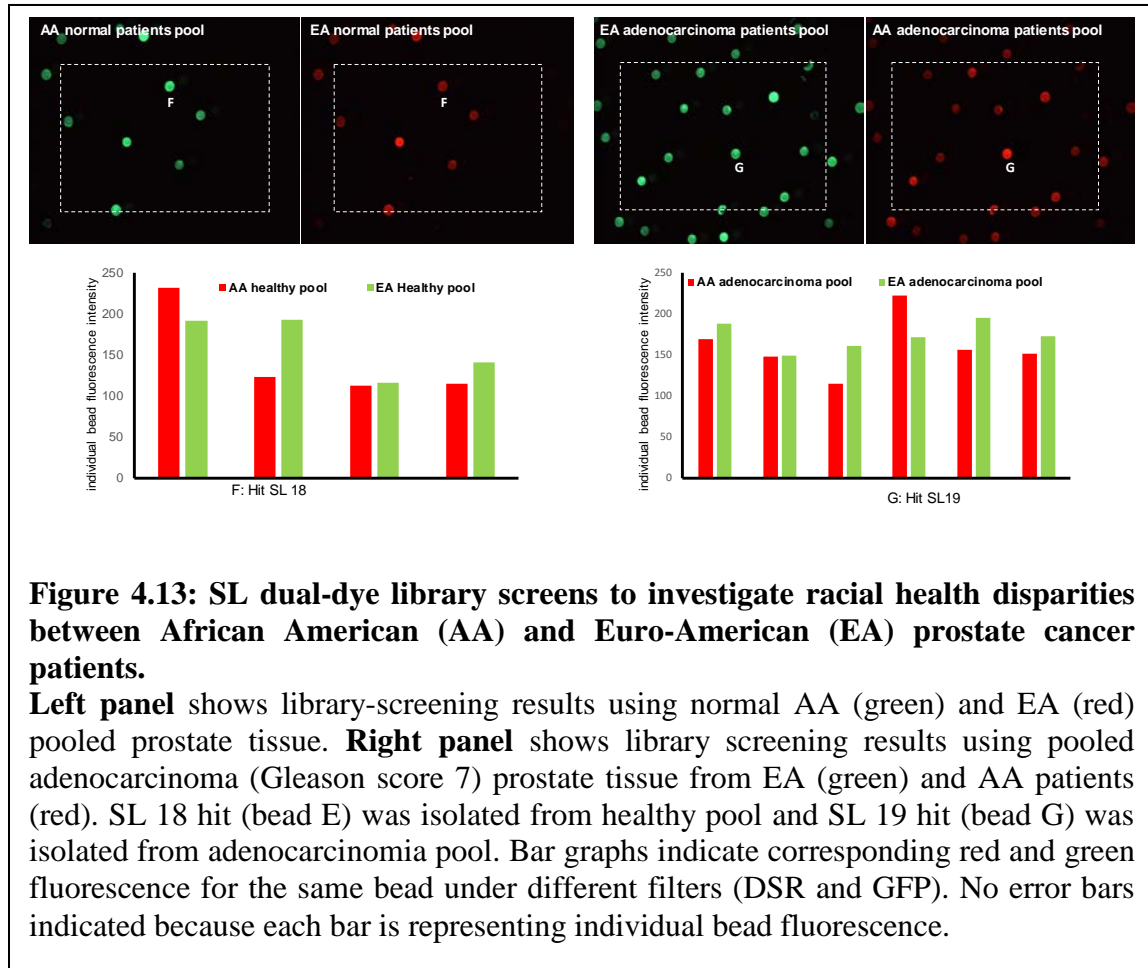


It was observed that, *more variable the fluorescent intensity of library beads, better suited is the library to find a 'hit'.* In **Figure 4.11**, a binning chart with bead population diversity has been shown by plotting fluorescent intensity of each bead in 3-BA and 2-BA SL libraries. Where, **Figure 4.11A** is SL library against human prostate metastatic (DU 145) and **Figure 4.11B** against human prostate normal cell line (RWPE-1). After calculating the fluorescent intensity of each individual bead in the library, it was observed (**Figure 4.11A**) that 3-BA library had greater binding variability with prostate cancer cells and 2-BA had better diversity with normal prostate cell lines (**Figure 4.11B**). This might be due to extra phenyl rings in 3-BA library, causing greater CH- π type interaction with metastatic prostate proteins (which have greater fucosylation expression than normal prostate cell lines).



Similarly, normal cell colon cell lines (CCD 841 CON) and metastatic (LoVo) were also labeled green and red, respectively. **Figure 4.12** shows a binning chart representing population diversity of bead fluorescent intensity in 3-BA and 2-BA SL libraries under the same dye (rhodamine) and same microscope filter set (DSR), with human colon metastatic (LoVo). More variable the fluorescent intensity of library beads, better suited the library is to find a 'hit'. This concludes that 2-BA library had better fluorescent signal diversity with metastatic colon cell proteins. Since increased virulence in colon cells is understood by change in sialylation, using a library with lesser number of phenyl rings (e.g. 2-BA) proves useful.

In **Figure 4.12** both libraries are interacting with proteins from metastatic colon cell line, thus showing that type of fluorescent dye used does not interfere with the binding affinity of SL hit with the analyte. SL 16 was isolated from colon metastatic 2-BA library. Dotted white squares over the image represent region-of interest under fluorescent microscope with maximum pixel information.



SL dual-dye library was also used to further investigate genetic health disparities amongst African American (AA) and Euro American (EA) prostate cancer patients (see **Figure 4.13**). Left pool shows proteins pool using normal prostate tissue of AA patients (green) and EA patients (red):

Table 4.2: Showing Malignancy specific SL ‘hits’.

Hits from 2BA and 3BA library screens using secreted proteins from human prostate or human colon cell lines, or membrane extracts from human prostate tissue samples (patient-matched normal and adenocarcinoma GS7), illustrating amino acid composite trends, e.g. charged amino acids, number of aromatic rings and boronic acids. (EA+AA): stands for Euro-American plus African-American patient sample. (2xAA): two African-American patient sample pool.

Name	#BA	Sequence	#(R)	#Ph rings	Polar S,T,N,Q	Library	Proteins from
SL15	2	NH ₂ -RT(Dab)*RGG(Dab)*TBBERM	3	2	2	colon cell	healthy colon
Bead B	2	NH ₂ -RS(Dab)*NLS(Dab)*QBBERM	2	2	4	colon cell	healthy colon
SL17	2	NH ₂ -RA(Dab)*NAQ(Dab)*NBBERM	2	2	3	prostate tissue	healthy prostate (EA+AA)
SL18	2	NH ₂ -RN(Dab)*VLS(Dab)*GBBERM	2	2	2	prostate tissue	healthy prostate (2XAA)
SL16	3	R*R(Dab)*AYR(Dab)*YBBERM	4	5	0	colon cell	metastatic colon
SL19	2	NH ₂ -RY(Dab)*RYF(Dab)*LBBERM	3	5	0	prostate tissue	metastatic prostate (2XAA)
SL20	2	NH ₂ -RY(Dab)*YYY(Dab)*RBBERM	3	6	0	prostate tissue	metastatic prostate (EA+AA)
Bead 1	3	R*G(Dab)*SGR(Dab)*QBBERM	3	3	1	prostate cell	metastatic prostate
Bead 3	2	NH ₂ -RY(Dab)*FFL(Dab)*RBBERM	3	5	0	prostate cell	metastatic prostate

SL hit 18 was found and isolated from healthy tissue pool and could be used for investigation of race-related normal health disparities in future. Right pool shows proteins pooled using adenocarcinoma (Gleason score 7) prostate tissue of EA patients (green) and AA patients (red): SL hit 19 was found and is unique in nature because it could discriminate between complex tissue protein mixtures of different cancer patients. This can be used as a future target to investigate personalized vectors for chemotherapeutic targets. Sequences in **Table 4.2**.

After sequencing all the hits (shown in **Table 4.2**) the hits, SL15-20 was designed, along with bead B, 1 and 3 (all 9 of these are colon and prostate specific SLs). The sequence analyses of these SL-hits surmised that specific amino acids play an important role to assist in “tissue-typing” the array. Several different glycoprotein sources were screened including those from secreted cell lines and ones from pooled clinical tissue samples, producing a new series of SL hits exhibiting some interesting trends. It was observed (**Table 4.2**) that SL hits for normal prostate and colon glycoproteins, had a higher ratio of polar amino acids e.g. S, T, N and Q (bold numerical values). SL hits for metastatic colon cancer had a greater number of arginine residues (R) (light grey). SL hits for metastatic prostate cancer had a greater number of aromatic rings (from amino acids like F and Y and/or from phenyl boronic acid) in the isolated SL sequences (dark grey).

4.6 CONCLUSIONS

The beginning of this chapter established some of the chemical targets of proteins/glycoproteins that bind to SLs to cause cancer virulence based discrimination. It has been shown that by using SLs sequences with greater number of Arginine residues and lesser number of phenyl groups, different colon cells lines can be discerned and a varied sialic acid concentration. It has been shown that varied sialic acid concentration is one of the statistical basis of discrimination among colon cell lines by the array.

However just the opposite is true for accurately discerning each prostate cell line where the SLs sequences containing greater number of phenyl rings and lesser number of Arginine residues, play an important role. Since the SLs bind through guanidinium groups to sialic acid residues, both in colon and prostate cell lines, it was proved that

stripping away of the sialic acid residues takes away the SL array's unique ability of colon cancer discrimination.

SL array uses positively charged SLs for pre-organization of negatively charged cancer-associated glycans (e.g. sLe^a and sLe^x) by increasing the overall binding of charged pieces. As the model derived from positively charged SLs retains excellent 'tightness/precision', it indicates that these positively charged SLs provide a microenvironment necessary for accurate binding. Besides that, the SL array uses SLs with N-terminal phenyl boronic acids to covalently target diols of cancer glycans like Le^a, sLe^a and sLe^x. These SLs also enhance CH- π type interaction through the large number of phenyl rings associated with them. One such SL (SL11) having high number of phenyl rings bound exceptionally well to fucose. The results have suggested that boronic acid residues are important to discriminate sLe^a, sLe^x and Le^a.

Guiding structural epitopes in SL peptide sequences are discussed to see if they can be used selectively to target colon and prostate cancers. In light of this, two new libraries were synthesized (3-BA and 2-BA). These libraries were screening against colon and prostate specific analytes (both cells and tissues). After sequencing the hits, SL15-20 was designed, along with bead B, 1 and 3 (all colon and prostate specific SLs).

The sequence analyses of these SL-hits surmised that specific amino acids play an important role to assist in "tissue-typing" the array. Polar amino acids (S, T, N and Q) target healthy analytes while R targets colon cancer through the tissue's proclivity to over-express sialylation and sialylated cancer-glycans. Increasing the number of phenyl rings (by increasing phenyl ring containing amino acids (F & Y), or by incorporating

phenyl-containing boronic acids), the SL-hits achieve an increased core-fucosylation of prostate cancer.

In total, these results validate our ability to differentiate structurally similar cancer associated glycans with high accuracy using a small, cross-reactive SL Array. At times it indicates the utility of SL array as prognostic tool by indicating cancer even when it became invisible to the surgeon.

4.7 FUTURE DIRECTIONS

The glycan motifs that are targeted by the SLs require evaluation. One way to approach the analysis of glycan would be to isolate and enrich the glycoprotein sequences that bind to the SLs (using hydrazide columns that can capture N-linked glycans) and then investigate them. A methodology is required that ensures the protein integrity is not compromised during the isolation of glycoprotein from the column. Using circular dichroism spectroscopy to evaluate protein stability as a function of added denaturants is suggested in future experimentation.

One of the challenges is to establish the stoichiometry of the SL-glycan interaction. This is because the SL works due to its multivalent display of peptides. Singular SL peptide sequence has a negligible binding ($\gg 10 \mu\text{M}$), therefore the peptide needs to be on the polymeric resin bead during interaction. One way to approach this question is to develop Biotinylated SLs on the beads, detach them from beads and then interact them with streptavidin platforms. This allows for calculation the number of binding events once the streptavidin-bound SL tetramer is interacted with a purified glycoprotein or a CAG of choice.

Currently, work on patient blood is underway, which is the next step to make headways into converting SL-array to a plausible diagnostic. Investigating the optimal blood sample size required to determine the virulence, is paramount considering that the K_d of SL beads is in the similar range as that of natural lectins.

4.8 EXPERIMENTAL PROCEDURES

Details on the complete procedure for SL synthesis, incubation of SLs with fluorescently tagged analytes, imaging of SL beads, data acquisition and statistical programs used for image analysis are listed under the experimental section after chapter 2.

Details on complete procedure for human colon and prostate cell-culture, protein extraction from cell membrane as well as concentration of cell-secreted proteins from cell growth media, protein quantification, fluorescently tagging and incubation with SLs are listed under the experimental section after chapter 3.

Procedure for Sialic Acid Quantitation

The Sialic Acid Quantitation Kit (Sigma Aldrich®) (Catalog Number SIALICQ) was used to provide a rapid and accurate determination of total N-acetylneuraminic acid, also known as sialic acid. The kit can provide a simple and reliable method for the quantitation of total sialic acid concentration in sera and specific glycoconjugates to study disease states and virulence potentials in many pathogens. The kit contains all reagents, including a proprietary α -(2→3, 6, 8, 9)-neuraminidase, which was required to quantitate the amount of sialic acid present in cell surface glycoproteins, polysialic acids, and capsular polysaccharides consisting of only polysialic acid. One unit of the neuraminidase enzyme catalyzes the release of 1 μ mole of 4-methylumbelliferone from 2'-

(4-methylumbelliferyl) alpha-D-N-acetylneuramin acid per minute. This kit is unsuitable for use with glycolipids or gangliosides. The α -(2→3, 6, 8, 9)-Neuraminidase cleaves all sialic acid linkages, including α -2→8 and α -2→9 linkages, as well as branched sialic acid. Variants of sialic acid such as N-glycolyl or O-acetylneuraminic acid were cleaved as well. Digestion with α -(2→3, 6, 8, 9)-neuraminidase generated total free sialic acid from glycoconjugates without incurring losses that are typically associated with chemical hydrolysis. The millimolar extinction coefficient of β -NADH is 6.22 at 340nm. Therefore, each nmole of sialic acid will cause a drop of 0.00622 absorbance units in a 1mL reaction, a quantity that can be reliably measured on a digital spectrophotometer. If the extinction coefficient level is exceeded (using too much protein), β -NADH will be exhausted from the reaction mix. In these experiments, 100 μ L of 0.01 mg/ml of crude cell-secreted protein sample was used.

To make Tris Reaction Buffer - Diluted the 1M Tris-HCl (Catalog Number T3946) 40-fold with distilled water to make a 25mM solution. The pH of Tris Reaction buffer is 7.5.

To make β -NADH Solution - Just prior to use, added 640 μ L of Tris Reaction Buffer to one 5mg vial of β -NADH (Catalog Number N8129) to make a 0.01M solution ($A_{340} = 62.2$).

For optimal performance, freshly prepared β -NADH solutions should be used. Prepared solutions should be discarded if the A_{340} drops 20% or the solution turns yellow.

