

2017

# Assessing Aberrant Glycosylation With Synthetic Lectins To Detect And Stage Prostate Cancer

Anna Veldkamp  
*University of South Carolina*

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

 Part of the [Chemistry Commons](#)

---

## Recommended Citation

Veldkamp, A. (2017). *Assessing Aberrant Glycosylation With Synthetic Lectins To Detect And Stage Prostate Cancer*. (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/etd/4005>

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](mailto:dillarda@mailbox.sc.edu).

ASSESSING ABERRANT GLYCOSYLATION WITH SYNTHETIC LECTINS TO  
DETECT AND STAGE PROSTATE CANCER

by

Anna Veldkamp

Bachelor of Science  
Georgia Southern 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

Caryn Outten, Chair, Examining Committee

Andrew Greytak, Committee Member

Marjorette Pena, Committee Member

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

© Copyright by Anna Veldkamp, 2016

All Rights Reserved.

## DEDICATION

This work is dedicated to my parents, Kenneth and Ruth Veldkamp, who are the strongest and most wonderful people I know. Without your guidance, love, and support, I would not have been able to finish this Ph.D.

## ACKNOWLEDGMENTS

Those of you who know me will know that this is the hardest part for me. There is no way I could ever show the gratitude and appreciation I have for everyone that has helped me throughout this degree (and life) but I will try to do my best.

To God, thank you for blessing me with the life I have. Thank you for all of the people mentioned in this acknowledgment and all of the wonderful things you have placed before me. It is through you that all things are possible.

Mom and Dad, thank you for always believing in me and never letting me give up. You have made me the person I am today. Thank you for loving me even with all of my mistakes and faults. I can never express my gratitude for everything you have done to make my life what it has been.

To Leah, thank you for showing me the love only a sister could. You have shown me that love can exist between two people no matter their differences or disagreements. You have always been the strong support I needed. I am amazed by your tenaciousness. You have taught me to always stay firm with what I believe and not let anyone change it. Thank you to James for being the best brother-in-law I could have hoped for and thank you for loving and supporting me just like my sister does.

To Matthew, thank you for choosing to do life with me. You are the person that sees the craziest of emotional train wrecks and still chooses to love me. No matter what we go through, I am always thankful to go through it with you. Thank you for always reminding

me that God is the reason for everything. I love you so much more than I could ever express.

To my friends throughout my life that has made life fun and worth living (Courtney, Erica, Vinnie, Matt, Alicia, Tony, Martina, Steven, Jung), thank you. I will cherish the good times always. Shawna, thank you for being my “ride or die.” You were my ray of sunshine through the darkest part of life. I wouldn’t have made it through without you.

To my lab members, thank you. Min, Shakena, Xiaoning, and Matt, thank you for your willingness to help me and teach me how to be a graduate student. Vinnie, I am so thankful we joined the same lab. It was less scary being new with you. To the girls (Kathy, Erin, and Tanya), thank you for putting up with my unwarranted mood swings. It is our little group of 4 (withstanding two to three hour subgroup meetings and lab “cleanups”) that I will remember most about lab. Kathy, thank you for your support. You always gave the comforting words of encouragement I needed through my candidacy “hurdles.” Tanya, thank you for being as sweet and supportive as you are and thank you for always finding food with me.

To Erin, thank you is not enough. I have never met someone who complements/completes the person I am better than you. Thank you for being whatever I needed (a rock, a shoulder to cry on, a planner, a cook, a statistician, a drinking buddy, an aunt to my fur babies) without me even telling you and for being a pseudo-adult with me. You are my second half and graduate school would not have been possible without you. God knew what he was doing when he placed you in my life. I never want to forget how seamless teaching is with you and I never want our telepathic abilities to cease.

To my committee members, thank you for always pushing me to be my best and for always having an open door policy. I am so thankful that each of you said yes when I asked for you to be on my committee.

And last but definitely not least, to Dr. Lavigne, thank you. Thank you for putting up with all of the tears. I'm sure I owe you many boxes