Proceeded by dissolving the glycoprotein or polysialic acid in 40 μ L of distilled water and then add 1 μ L of Sialidase Buffer (250mM sodium phosphate, pH 5.0, Catalog Number S7189 ready-to-use). Added 1mL of α -(2→3, 6, 8, 9)-Neuraminidase and incubated for at

least 3-hours in a 37°C water bath. Added 930µL of Tris Reaction buffer to the tube. Pipetted the reaction mixture into a cuvette and blank the spectrophotometer. Added 20mL of the β-NADH Solution and mix by inversion several times. Read and recorded the initial β-NADH absorbance (triplicate). Initial A_{340} should read ~1.25. The absorbance according to Beer's law directly correlates to the path length of the cuvette. Therefore, if using 1mL cuvette, A_{340} should read out 1.25, but if using 0.5 mL cuvette, A_{340} should read out ~0.625. Returned the reaction mixture to the original Eppendorf tube. Added 1µL of N-Acetylneuraminic Acid Aldolase (Catalog Number A0849, ready-to-use) and 1µL of Lactic Dehydrogenase (Catalog Number L9889, ready-to-use) and mixed by inversion several times. Incubated in a 37°C water bath for a minimum of 1-hour. This can go up to 3-hours. Pipetted the reaction mixture back into the cuvette. Read and recorded the final A_{340} . Calculated the nmoles of sialic acid using the formula:

$$\text{nano moles of total cleaved sialic acid} = \{(A_{340} \text{ Initial} - A_{340} \text{ Final}) \times 1,000\} / 6.22$$

To prevent any carryover of lactic dehydrogenase, which might result in a low initial absorbance reading, always wash the cuvette thoroughly between assays. Positive glycoprotein control of Bovine Fetuin (Catalog Number F4301) sample was used to perform positive controls before proceeding with analysis of unknowns. When the entire sample of Bovine Fetuin was digested with α-(2→3, 6, 8, 9)-neuraminidase, the amount of sialic acid should be ~48nmoles.

Labeling protocol for secreted proteins after subjection to sialidase enzyme leach

After sialic acid quantitation was completed using UV-spectroscope, each tube contained intact protein samples from each cell line in Tris reaction buffer along with total free sialic acid residues, 1µM quantity of each sialidase, aldolase and dehydrogenase

enzymes. To get rid of the total free sialic acid from the leftover intact protein sample and consequently tag the leftover proteins, we added 10 μ L of FITC dissolved in 10:1 ratio of PBS: DMF. To make FITC solution, weighed out 2.8 mg of FITC, dissolved it in 100 μ L of 99% DMF, and added 900 μ L PBS. 10 μ L of this FITC solution was added to 1 ml leftover intact protein sample (with a concentration of 0.01mg/ml and volume of 100 μ L in Tris reaction buffer). Therefore, the protein: FITC ratio was 10:1. This protein solution was covered and tumbled at 37°C for 1-hour. Later, the protein solution was centrifuged at 4000 RPM using Amicon[®] 10-KDa centrifuge tubes. The buffer was exchanged from Tris buffer (pH 7.5) to PBS (pH 7.3) using the buffer exchange tube, by addition of 1.5 mL of PBS. The leftover intact-labeled protein was collected and a BCA assay was run before using the labeled protein for further experimentation.

General Screening Procedure for Biotinylated-Cancer Associated Glycans

Approximately 2mg of SL polymeric beads (from freezer) were incubated in Phosphate Buffer Saline (PBS) without glycerol for 10 minutes at RT. Then the SL beads were pre-blocked with 1mg/mL BSA and 10% glycerol containing PBS (Screening Buffer) for 15-minutes at RT. The resin beads were later incubated with 0.01 mg/mL biotinylated-glycans for 12-19 hours at 37°C. Just 2-hours before the completion of the incubation, 6 μ L of 100nmol/L FITC-streptavidin was added and tumbling was continued at 37°C. After washing with PBS three times, the beads were imaged using a fluorescent microscope and analyzed as previously described. Assuming that 2mg of 300-micron tentagel macrobeads contain roughly 100 beads and each bead has roughly 0.004 μ mole capacity. Further, assuming that each SL copy on the bead can bind to five glycans (this was done by assuming that each SL has a maximum of +5 charge), it was calculated that

2mg of beads have a bead capacity of 0.002mmoles. Therefore, 0.01mg/mL of biotinylated-glycans were used instead of 0.1mg/mL because even by using 0.01mg/mL concentration, the glycan concentration was still roughly 10-times higher than concentration of the number of binding sites calculated per peptide per polymeric bead. The reduction in the fold-concentration of biotinylated glycans was done because at the time a small quantity was available.

Frozen Prostate Tissue Procurement, Handling and Fixation

Obtained eight pairs of patient-matched frozen prostate sample (Courtesy: Ms. Ella Weinkle at Prostate tissue core facility, CLS 508). A confirmed histology report was provided (Courtesy: Dr. Loulia Chatzistamou, Instrumentation Resource Facility Building 1 room # B-60, USC School of Medicine). The paraffin tissue blocks needed to be made (courtesy: Benny Davidson IRF, USC School of Medicine) for sectioning. Before sending the tissues for sectioning, each frozen tissue sample required fixing using neutral buffered formaldehyde (NBF) for 24-hours.

Preparation of neutral buffered formaldehyde (NBF)- Mixed 100ml formalin (37-40% stock solution) with 900ml deionized water, 4g/L NaH_2PO_4 (monobasic) and 6.5g/L Na_2HPO_4 (dibasic and anhydrous) at RT. The stock was stored at 4°C. It is suggested to prepare fresh NBF solution each time, to maintain 10% w/v ratio of formaldehyde gas: water. Do not use paraformaldehyde because, it is not a fixative.

Alternatively, paraformaldehyde can be broken down into its basic building block formaldehyde and be used for the same purpose. This could be done by heating or basic conditions until paraformaldehyde became solubilized in order to act as a fixative.

To slice the frozen tissue, -20°C microtome at PSC 621 (courtesy: Dr. Hexin Chen) was used, so that the un-fixed frozen tissue does not thaw during the slicing and NBF fixing. The tissue was sliced using a scalpel and immediately placed in a 15-mL falcon tube containing 5mL of NBF at RT. The other half of each tissue was saved in the original tubes and carried using dry ice. Unused tissue, required for actual experimentation was stored under -80°C conditions (Dr. Mythreye Karthikeyan Lab).

Almost 50% of prostate tissue from each slice was used for fixing and subsequent paraffin-blocking and sectioning to determine its histology. Each piece of tissue makes one paraffin block. This was because of the nature of prostate cancer, as it exists in small discrete areas in the tissue. The number of sections mounted per slide were two and were further H&E stained by IRF for evaluation. (Leftover paraffin blocks were stored in GSRC435 refrigerator at 4°C).

Procedure for Prostate Tissue Pulverization Using Liquid Nitrogen

All the pulverization was handled under the sterilized fume-hood cabinet. A liquid 4L Nitrogen tank was set nearby. Tin foil was applied all over the fume-hood to ensure a safer clean up after working with tissue (potential biohazard). All the necessary supplies were kept under the hood after spraying with 70% ethanol mixture (sharps container, box of ice, autoclaved-pipette tips, 1000 µL pipette, towels/tampons, 16 G needles, mortar and pestle, lysis buffer with protease inhibitor). The mortar and pestle was pre-chilled by pouring in liquid nitrogen and allowing it to evaporate. The tissue was place along with a little amount of nitrogen; the tissue must remain in nitrogen at all the times. Before, pulverization by twisting motion, it was ensured that the tissue had completely frozen or attained a stone like form. Then it was lightly crushed by tapping from the top (this

needed to be done carefully, else the tissue piece could bounce out of the mortar. Used tap and turn technique for pulverization while adding small aliquots of nitrogen (extreme care!).

Scooped up all the pulverized tissue using the scalpel and transferred into an Eppendorf containing 1mL of cold lysis buffer with protease inhibitors. The crushed sample was passed through the 16 G needle several times, to break any larger chunks of tissues. This step was done to homogenize the tissue slurry for greater action of lysis buffer. The sample was rested in the buffer, meanwhile the next tissue piece was pulverized. Mortar and pestle were bleached each time before working with a new tissue to avoid contaminations.

The tissue slurry in the lysis buffer was centrifuged using 4°C centrifuge for 20 minutes. The supernatant was removed carefully leaving the tissue pellet behind and the supernatant was further used for membrane protein extraction.

Prostate tissue may contain a layer of adipose tissue, which floats as fat layer on the top of the supernatant during the centrifuge. In that case, carefully position the pipette to only collect the middle layer of supernatant.

Extraction of Protein from Colon Cell Membrane and Prostate Frozen Tissue and its Corresponding Fluorescent Labeling using FITC

The Qiagen® Plasma Membrane Protein Kit was used to extract plasma membrane proteins and glycoproteins from cells as well as tissues; this is because out of five commercially available membrane extraction kits, it gave the highest membrane protein yield and the second least amount of cytosolic protein contamination.¹⁴ The starting material for one fractionation procedure is 1×10^7 adherent cells. That corresponded to

four 75-cm² cell flasks at 70% confluence. Cells were incubated in a hypotonic lysis buffer, causing them to swell. After the addition of a mild detergent, the cells burst and the resulting cell suspension was homogenized by mechanical disruption using a needle and syringe. Intact cells, cell debris, nuclei and the major organelles were removed by centrifugation under 4°C temperature at 12000X g for 20 minutes. The resulting supernatant contained cytosolic proteins and microsomes (i.e. small vesicles (20–200 nm in diameter) formed from the endoplasmic reticulum, Golgi vesicles, and plasma membranes). In case of tissue samples, skip this step containing lysis buffer and centrifugation and resumed with the next step.

A ligand specific for commonly expressed antibody molecules on the plasma membrane was added to the supernatant. The ligand bound to the plasma membrane vesicles and the ligand–vesicle complexes were precipitated using silane coated magnetic beads that bound to the ligand. Silane coating prevents non-specific binding. The 1-4µm diameter of magnetic beads provided for enough surface area to volume ratio for the ligand to bind to the magnetic beads. The plasma membrane vesicles were captured on to the bead surface. After washing, plasma membrane vesicles were eluted under native conditions and the ligand remained bound to the beads. The membrane protein remained in elution buffer and could be stored in -20°C temperature until required.

The protein extract was centrifuged and buffer exchanged into carbonate buffer pH 9.8 using Amicon® 10-KDa spin tubes. Later, a BCA assay was performed to obtain the concentration of the protein extract. The required amount of protein was added to carbonate buffer to make up a total volume of 600 µL. Depending on which dye was being used to label, 0.07 µmol FITC or NHS-rhodamine was added and the solution was

incubated at 37°C and tumbled for 1-hour. After an hour, the labeled protein was spun down using a second Amicon® 10-KDa spin tubes at 4000X g for approximately 15 minutes. The buffer was exchanged to PBS at RT. The final labeled protein sat in PBS at pH of 7.3.

General Incubation Protocol of Labeled Proteins Extracted from Prostate Tissue

Approximately 2mg of library resin was tumbled in PBS twice at RT and then pre-incubated with 1% BSA in PBS containing 10% glycerol for 15-minutes at RT. This pre-incubation was performed, to reduce the nonspecific binding between the SL and proteins. The 1% BSA in PBS with 10% glycerol solution was then removed from the beads and 1mL of a 0.0002mg/mL solution of the fluorescently labeled prostate tissue extract in PBS was incubated with the beads for 12-hours at 37°C.

Procedure for (BOC)₂O Protection to Design 2-BA Library

The two designed libraries (2-BA and 3-BA) were both fixed position libraries. The 2-BA library had a free –NH₂ at the N-terminus. To keep the N-terminus –NH₂ free even during the ivDde de-protection (basic de-protection), the N-terminus was protected using (BOC)₂O protecting group. The ivDde were the protecting groups on the amino side chains of diamino butanoic acids, which were removed for the subsequent attachment of phenyl boronic acid on the diamino butanoic acid. Not protecting the N-terminal –NH₂, causes the attachment of a third phenyl boronic acid residue on the N-terminus. Correspondingly, the (BOC)₂O protection and avoiding of (BOC)₂O protection led to formation of 2-BA and 3-BA libraries respectively. Where “2” stood for two of the phenyl boronic acid (BA) attachments per SL sequence.

Calculate the amount of (BOC)₂O (Di-tert-butyl dicarbonate) and DIEA/DIPEA/Hünich's required. The ratio of tentagel beads (bead capacity 0.25g/mmol): (BOC)₂O: DIEA is 1eq: 15eq: 15eq. 15eq was used because (BOC)₂O and DIEA are quite cheap and the ratio of tentagel beads to bulkiest amino acid (R) was 1eq: 10eq.

After calculating for 1000mg of tentagel beads (bead capacity 0.25g/mmol) used 818.44mg of (BOC)₂O and 653.16μL DIEA. After transferring the tentagel beads with peptide (free -NH₂ group on N-terminus) into falcon tube poured in 6ml solution containing (BOC)₂O. Added approximately 30 molecular sieves (4X8 mesh size). Using a syringe under nitrogen gas environment added the required amount of DIEA. Quickly cap and tumble the tube for 3-hours. Be sure to vent the tube initially after 15-20 minutes. The beads were washed using DMF followed by MeOH and DMF again. To ensure complete (BOC)₂O coupling, performed ninhydrin test. The bead should be clear/ yellow tinge. If the bead is purple, repeat the coupling again before proceeding to ivDde de-protection step.

Library Screening for Dual Dye Analytes

The library screening for dual dye analytes followed the same protocol in terms of sample concentrations, be it for colon cell membrane extracts, prostate cell secreted proteins or prostate tissue proteins. Typically, healthy protein extract was labeled using FITC and cancerous protein extract was labeled using NHS-Rhodamine. 0.0001 mg of each labeled protein analyte was used to make a total concentration of 0.0002 mg/mL per tube. This dual labeled protein mix was then incubated with 2-5 mg of either of the

libraries for 12 hours. The standard wash protocol using PBSX 3 times and imaging under microscope was done.

4.9 REFERENCES

1. Dube, D. H.; Bertozzi, C. R. Glycans in cancer and inflammation, potential for therapeutics and diagnostics. *Nature Reviews* **2005**, *4*, 477-488.
2. Saitoh, O.; Wang, W. C.; Lotan, R.; Fukuda, M. Differential glycosylation and cell surface expression of lysosomal membrane glycoproteins in sublines of a human colon cancer exhibiting distinct metastatic potentials. *J. Biol. Chem.* **1992**, *267*, 5700-5711.
3. Garcia, d. A. R.; Nakamura, C.V.; de Souza, W.; Morgado-Díaz, J. A. Differential expression of sialic acid and N-acetylgalactosamine residues on the cell surface of intestinal epithelial cells according to normal or metastatic potential. *J. Histochem. Cytochem.* **2004**, *52*, 629-640.
4. Calvin, R.; Justus, N. L.; Ruiz-Echevarria, M.; Yang, L.V. In vitro Cell Migration and Invasion Assays. *J. Vis. Exp.* **2014**, 88.
5. de Albuquerque, G. R.; Nakamura, C. V.; de Souza, W.; Morgado-Díaz, J. A. Differential expression of sialic acid and N-acetylgalactosamine residues on the cell surface of intestinal epithelial cells according to normal or metastatic potential. *J Histochem Cytochem.* **2004**, *52* (5), 629-40.
6. Höbarth, K; Hofbauer, J; Fang-Kircher, S. Plasma sialic acid in patients with prostate cancer. *Br J Urol.* **1993**, *72* (5), 621-4.
7. Listinsky, J.J. Siegal, G.P. Listinsky, C.M. The emerging importance of α -L-fucose in human breast cancer: a review. *Am. J. Transl. Res.* **2011**, *3* (4), 292–322
8. Mannori, G. Santoro, G. Inhibition of colon carcinoma cell lung colony formation by a soluble form of E-selectin. *Am. J. Path.* **1997**, *151*, (1), 233-243.
9. Epstein, J. I.; Allsbrook, W. C. Jr.; Amin, M. B.; Egevad, L. L.; ISUP Grading Committee. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma. *Am. J. Surg. Pathol.* **2005**, *29* (9),1228-42.
10. Wang, X.; Chen, J.; Li, Q. K.; Peskoe, S. B.; Zhang, B.; Choi, C. Overexpression of α (1,6) fucosyltransferase associated with aggressive prostate cancer. *Glycobiology* **2014**, *24* (10), 935–944.
11. Saldova, R.; Dempsey, E.; Perez-Garay, M.; Marino, K.; Watson, J. A.; Blanco-Fernandez, A. 5-AZA-2'-deoxycytidine induced demethylation influences N-glycosylation of secreted glycoproteins in ovarian cancer. *Epigenetics: Official Journal of the DNA Methylation Society* **2011**, *6*, 1362–1372.
12. Uozumi, N.; Teshima, T.; Yamamoto, T.; Nishikawa, A.; Gao, Y. E.; Miyoshi, E. A fluorescent assay method for GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha 1-6 fucosyltransferase activity, involving high performance liquid chromatography. *J. of Biochem.* **1996**, *120*, 385–392.
13. Simon, J.; Salzbrunn, S.; Prakash, G. K. S.; Petasis, N. A.; Olah, G. A. Regioselective conversion of arylboronic acids to phenols and subsequent coupling to symmetrical diaryl ethers. *J. Org. Chem.* **2001**, *66* (2), 633-634.
14. Bünger, S.; Roblick, U. J.; Habermann, J. K. Comparison of five commercial extraction kits for subsequent membrane protein profiling. *Cytotechnology* **2009**, *61* (3), 153–159.

BIBLIOGRAPHY:

- Amerongen, A. V. N.; Bolscher, J. G. M.; Veerman, E. C. I., Salivary mucins: Protective functions in relation to their diversity. *Glycobiology* **1995**, *5* (8), 733-740.
- An, H. J.; Miyamoto, S.; Lancaster, K. S.; Kirmiz, C.; Li, B.; Lam, K. S.; Leiserowitz, G. S.; Lebrilla, C. B. Profiling of glycans in serum for the discovery of potential biomarkers for ovarian cancer. *J. Proteome Res.* **2006**, *5*, 1626–1635.
- Anslyn, E. V.; McCleskey, S. C.; Griffin, M. J.; Schneider, S. E.; McDevitt, J. T., Differential receptors create patterns diagnostic for ATP and GTP. *J. Am. Chem. Soc.* **2003**, *125* (5), 1114-1115.
- Anslyn, E. V.; Wright, A. T.; Griffin, M. J.; Zhong, Z. L.; McCleskey, S. C.; McDevitt, J. T., Differential receptors create patterns that distinguish various proteins. *Angew. Chem. Int. Ed.* **2005**, *44* (39), 6375-6378.
- Anslyn, E.V.; Lavigne, J.J. Boronic acid based peptidic receptors for pattern-based saccharide sensing in neutral aqueous media, an application in real-life samples. *J. Am. Chem. Soc.* **2007**, *129*, 13575-13583.
- Aragay, G.; Hernández, D.; Verdejo, B.; Escudero-Adán, E.C.; Martínez, M.; Ballester, P. Quantification of CH- π Interactions Using Calix(4)pyrrole Receptors as Model Systems. *Molecules.* **2015**, *20*, 16672-16686.
- Arimori, S.; Ushiroda, S.; Peter, L. M.; Jenkins, A. T. A.; James, T. D., A modular electrochemical sensor for saccharides. *Chemical Communications* **2002**, (20), 2368-2369.
- Ballester, P.; Biros, S.M. CH- π and π - π interactions as contributors to the guest binding in reversible inclusion and encapsulation complexes. In *The Importance of Pi-Interactions in Crystal Engineering*; John Wiley & Sons, Ltd: Chichester, West Sussex, UK, **2012**, 79–107.
- Barrabes, S. Glycosylation of serum ribonuclease I indicates a major endothelial origin and reveals an increase in core-fucosylation in pancreatic cancer, *Cancer Res.* **2007**, *17*, 388-400.
- Barry, M. J., Prostate-Specific–Antigen Testing for Early Diagnosis of Prostate Cancer. *New England Journal of Medicine* **2001**, *344* (18), 1373-1377.
- Barry, M. J., Prostate-specific-antigen testing for the early diagnosis of prostate cancer. *N. Engl. J. Med.* **2001**, *344*, 1373-1377.
- Bast, R. C., Jr.; Ravdin, P.; Hayes, D. F.; Bates, S.; Fritsche, H., Jr.; Jessup, J. M.; Kemeny, N.; Locker, G. Y.; Mennel, R. G.; Somerfield, M. R. 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical

practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol.* **2001**, *19* (6), 1865-78.

- Berg, J. M.; Stryer, L. Biochemistry. In Biochemistry, 5th ed.; W.H. Freeman: New York **2002**, (5).
- Berube, M.; Dowlut, M.; Hall, D. G. Benzoboroxoles as efficient glycopyranoside-binding agents in physiological conditions: structure and selectivity of complex formation. *J. Org. Chem.* **2008**, *73*, 6471–6479.
- Bicker, K.L.; Sun, J.; Lavigne, J.J. Boronic acid functionalized peptidyl synthetic lectins: combinatorial library design, Peptide sequencing, and selective glycoprotein recognition. *ACS Comb. Sci.* **2011**, *13*, 232-243.
- Bicker, K.L.; Sun, J.; Lavigne, J.J. Synthetic lectin arrays for the detection and discrimination of cancer associated glycans and cell lines. *Chem. Sci.* **2012**, *3*, 1147-1156.
- Brandl, M.; Weiss, M.S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. C-H. π -interactions in proteins. *J Mol Biol.* **2001**, *307*, 357-77.
- Brattain, M. D.; Fine, W. D.; Khaled, F. M.; Thompson, J.; Brattain, D. E. Heterogeneity of malignant cells from a human colonic carcinoma. *Cancer Res.* **1981**, *41* (5), 1751-6.
- Brattoli, M.; Gennaro, G.D. Odour detection methods: olfactometry and chemical sensors, *Sensors* **2011**, *115*, 5290-5322.
- Brawer, M. K.; Lange, P. H. Prostate-specific antigen in management of prostatic carcinoma. *Urology* **1989**, *33* (5), 11–16.
- Bünger, S.; Roblick, U. J.; Habermann, J. K. Comparison of five commercial extraction kits for subsequent membrane protein profiling. *Cytotechnology* **2009**, *61* (3), 153–159.
- Calvin, R.; Justus, N. L.; Ruiz-Echevarria, M.; Yang, L.V. In vitro Cell Migration and Invasion Assays. *J. Vis. Exp.* **2014**, 88.
- Candas, B.; Cusan, L.; Gomez, J. L.; Diamond, P.; Suburu, R. E.; Lévesque, J. Evaluation of prostatic specific antigen and digital rectal examination as screening tests for prostate cancer. *Prostate* **2000**, *45*, 19–35.
- Cao, Y.; Karsten, U. R.; Liebrich, W.; Haensch, W.; Springer, G. F.; Schlag, P. M. Expression of Thomsen-Friedenreich-related antigens in primary and metastatic colorectal carcinomas. A reevaluation. *Cancer* **1995**, *76* (10), 1700–1708.
- Collins, B. E.; Sorey, S.; Hargrove, A. E.; Shabbir, S. H.; Lynch, V. M.; Anslyn, E. V., Probing Intramolecular B–N Interactions in Ortho-Aminomethyl Arylboronic Acids. *J. Org. Chem.* **2009**, *74* (11), 4055-4060.
- de Albuquerque, G. R; Nakamura, C. V.; de Souza, W.; Morgado-Díaz, J. A. Differential expression of sialic acid and N-acetylgalactosamine residues on the cell surface of intestinal epithelial cells according to normal or metastatic potential. *J Histochem Cytochem.* **2004**, *52* (5), 629-40.

- Dell, A.; Galadari, A.; Sastre, F.; Hitchen, P., Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *International Journal of Microbiology* **2010**, *2010*, 1-14.
- DiCesare, N.; Lakowicz, J. R., Spectral Properties of Fluorophores Combining the Boronic Acid Group with Electron Donor or Withdrawing Groups. Implication in the Development of Fluorescence Probes for Saccharides. *The Journal of Physical Chemistry A* **2001**, *105* (28), 6834-6840.
- Dougherty, D.A. Cation- π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science*, **1996**, *271*, 163-168.
- Dube, D. H.; Bertozzi, C. R. Glycans in cancer and inflammation, potential for therapeutics and diagnostics. *Nature Reviews* **2005**, *4*, 477-488.
- Dube, D.H.; Bertozzi, C.R. Glycans in cancer and inflammation, potential for therapeutics and diagnostics. *Nature Reviews* **2005**, *4*, 477-488.
- Durrant, L. G.; Noble, P.; Spendlove, I. Immunology in the clinic review series; focus on cancer: Glycolipids as targets for tumour immunotherapy. *Clin. and Exp. Immunology* **2012**, *167* (2), 206–215.
- Edwards, N. Y.; Sager, T. W.; McDevitt, J. T.; Anslyn, E. V., Boronic acid based peptidic receptors for pattern-based saccharide sensing in neutral aqueous media, an application in real-life samples. *J. Am. Chem. Soc.* **2007**, *129*, 13575-13583.
- Epstein, J. I.; Allsbrook, W. C. Jr.; Amin, M. B.; Egevad, L. L.; ISUP Grading Committee. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma. *Am. J. Surg. Pathol.* **2005**, *29* (9),1228-42.
- Fakih, M.G.; Aruna, P. CEA monitoring in colorectal cancer: what you should know. *Oncology* **2006**, *20*, 579-587.
- Ferlay, J.; Soerjomataram, I.; Ervik, M.; Dikshit, R.; Eser, S.; Mathers, C. Cancer Incidence and Mortality Worldwide: IARC. *International Agency for Research on Cancer GLOBOCAN* **2012**.
- Fink, K.; Boratyński, J. Noncovalent cation- π interactions--their role in nature. *Postepy. Hig. Med. Dosw. (Online)*. **2014**, *68*, 1276-86.
- Gao, X.; Zhang, Y.; Wang, B., A highly fluorescent water-soluble boronic acid reporter for saccharide sensing that shows ratiometric UV changes and significant fluorescence changes. *Tetrahedron* **2005**, *61* (38), 9111-9117.
- Gao, X.; Zhang, Y.; Wang, B., Naphthalene-based water-soluble fluorescent boronic acid isomers suitable for ratiometric and off-on sensing of saccharides at physiological pH. *New Journal of Chemistry* **2005**, *29* (4), 579-586.
- Gao, X.; Zhang, Y.; Wang, B., New Boronic Acid Fluorescent Reporter Compounds. 2. A Naphthalene-Based On–Off Sensor Functional at Physiological pH. *Organic Letters* **2003**, *5* (24), 4615-4618.
- Garcia, d. A. R.; Nakamura, C.V.; de Souza, W.; Morgado-Díaz, J. A. Differential expression of sialic acid and N-acetylgalactosamine residues on the cell surface of

intestinal epithelial cells according to normal or metastatic potential. *J. Histochem. Cytochem.* **2004**, *52*, 629-640.

- Gu, J.; Taniguchi, N. Potential of N-glycan in cell adhesion and migration as either a positive or negative regulator, *Cell Adh Migr.* **2008**, *2*, (4), 243–245.
- Gupta, S. H., E.A.; Rockey, D.C.; Hammons, M.; Koch, M.; Carter, E.; Valdez, L.; Tony, L.; Ahn, C.; Kashner, M.; Argenbright, K.; Tiro, J.; Geng, Z.; Pruitt, S.; Skinner, C., Comparative effectiveness of fecal immunochemical test outreach, colonoscopy outreach, and usual care for boosting colorectal cancer screening among the underserved. *JAMA Internal Medicine* **2013**, *173* (18), 1725-1732.
- Gupta, S. S., D.A.; Doubenia, C.A.; Anderson, D.S.; Lukejohn, D.; Deshpande, A.R.; Elmunzer, B.J.; Laiyemo, A.O.; Mendez, J.; Somsouk, M.; Allison, J.; Bhuket, T.I.; Geng, Z.; Green, B.B.; Itzkowitz, S.H.; Martinez, M.E., Challenges and possible solutions to colorectal cancer screening for the underserved. *J. of the National Cancer Institute* **2014**, *106* (4).
- Hall, D. G.; Pal, A.; Berube, M., Design, Synthesis, and Screening of a Library of Peptidyl Bis(Boroxoles) as Oligosaccharide Receptors in Water: Identification of a Receptor for the Tumor Marker TF-Antigen Disaccharide. *Angew. Chem. Int. Ed.* **2010**, *49* (8), 1492-1495.
- Han–Mo Chiu, Y. C. L., Chia–Hung Tu, Chien–Chuan Chen, Ping–Huei Tseng, Jin–Tung Liang, Chia–Tung Shun, Jaw–Town Lin, Ming–Shiang Wu, Association between early stage colon neoplasms and false-negative results from the fecal immunochemical test. *Clinical Gastroenterology and Hepatology* **2013**, *11* (7), 832-838.
- Höbarth, K; Hofbauer, J; Fang-Kircher, S. Plasma sialic acid in patients with prostate cancer. *Br J Urol.* **1993**, *72* (5), 621-4.
- Hollingsworth, M.A; Swanson, B.J. Mucins in cancer: Protection and control of the cell surface. *Nature Rev. Cancer* **2004**, *4*, 45-60.
- Hua, S.; An, H. J.; Ozcan, S.; Ro, G. S.; Soares, S.; DeVere-White, R.; Lebrilla, C. B. Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers. *Analyst*, **2011**, *136*, 3663–3671.
- Hudson, K. L.; Bartlet, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. Carbohydrate-Aromatic Interactions in Proteins. *J Am Chem Soc.* **2015**, *137*(48), 15152-60.
- James, T. D.; Samankumara Sandanayake, K. R. A.; Shinkai, S., Chiral discrimination of monosaccharides using a fluorescent molecular sensor. *Nature* **1995**, *374* (6520), 345-347.
- James, T. D.; Sandanayake, K. R. A. S.; Iguchi, R.; Shinkai, S., Novel Saccharide-Photoinduced Electron Transfer Sensors Based on the Interaction of Boronic Acid and Amine. *Journal of the American Chemical Society* **1995**, *117* (35), 8982-8987.
- Kirmiz, C.; Li, B.; An, H. J.; Clowers, B. H.; Chew, H. K.; Lam, K. S.; Ferrige, A.; Alecio, R.; Borowsky, A. D.; Sulaimon, S. A serum glycomics approach to breast cancer biomarkers. *Mol. Cell Proteomics* **2007**, *6*, 43–55.

- Klarmann, G. J.; Hurt, E. M.; Mathews, L. A.; Zhang, X.; Duhagon, M. A.; Mistree, T.; Thomas, S. B.; Farrar, W. L. Invasive Prostate Cancer Cells Are Tumor Initiating Cells That Have A Stem Cell-Like Genomic Signature. *Clin. Exp. Metastasis* **2009**, *26* (5), 433–446.
- Koprowski, H.; Herlyn, M.; Stepkowski, Z.; Sears, H. F., Specific antigen in serum of patients with colon carcinoma. *Science* **1981**, *212* (4490), 53-55.
- Kyselova, Z.; Mechref, Y.; Kang, P.; Goetz, J. A.; Dobrolecki, L. E.; Sledge, G. W.; Schnaper, L.; Hickey, R. J.; Malkas, L. H.; Novotny, M.V. Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. *Clin Chem.* **2008**, *54*, 1166–1175.
- Lavigne, J. J. A., E. V. Sensing a paradigm shift in the field of molecular recognition: from selective to differential receptors. *Angew. Chem Int. Ed* **2001**, *40*, 3118-3130.
- Lavigne, J. J.; Anslyn, E.V. Solution-based analysis of multiple analytes by a sensor array: toward the development of an “electronic tongue”. *J. Am. Chem. Soc.* **1998**, *120*, 6429-6430.
- Levi, Z. B., S.; Vilkin, A.; Bar-Chana, M.; Lifshitz, I.; Chared, M.; Maoz, E.; Niv, Y. A higher detection rate for colorectal cancer and advanced adenomatous polyp for screening with immunochemical fecal occult blood test than guaiac fecal occult blood test, despite lower compliance rate. A prospective, controlled, feasibility study. *International J. of Cancer* **2011**, *128* (10), 2415-2424.
- Lin, N.; Mascarenhas, J.; Sealover, N.R.; George, H.J.; Brooks, J.; Kayser, K.J.; Gau, B.; Yasa, I.; Azadi, P.; and Archer-Hartmann, S. Chinese hamster ovary (cho) host cell engineering to increase sialylation of recombinant therapeutic proteins by modulating sialyltransferase expression. *Biotechnol. Prog.* **2015**, *31*, 334–346.
- Listinsky, J.J. Siegal, G.P. Listinsky, C.M. The emerging importance of α -L-fucose in human breast cancer: a review. *Am. J. Transl. Res.*, **2011**, *3* (4), 292–322.
- Listinsky, J.J. Siegal, G.P. Listinsky, C.M. The emerging importance of α -L-fucose in human breast cancer: a review. *Am. J. Transl. Res.* **2011**, *3* (4), 292–322
- Magnani, J. L.; Stepkowski, Z.; Koprowski, H.; Ginsburg, V., Identification of the gastrointestinal and pancreatic cancer-associated antigen detected by monoclonal antibody 19-9 in the sera of patients as a mucin. *Cancer Res.* **1983**, *43* (11), 5489-92.
- Magnani, J.L. *Arch. Biochem. Biophys.* **2004**, *426*, 122-131.
- Mammen, M.; Choi, S.-K.; Whitesides, G. M., Polyvalent interaction in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew. Chem. Int. Ed.* **1998**, *37*, 2754-2794.
- Mannori, G. Santoro, G. Inhibition of colon carcinoma cell lung colony formation by a soluble form of E-selectin. *Am. J. Path.* **1997**, *151*, (1), 233-243.
- Mechref, Y.; Hu, Y.; Garcia, A.; Zhou, S.; Desantos-Garcia, J. L.; Hussein, A. Defining putative glycan cancer biomarkers by MS. *Bioanalysis* **2012**, *4*, 2457–2469.
- Meezan, E.; Wu, H.C.; Black, P.H.; Robbins, P.W. Comparative studies on the carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. *J.Am.Chem. Soc.* **1969**, *8*, 2518-2524.

- Mody, R.; Joshi, S.; Chaney, W. Use of lectins as diagnostic and therapeutic tools for cancer. *J. Pharm. & Toxic. Methods* **1995**, 33.
- Moerke, N. J., Fluorescence polarization (FP) assay for monitoring peptide-protein or nucleic acid-protein binding. *Curr. Protoc. Chem. Biol.* **2009**, 1, 1-15.
- Moertel, C. CEA monitoring among patients in multi-institutional adjuvant G.I therapy protocols. *Ann. Surg.* **1982**, 196 (2), 162-169.
- Munkley, J. E.; Hallmarks of glycosylation in cancer. *Oncotarget* **2016**, 7 (23), 35478-35483.
- Munkley, J.; Elliott, D. J. Hallmarks of glycosylation in cancer. *Oncotarget* **2016**, 7 (23), 35478-89.
- Nishio, M.; Hirota, M.; Umezawa, Y. The CH/Pi Interaction. Evidence, Nature, and Consequences; Wiley-VCH: New York, NY, USA, **1998**.
- Nishio, M.; Umezawa, Y.; Fantini, J.; Weiss, M.S.; Chakrabarti, P. CH/ π hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.* **2014**, 16, 12648–12683.
- Nuti, M.; Teramoto, Y. A.; Mariani-Costantini, R.; Hand, P. H.; Colcher, D.; Schlom, J. A monoclonal antibody (B72.3) defines patterns of distribution of a novel tumor associated antigen in human mammary carcinoma cell populations. *Int. J. Cancer* **1982**, 29 (5), 539–545.
- Oshovsky, G. V.; Reinhoudt, D. N.; Verboom, W., Supramolecular Chemistry in Water. *Angew. Chem. Int. Ed.* **2007**, 46 (14), 2366-2393.
- Osumi, D.; Takahashi, M.; Miyoshi, E.; Yokoe, S.; Lee, S. H.; Noda, K. Core-fucosylation of E-cadherin enhances cell-cell adhesion in human colon carcinoma WiDr cells. *Cancer Science*, **2009**, 100, 888–895.
- Pal, A.; Bérubé, M.; Hall, D. G., Design, Synthesis, and Screening of a Library of Peptidyl Bis(Boroxoles) as Oligosaccharide Receptors in Water: Identification of a Receptor for the Tumor Marker TF-Antigen Disaccharide. *Angewandte Chemie International Edition* **2010**, 49 (8), 1492-1495.
- Pal, A.; Hall, D.G. Design, Synthesis and screening of a library of peptidyl bis(boroxoles) as oligosaccharide receptors in water: Identification of a receptor for the tumor marker TF-Antigen disaccharide. *Angew. Chem. Int. Ed.* **2010**, 49, 1492-1495.
- Palit, D. K.; Pal, H.; Mukherjee, T.; Mittal, J. P., Photodynamics of the S1 state of some hydroxy- and amino-substituted naphthoquinones and anthraquinones. *J. Chem. Soc. Faraday Trans.* **1990**, 86 (23), 3861-3869.
- Passerini, R.; Cassatella, M. D.; Boveri, S.; Salvatici, M.; Radice, D.; Zorzino, L.; Galli, C.; Sandri, M. T. The Pitfalls of CA19-9: Routine Testing and Comparison of Two Automated Immunoassays in a Reference Oncology Center. *Am J Clin Pathol* **2012**, 138 (2): 281-287.

- Patwa, T. H. Z., J.; Anderson, M.A.; Simeone, D.M.; Lubman, D.M., Screening of Glycosylation Patterns in Serum Using Natural Glycoprotein Microarrays and Multi-Lectin Fluorescence Detection. *Anal. Chem.* **2006**, *78*, 6411-6421.
- Paul, H.; Reginato, A. J.; Schumacher, H. R., Alizarin Red S Staining as a Screening Test to Detect Calcium Compounds in Synovial Fluid. *Arthritis Rheum.* **1983**, *26* (2), 191-200.
- Peracaula, R.; Barrabes, S.; Sarrats, A.; Rudd, P. M.; de Llorens R. Altered glycosylation in tumours focused to cancer diagnosis. *Dis Markers.* **2008**, *25*, 207–218.
- Peracaula, R.; Tabares, G.; Royle, L.; Harvey, D.J., Dwek, R.A.; Rudd, P.M.; de Llorens, R. Altered glycosylation pattern allows the distinction between prostate specific antigen from normal and tumour origins. *Glycobiology* **2003**, *13*, 457-470.
- Pfeiffer, M. J.; Schalken, J. A. Stem Cell Characteristics in Prostate Cancer Cell Lines. *European Urology*, **2009**.
- Pinho, S. S.; Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer* **2015**, *15*, 540–555.
- Postma, R.; Schroder, F. H.; van Leenders, G. J.; Hoedemaeker, R. F.; Vis A. N.; Roobol, M. J. Cancer detection and cancer characteristics in the European Randomized Study of Screening for Prostate Cancer (ERSPC) – Section Rotterdam. A comparison of two rounds of screening. *Eur Urol.* **2007**, *52*, 89–97.
- Quintero, E. C., A.; Bujanda, L.; Cubiella, J. et al, Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. *New England Journal of Medicine* **2012**, *366*, 697-706.
- Ries, L.; Melbert, D.; Krapcho, M. SEER Cancer Statistics Review, 1975–2004, *National Cancer Institute*, **2007**.
- Ruhaak, L. R.; Miyamoto, S.; Lebrilla, C. B. Developments in the Identification of Glycan Biomarkers for the Detection of Cancer. *Mol. Cell Proteomics.* **2013**, *12* (4), 846–855.
- Saitoh, O.; Wang, W. C.; Lotan, R.; Fukuda, M. Differential glycosylation and cell surface expression of lysosomal membrane glycoproteins in sublines of a human colon cancer exhibiting distinct metastatic potentials. *J. Biol. Chem.* **1992**, *267*, 5700-5711.
- Saldova R.; Fan, Y.; Fitzpatrick J. M. ; William R. ; Watson, G. ; Rudd, P. M. Core fucosylation and α 2-3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia. *Glycobiology* **2011**, *21* (2), 195–205.
- Saldova, R.; Dempsey, E.; Perez-Garay, M.; Marino, K.; Watson, J. A.; Blanco-Fernandez, A. 5-AZA-2'-deoxycytidine induced demethylation influences N-glycosylation of secreted glycoproteins in ovarian cancer. *Epigenetics: Official Journal of the DNA Methylation Society* **2011**, *6*, 1362–1372.
- Schröder, F. H.; van der Crujisen-Koeter, I.; de Koning H. J.; Vis A. N.; Hoedemaeker, R. F.; Kranse, R.; Prostate cancer detection at low prostate specific antigen. *J Urol.* **2000**, *163*, 806–812.

- Schroder, F. H.; van der Maas, P.; Beemsterboer, P.; Kruger, A. B.; Hoedemaeker, R.; Rietbergen, J. Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst.* **1998**, 1817–1823.
- Shin, S. Y.; Kim, C. G.; Jung, Y. J.; Lim, Y. L.; Lee, Y. H. The UPR inducer DPP23 inhibits the metastatic potential of MDA-MB-231 human breast cancer cells by targeting the Akt–IKK–NF- κ B–MMP-9 axis. *Nature Scientific Reports* **2016**, *6*, 1-12.
- Shinkai, M.D.; Boronic acids in saccharide recognition. Royal *Society of Chemistry Publishing*; **2006**, 14-16.
- Shui, W.; Li, Z. Glycoproteomic analysis of tissues from patients with colon cancer using lectin microarrays and nanoLC-MS/MS. *Mol. BioSyst.* **2013**, *9*, 1877.
- Simon, J.; Salzbrunn, S.; Prakash, G. K. S.; Petasis, N. A.; Olah, G. A. Regioselective conversion of arylboronic acids to phenols and subsequent coupling to symmetrical diaryl ethers. *J. Org. Chem.* **2001**, *66* (2), 633-634.
- Singh, H. N. Z.; Demers, A.A.; Kliewer, E.V.; Mahmud, S.M.; Bernstein, C.N. The reduction in colorectal cancer mortality after colonoscopy varies by site of the cancer. *Gastroenterology* **2010**, *139*, 1128-1137.
- Society, A. C., Cancer Facts and Figures **2017**, pp 1-76.
- Springsteen, G.; Wang, B. Alizarin red as a general fluorescent reporter for studying the binding of boronic acids and carbohydrates. *Chem. Commun.* **2001**, 1608–1609.
- Springsteen, G.; Wang, B., A detailed examination of boronic acid-diol complexation. *Tetrahedron* **2002**, *58*, 5291-5300.
- Stones, D.; Manku, S.; Lu, X.; Hall, D. G., Modular Solid-Phase Synthetic Approach To Optimize Structural and Electronic Properties of Oligoboronic Acid Receptors and Sensors for the Aqueous Recognition of Oligosaccharides. *Chemistry – A European Journal* **2004**, *10* (1), 92-100.
- Storr, S. J.; Royle, L.; Chapman, C. J.; Hamid, U. M.; Robertson, J. F.; Murray, A.; Dwek, R. A.; Rudd, P. M. The O-linked glycosylation of secretory/shed MUC1 from an advanced breast cancer patient’s serum. *Glycobiology*, **2008**, *18*, 456–462.
- Sul, J. Y.; Song, I. S.; Bae, C. S.; Chang, E. S.; Yoon, W. H. Metastatic Potential of Human Colon Cancer Cell Lines, LoVo and SW480. *J. Korean Cancer Assoc.* **1995**, *27* (2), 209-223.
- Taniguchi, N., Korekane, H. Branched N-glycans and their implications for cell adhesion, signaling and clinical applications for cancer biomarkers and in therapeutics. *BMB Rep.* **2011**, *44* (12), 772-81.
- Thor, A.; Ohuchi, N.; Szpak, C. A.; Johnston, W. W.; Schlom, J. Distribution of oncofetal antigen tumor-associated glycoprotein-72 defined by monoclonal antibody B72.3. *Cancer Res.* **1986**, *46* (6), 3118–3124.
- Turner, G.A. N-glycosylation of serum proteins in disease and its investigation using lectins, *Clinica Chimica Acta.* **1992**, *3*, 149–171.
- Uozumi, N.; Teshima, T.; Yamamoto, T.; Nishikawa, A.; Gao, Y. E.; Miyoshi, E. A fluorescent assay method for GDP-L-Fuc: N-acetyl-beta-D-glucosaminide alpha 1-6

fucosyltransferase activity, involving high performance liquid chromatography. *J. of Biochem.* **1996**, *120*, 385–392.

- Wang, B. Water-soluble fluorescent boronic acid compounds for saccharide sensing. *Chem. Eur. J.* **2006**, *12*, 1377–1384.
- Wang, Q. Y.; Lu, J.; Liao, S. M.; Du, Q. S.; Huang, R.B. Unconventional interaction forces in protein and protein-ligand systems and their impacts to drug design. *Curr Top Med Chem.* **2013**, *13* (10), 1141-51.
- Wang, W.; Gao, X.; Wang, B. Boronic acid-based sensors. *Current Organic Chemistry* **2002**, *6*, 1285-1317.
- Wang, X.; Chen, J.; Li, Q. K.; Peskoe, S. B.; Zhang, B.; Choi, C. Overexpression of α (1,6) fucosyltransferase associated with aggressive prostate cancer. *Glycobiology* **2014**, *24* (10), 935–944.
- Webber, M. M.; Quader, S. T. A.; Kleinman, H.K.; Bello-DeOcampo, D.; Storto, P. D.; Bice, G.; DeMendonca-Calaca, W.; Williams, D. E. Human Cell Lines as an InVitro/InVivo Model for Prostate Carcinogenesis and Progression. *The Prostate*, **2001**, *47*, 1-13.
- Wright, A. T.; Anslyn, E. V., Differential receptor arrays for solution-based molecular recognition. *Chem. Soc. Rev.* **2006**, *35*, 14-28.
- Wright, A. T.; Griffin, M. J.; Zhenlin, Z.; McCleskey, S. C.; Anslyn, E. V.; McDevitt, J. T., Differential receptors create patterns that distinguish various proteins. *Angew. Chem. Int. Ed.* **2005**, *44* (39), 6375-6378.
- Yang, W.; Carson, J.; Weston, B.; Wang, B. The first fluorescent diboronic acid sensor specific for hepatocellular carcinoma cells expressing sialyl Lewis X. *Chem. Biol.* **2004**, *11*, 439–448.
- Yang, W.; Fan, H.; Gao, X.; Gao, S.; Karnati, V. V. R.; Ni, W.; Hooks, W. B.; Carson, J.; Weston, B.; Wang, B., The First Fluorescent Diboronic Acid Sensor Specific for Hepatocellular Carcinoma Cells Expressing Sialyl Lewis X. *Chemistry & Biology* **2002**, *11* (4), 439-448.
- Yang, W.; Gao, S.; Gao, X.; Karnati, V. V. R.; Ni, W.; Wang, B.; Hooks, W. B.; Carson, J.; Weston, B., Diboronic acids as fluorescent probes for cells expressing sialyl lewis X. *Bioorganic & Medicinal Chemistry Letters* **2002**, *12* (16), 2175-2177.
- Yang, W.; Gao, S.; Weston, B. Diboronic acids as fluorescent probes for cells expressing sialyl lewis X. *Bioorg. Med. Chem. Lett.*, **2002**, *12*, 2175–2177.
- Yang, W.; Yan, J.; Springsteen, G.; Deeter, S.; Wang, B., A novel type of fluorescent boronic acid that shows large fluorescence intensity changes upon binding with a carbohydrate in aqueous solution at physiological pH. *Bioorganic & Medicinal Chemistry Letters* **2003**, *13* (6), 1019-1022.
- Yoon, J.; Czarnik, A. W., Fluorescent chemosensors of carbohydrates. A means of chemically communicating the binding of polyols in water based on chelation-enhanced quenching. *J. Am. Chem. Soc.* **1992**, *114* (14), 5874-5875.

- Zhang, H.; Zhang, Y.; Duan, H. O.; Kirley, S. D.; Lin, S. X.; McDougal, W. S.; Xiao, H.; Chin-Lee Wu. TIP30 is associated with progression and metastasis of prostate cancer. *Int. J. Cancer* **2008**, *123*, 810–816.
- Zhang, Y.; Gao, X.; Hardcastle, K.; Wang, B., Water-Soluble Fluorescent Boronic Acid Compounds for Saccharide Sensing: Substituent Effects on Their Fluorescence Properties. *Chemistry – A European Journal* **2006**, *12* (5), 1377-1384.
- Zhang, Z.; Lang, J.; Cao, Z.; Li, R.; Wang, X.; Wang' W. Radiation-induced SOD2 overexpression sensitizes colorectal cancer to radiation while protecting normal tissue. *Oncotarget*. **2017**, *8* (5), 7791–7800.
- Zhao, J.; H. Z., J.; Anderson, M.A.; Simeone, D.M.; Lubman, D.M., Glycoprotein Microarrays with Multi-Lectin Detection: Unique Lectin Binding Patterns as a Tool for Classifying Normal, Chronic Pancreatitis and Pancreatic Cancer Sera *Anal. Chem.* **2006**, *78*, 6411-6421.
- Zhao, Y. P., Ruan, C. P., Wang, H., Hu, Z. Q., Fang, M., Gu, X. Identification and assessment of new biomarkers for colorectal cancer with serum N-glycan profiling. *Cancer* **2012**, *118* (3), 639–650.
- Zondlo, N.J. Aromatic-proline interactions: Electronically tunable ch/π interactions. *Acc. Chem. Res.* **2013**, *46*, 1039–1049.
- Zou, Y.; Broughton, D. L.; Thompson, P. R.; Lavigne, J. J., Peptide Borono-Lectins (PBLs): A New Tool for Glycomics and Cancer Diagnostics. *ChemBioChem.* **2007**, *8* (17), 2048-2051.
- Zuckerman, D.R.; Rockey, D.C. A prospective multicenter evaluation of new fecal occult blood tests in patients undergoing colonoscopy. *Am. J. Gastroenterol.* **2000**, *95*, 1331-1338.

APPENDIX A: LC-MS OF PEPTIDE SEQUENCES OF SELECT SLS

LC-MS OF PEPTIDE SEQUENCES OF SELECT SLS WITHOUT BORONIC ACIDS,

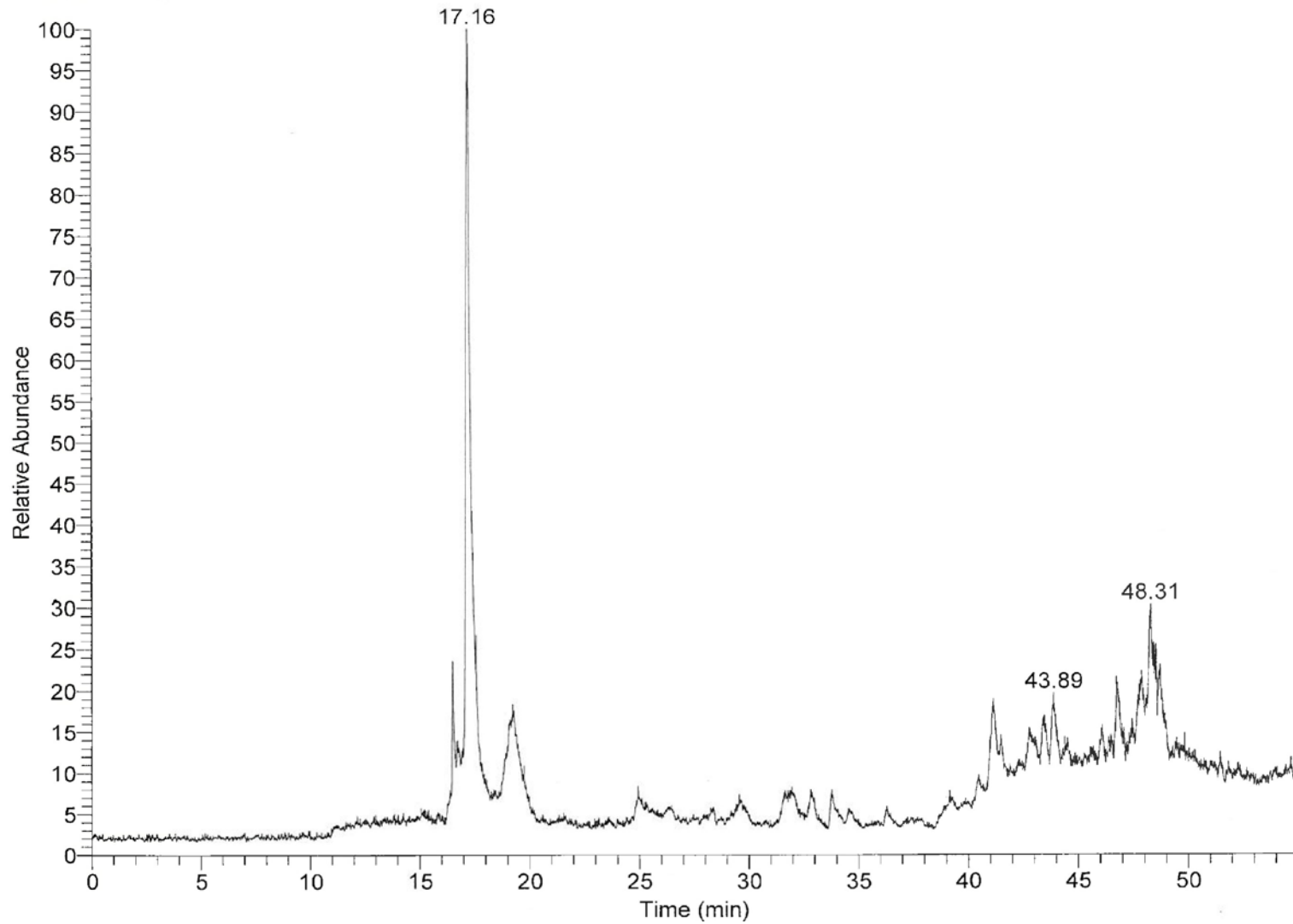
IVDDE OR ACID LABILE SIDE CHAIN GROUPS

- | | | |
|----------------|--------------|------------------------|
| a) Peptide# 91 | SL5 | Molecular Weight: 1325 |
| b) Peptide#100 | SL5-RRAc | Molecular Weight: 1637 |
| c) Peptide#114 | SL5-R1,5,11A | Molecular Weight: 1069 |
| d) Peptide#131 | 4R | Molecular Weight: 1265 |
| e) Peptide#132 | 5R | Molecular Weight: 1352 |

C:\Xcalibur\...\Peptide_2013\Peptide91a
Peptide #91 1325 add 100 then 2ul into 100 ul shoot 0.2 ul
RT: 0.00 - 54.99

11/13/2013 3:39:55 AM

NL:
2.87E7
TIC MS
Peptide91

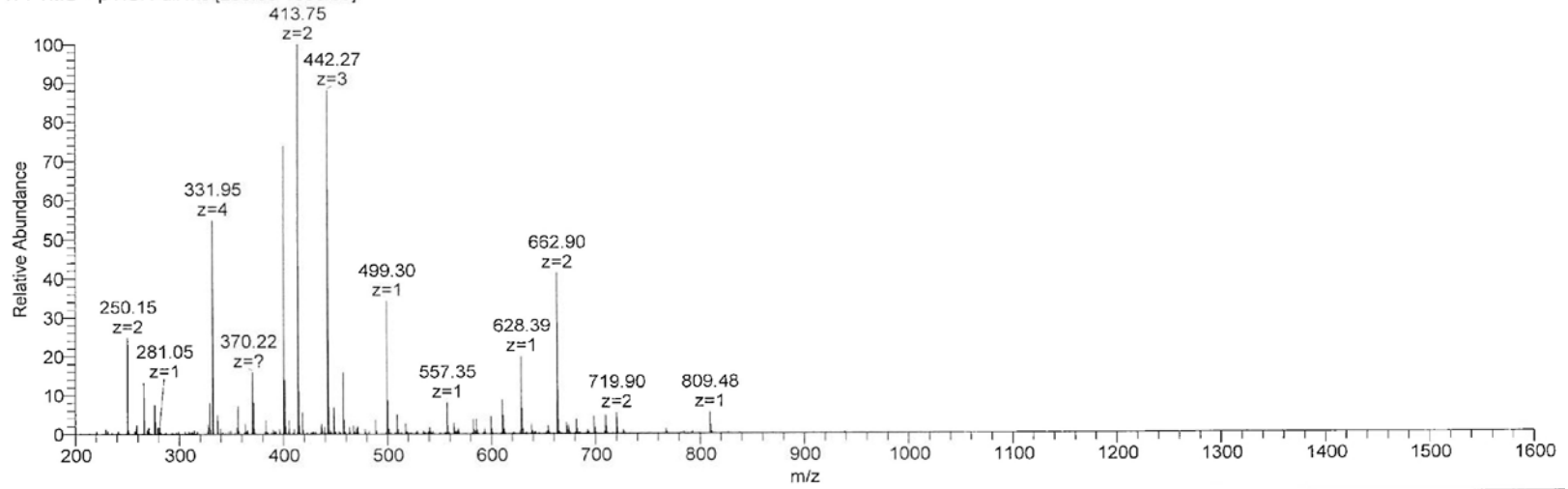


168

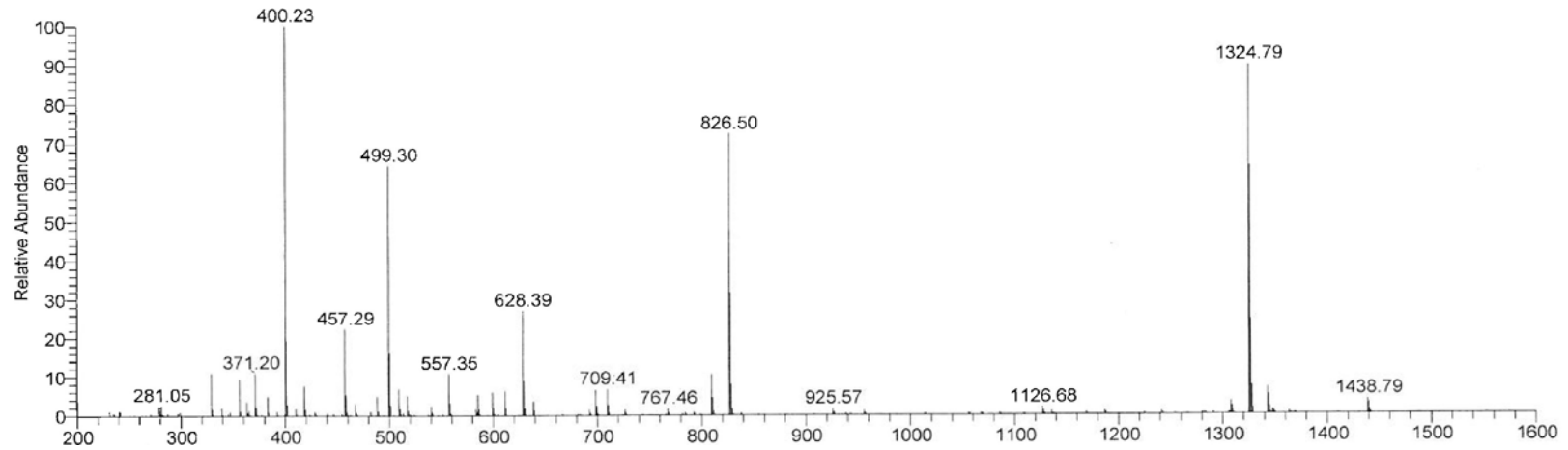
C:\Xcalibur...\Peptide_2013\Peptide91a
Peptide #91 1325 add 100 then 2ul into 100 ul shoot 0.2 ul

11/13/2013 3:39:55 AM

Peptide91a #1440-1585 RT: 17.03-17.71 AV: 146 NL: 1.15E6
T: FTMS + p NSI Full ms [200.00-1800.00]

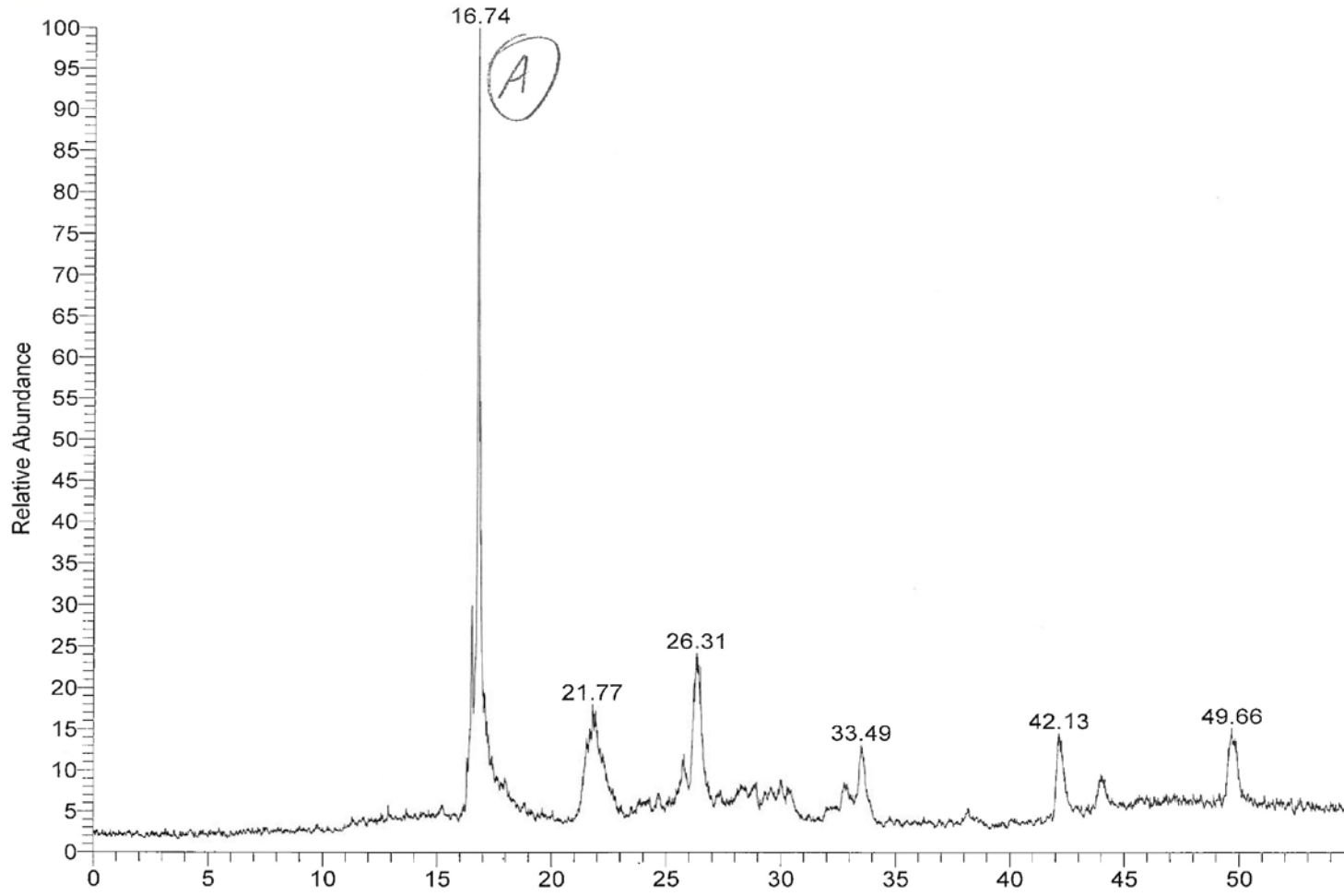


Peptide91a_XT_00001_MHp_131113071303#1 RT: 1.00 AV: 1 NL: 8.39E5
T: FTMS + p NSI Full ms [200.00-1800.00]



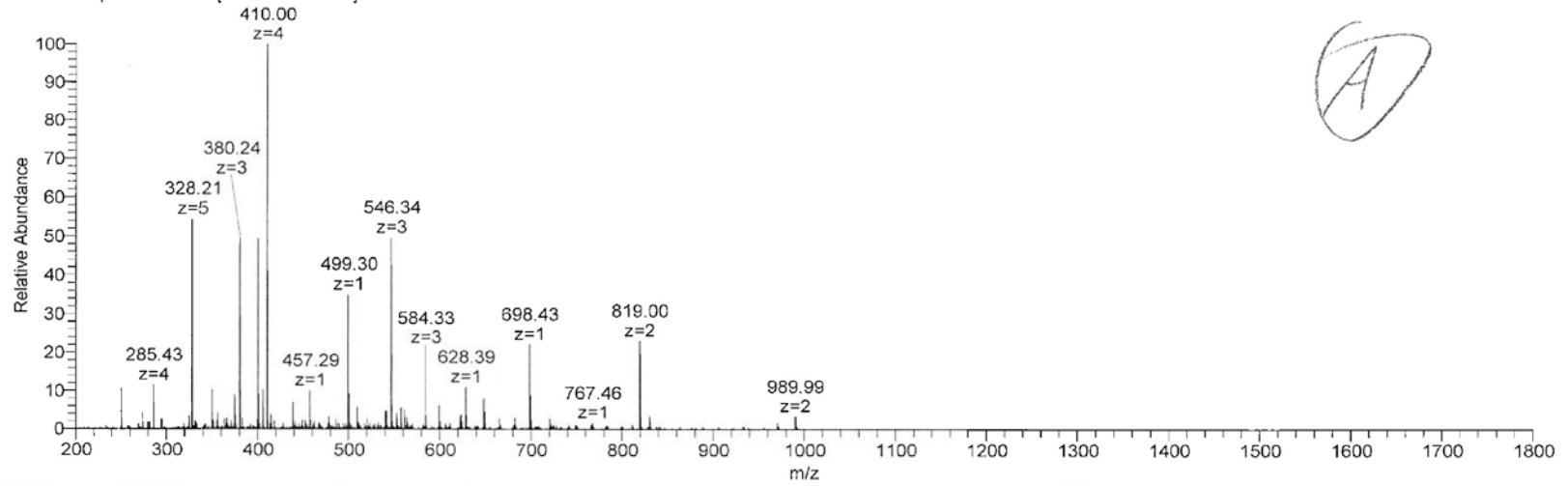
RT: 0.00 - 55.00

NL:
2.93E7
TIC MS
Peptide10
a

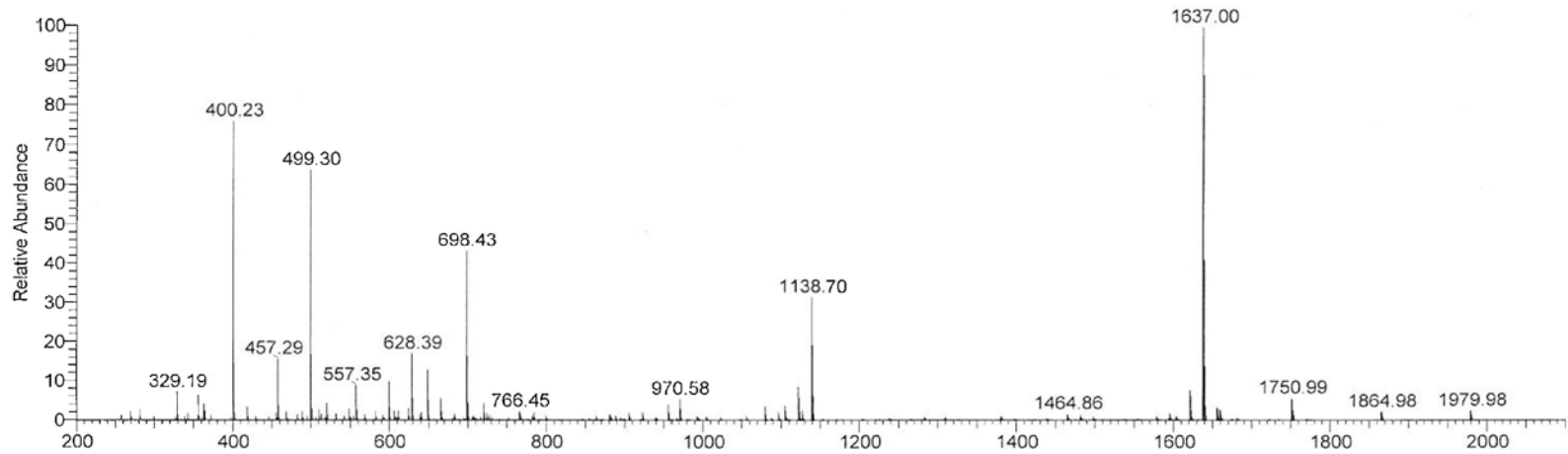


170

Peptide100a#1364-1428 RT: 16.66-16.93 AV: 65 NL: 1.03E6
T: FTMS + p NSI Full ms [200.00-1800.00]

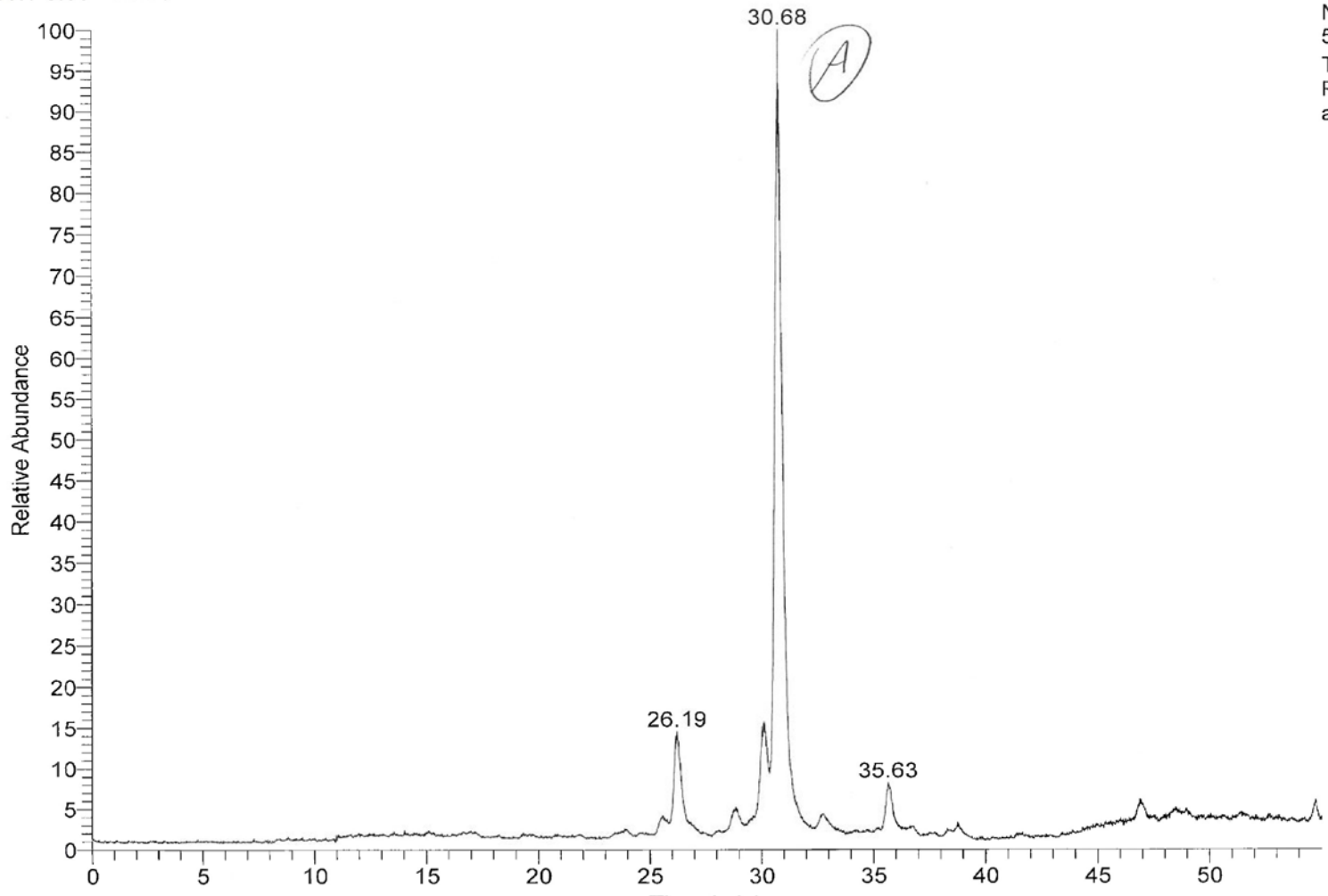


Peptide100a_XT_00001_MHp_#1 RT: 1.00 AV: 1 NL: 6.55E5
T: FTMS + p NSI Full ms [200.00-1800.00]



C:\Xcalibur...\Peptide_2013\Peptide114a *SL5 R15, 114* 12/11/2013 12:25:59 AM
Peptide #114 1064 add 100 ul then 2 into 100 shoot 0.2 ul

RT: 0.00 - 55.00



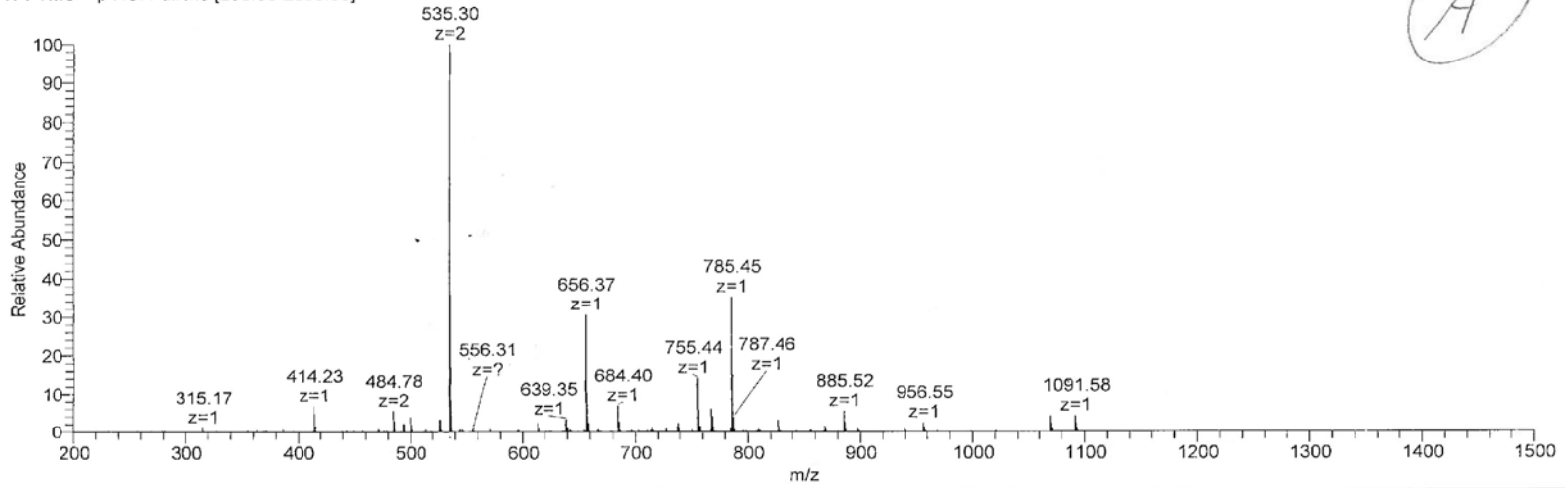
NL:
5.46E7
TIC MS
Peptide114
a

172

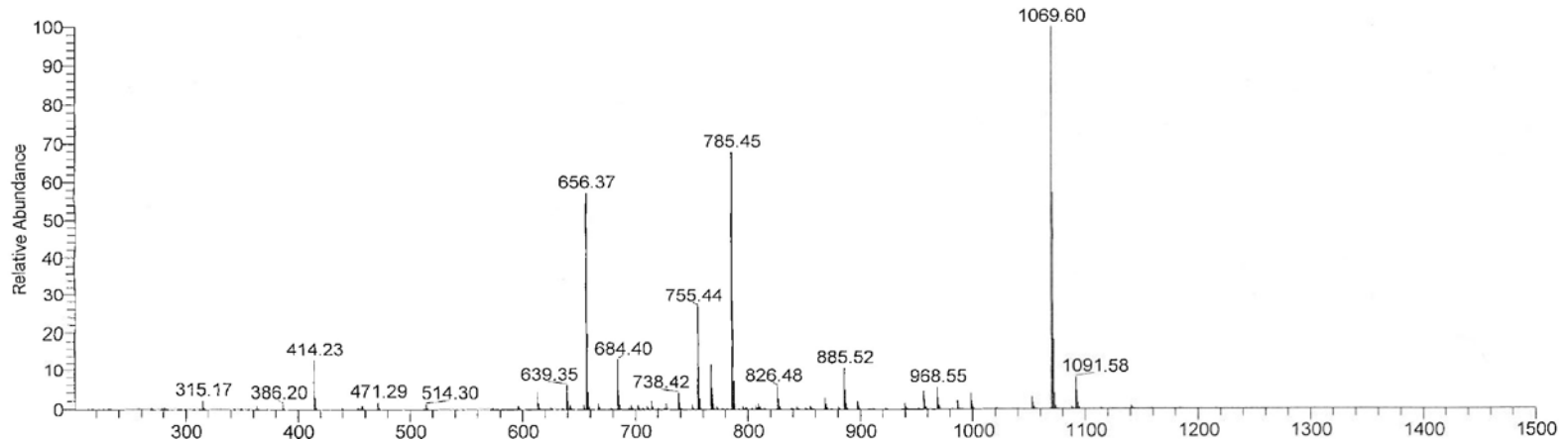
C:\Xcalibur...\Peptide_2013\Peptide114a
Peptide #114 1064 add 100 ul then 2 into 100 shoot 0.2 ul
Peptide114a #2258-2378 RT: 30.54-31.08 AV: 121 NL: 6.72E6
T: FTMS + p NSI Full ms [200.00-2000.00]

12/11/2013 12:25:59 AM

A



Peptide114a_XT_00001_MHp_#1 RT: 1.00 AV: 1 NL: 3.57E6
T: FTMS + p NSI Full ms [200.00-2000.00]



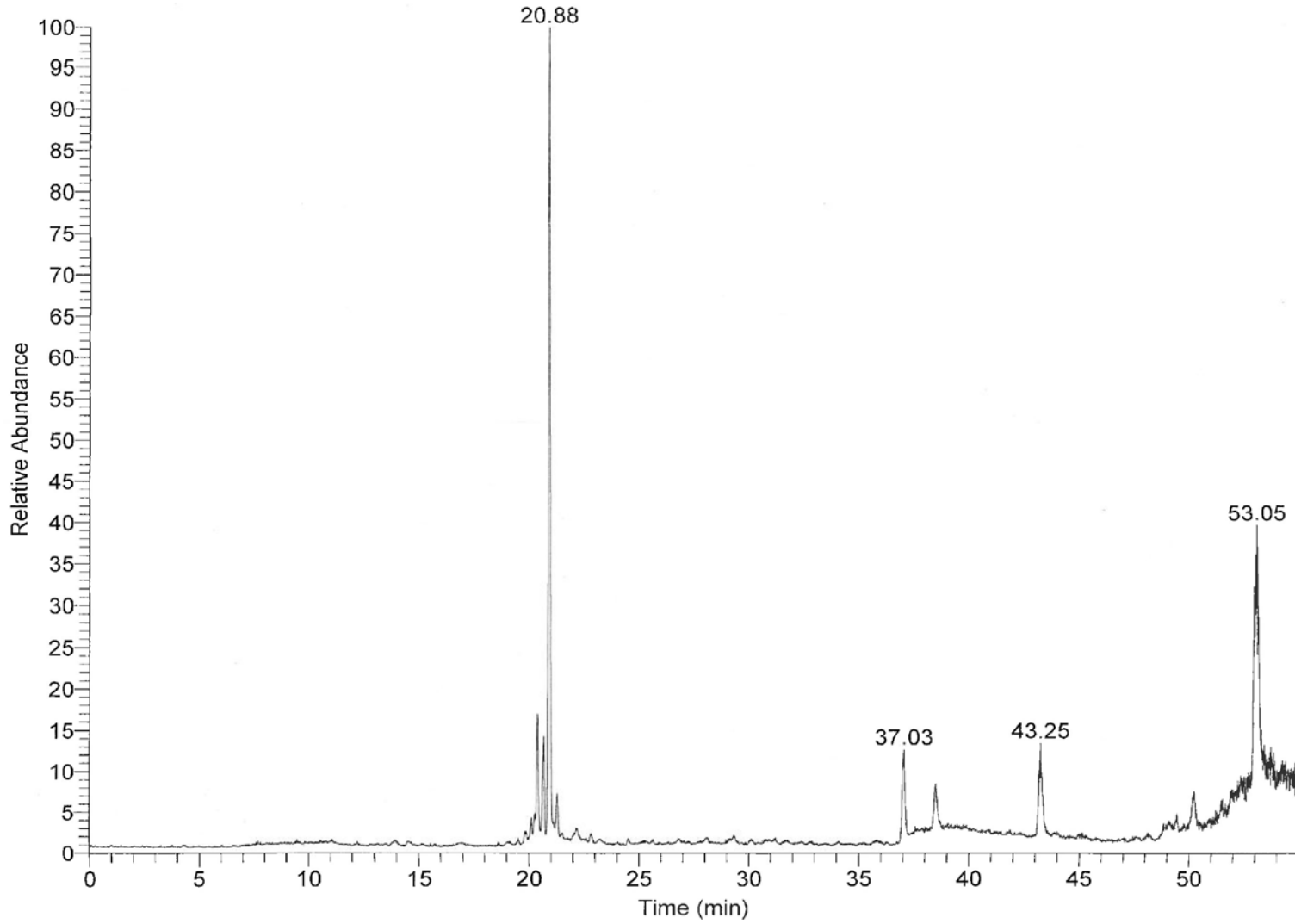
173

C:\Xcalibur\...MarApr\Peptide131a 3/19/2014 1:10:01 AM
Peptide #131 3/18/14 1269 add 100ul then 2 into 100 shoot 0.2

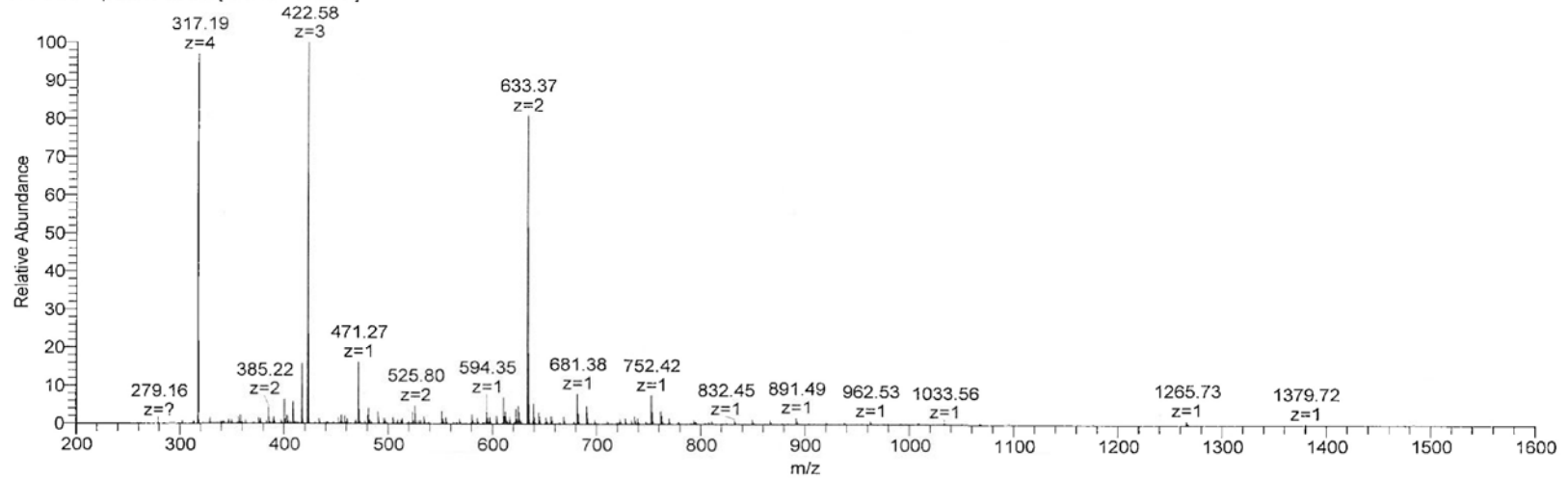
RT: 0.00 - 55.00

NL:
5.77E7
TIC MS
Peptide131
a

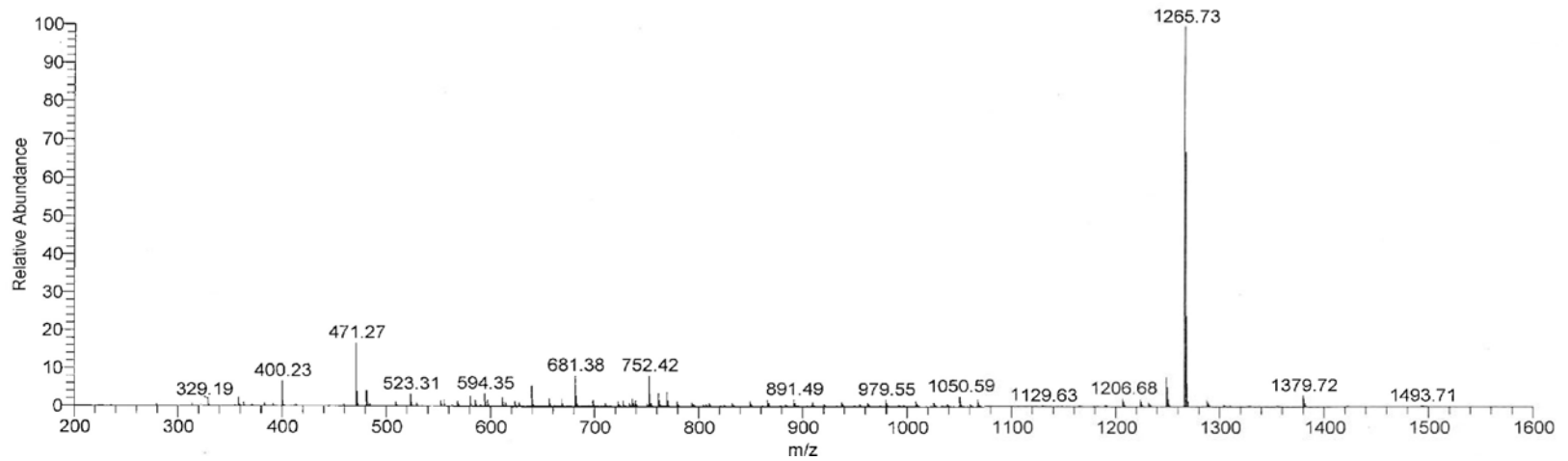
174



Peptide131a #1448-1511 RT: 20.78-21.03 AV: 64 NL: 2.91E6
T: FTMS + p NSI Full ms [200.00-2000.00]



Peptide131a_XT_00001_MHp_#1 RT: 1.00 AV: 1 NL: 2.80E6
T: FTMS + p NSI Full ms [200.00-2000.00]

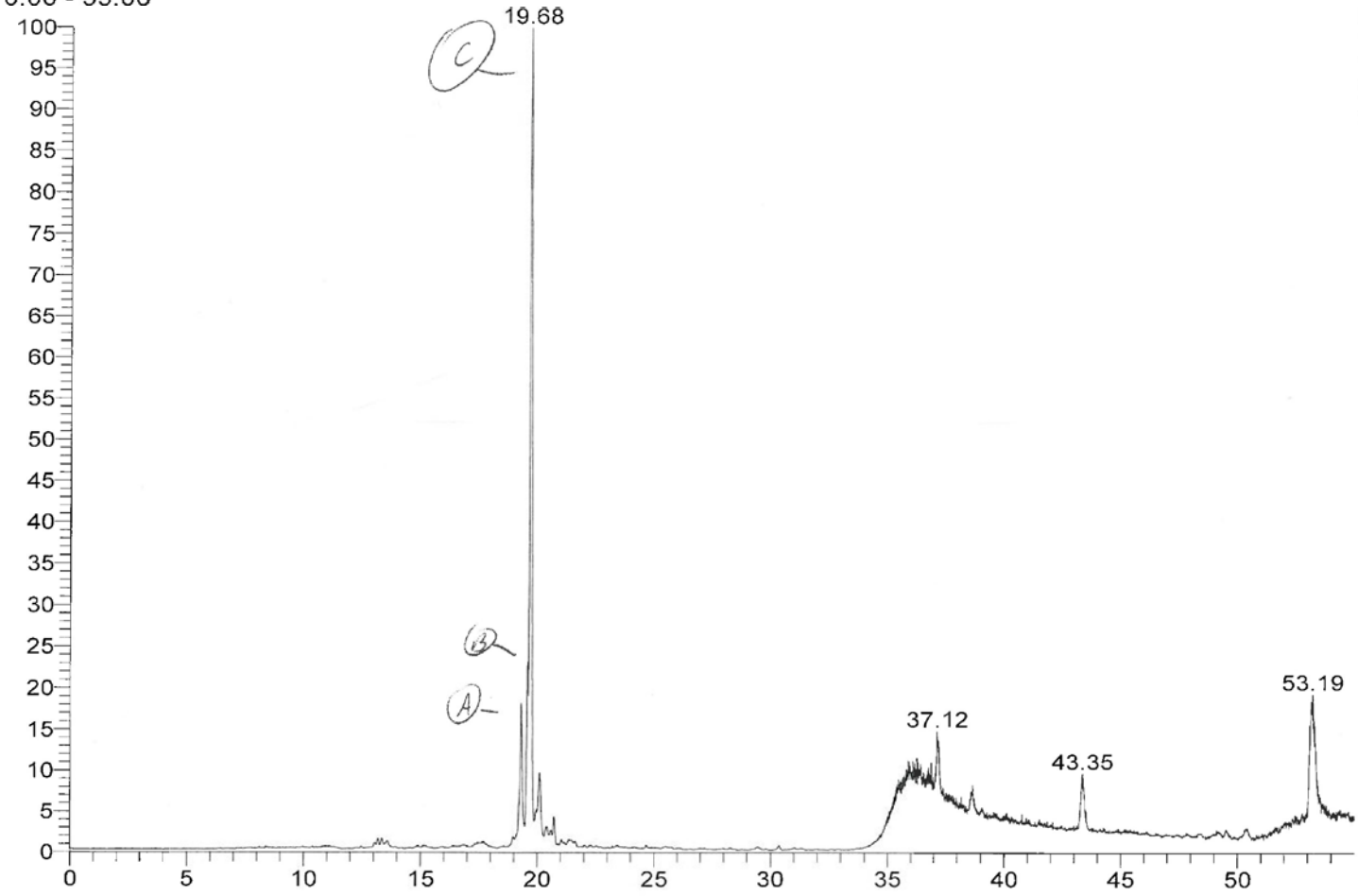


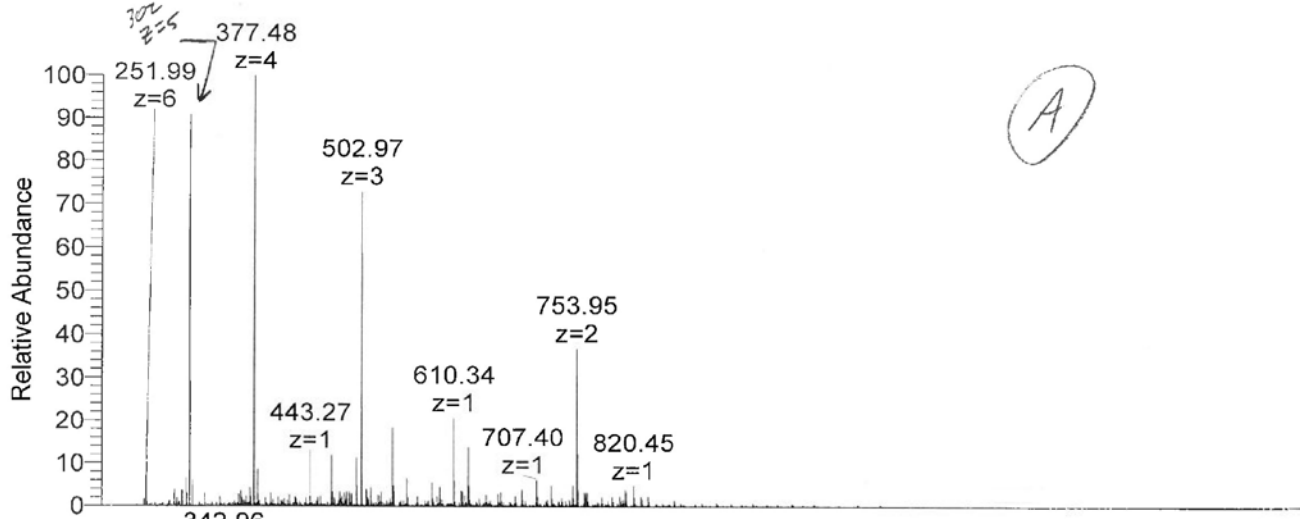
C:\Xcalibur...\MarApr\Peptide132a 3/25/2014 12:24:05 AM
Peptide #132 3/24/14 1355 add 100ul then 2 into 100 shoot 0.2

RT: 0.00 - 55.00

NL:
1.03E8
TIC MS
Peptide13
a

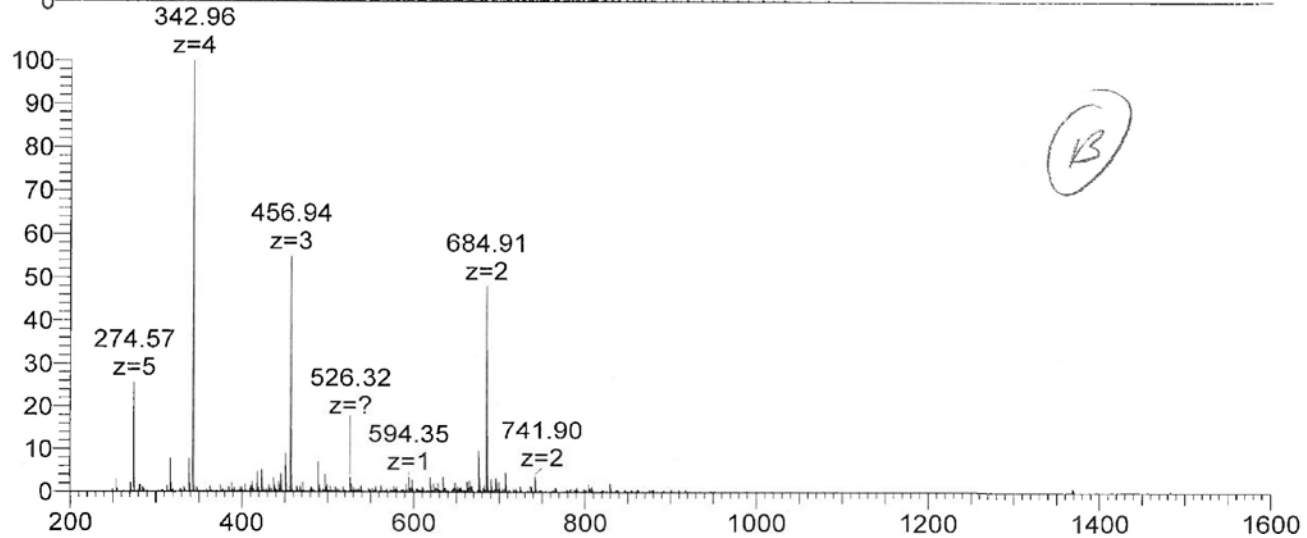
176





NL: 8.14E5
Peptide132a#1281-
1317 RT: 19.24-19.3
AV: 37 T: FTMS + p
NSI Full ms
[200.00-2000.00]

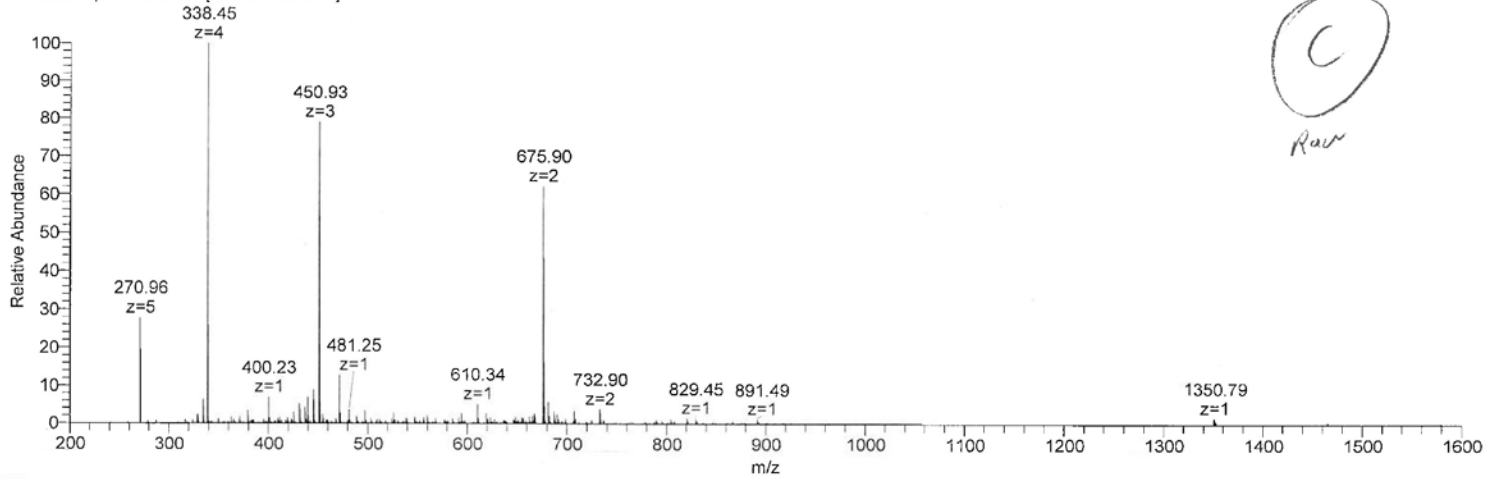
177



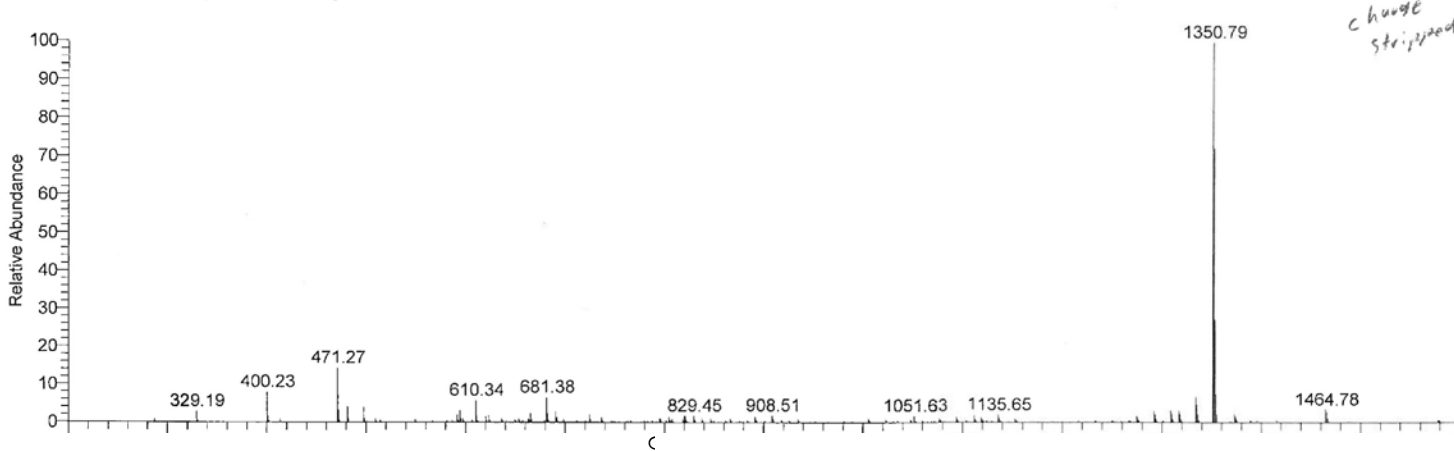
NL: 1.84E6
Peptide132a#1336-
1362 RT: 19.51-19.6
AV: 27 T: FTMS + p
NSI Full ms
[200.00-2000.00]

C:\Xcalibur...MarApr\Peptide132a 3/25/2014 12:24:05 AM
Peptide #132 3/24/14 1355 add 100ul then 2 into 100 shoot 0.2

Peptide132a #1376-1413 RT: 19.65-19.75 AV: 38 NL: 8.40E6
T: FTMS + p NSI Full ms [200.00-2000.00]



Peptide132a_XT_00001_MHp_#1 RT: 1.00 AV: 1 NL: 7.39E6
T: FTMS + p NSI Full ms [200.00-2000.00]



178

APPENDIX B: 22-UNIT SL ARRAY

22-UNIT SL ARRAY USED FOR PROSTATE TISSUE INVESTIGATION

SL Name	Amino acid sequence and modifications	#R	#2-PBA	#Ph
SL1	Ac-RG(Dab)*VTF(Dab)*RBBRM	3	2	3
SL2	Ac-RT(Dab)*RFL(Dab)*VBBRM	3	2	3
SL3	Ac-RS(Dab)*VTT(Dab)*RBBRM	3	2	2
SL4	Ac-RR(Dab)*TQT(Dab)*QBBRM	3	2	2
SL5	Ac-RA(Dab)*TRV(Dab)*VBBRM	3	2	2
SL6	Ac-RT(Dab)*NRN(Dab)*FBBRM	3	2	3
SL7	Ac-RS(Dab)*YFT(Dab)*QBBRM	2	2	4
SL8	Ac-RT(Dab)*YGN(Dab)*NBBRM	2	2	3
SL9	Ac-RT(Dab)*YQV(Dab)*ABBRM	2	2	3
SL11	R*L(Dab)*YLT(Dab)*RBBRM	3	3	4
SL11-R8A	R*L(Dab)*YLT(Dab)*ABBRM	2	3	4
SL11-T6A	R*L(Dab)*YLA(Dab)*RBBRM	3	3	4
SL11-Y4A	R*L(Dab)*ALT(Dab)*RBBRM	3	3	3
SL11-Y4F	R*L(Dab)*FLT(Dab)*RBBRM	3	3	4
SL11-T6Abu	R*L(Dab)*YL(Abu)(Dab)*RBBRM	3	3	4
SL5-R1,5,11A	Ac-AA(Dab)*TAV(Dab)*VBBAM	0	2	2
SL5-Dab	Ac-RA(Dab)TRV(Dab)VBBRM	3	0	0
SL5-RRAc	Ac-RRRA(Dab)*TRV(Dab)*VBBRM	5	2	2
4R	Ac-RAARAARA-BBRM	4	0	0
5R	Ac-RARARARA-BBRM	5	0	0
SL15	NH ₂ -RT(Dab)*RGG(Dab)*TBBRM	3	2	2
SL16	R*R(Dab)*AYR(Dab)*YBBRM	4	3	5

R* indicates N-terminal Arginine residue having 2-fprmyl phenyl boronic acid attachment Ac indicates acylated N-terminus
NH₂ indicates free N-terminus

APPENDIX C: TOP SEVEN SLS FOR ALL TISSUES TOGETHER

TABLE SHOWING TOP SEVEN SLS THAT CLASSIFY 13 DIFFERENT CELL LINES FROM DIFFERENT TISSUES.

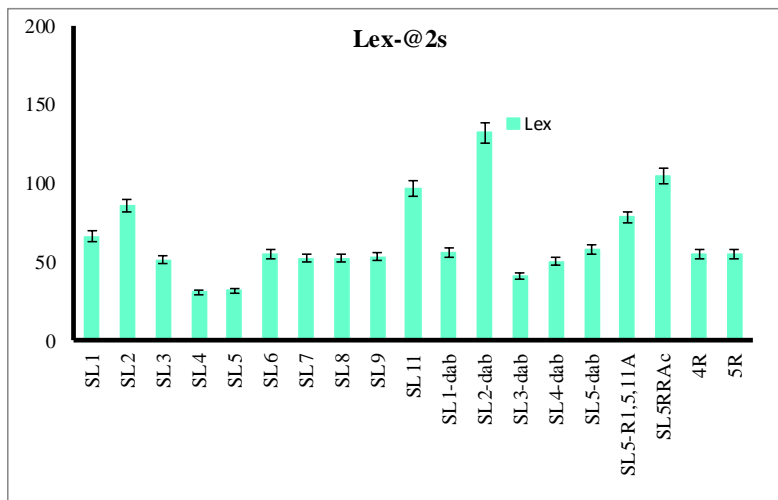
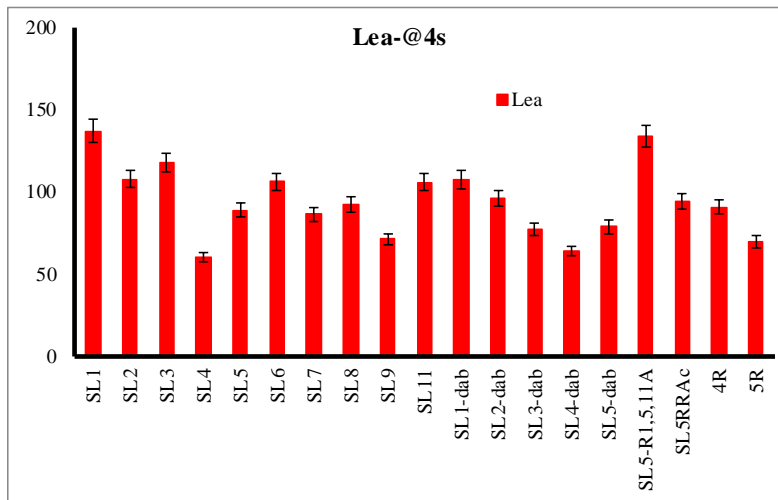
Table showing top 7 (out of 20) SLS that significantly classified 13 cell lines (from colon prostate and breast) according to the cell line metastatic potential. F-scores are used as statistical markers of important classifier. (SLS with F-score in single decimal place and below were not considered significant).

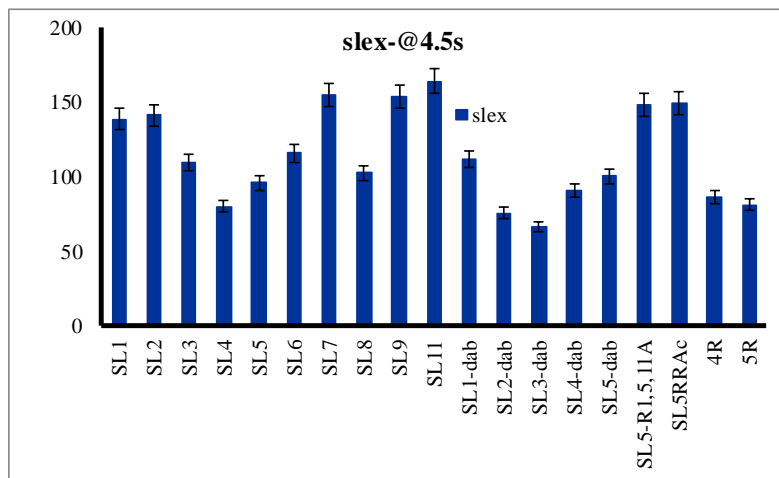
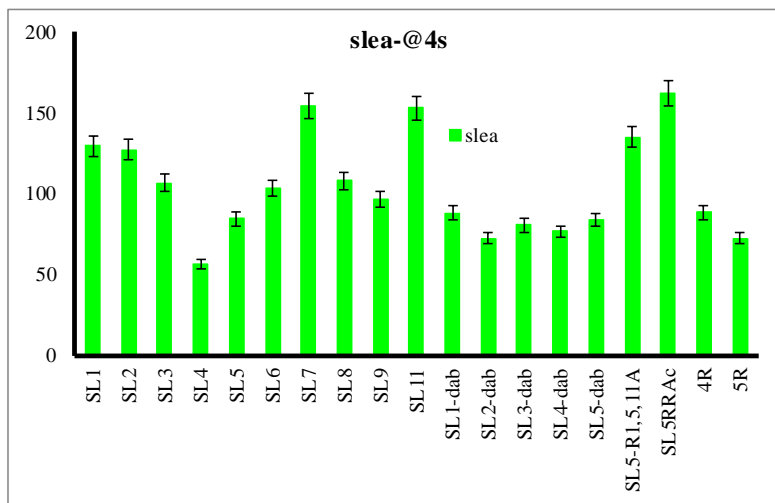
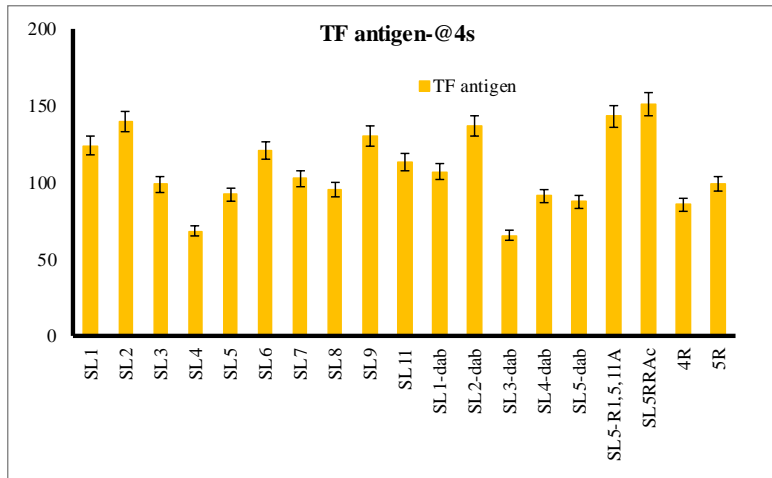
SL Name	F-Scores	Number of positive charges (from R and free Dab)	Number of Ph rings (from PBA and peptides)
SL4	64.09	3	2
4R	26.13	4	0
SL5-Dab	21.28	5	0
SL5	14.88	3	2
SL11-Y4A	14.66	3	3
5R	14.40	5	0
SL11-R8A	10.41	2	4

APPENDIX D: BAR GRAPHS OF OTHER FIVE CAGS WITH 20 SLS

BAR-GRAPHS OUTPUTS OF OTHER FIVE CANCER-ASSOCIATED GLYCANS

WITH 20 SLS





Raw average fluorescent intensity values. Standard deviation error bars. Exposure times used for each CAG were different and mentioned above in title of each bar-graph. Colors correspond to LDA color scheme in **Figure 4.5**.

APPENDIX E: BEAD INTENSITY BAR GRAPH FOR SL17 AND SL20

INDIVIDUAL BEAD INTENSITY BAR GRAPH FOR SL17 AND SL20.

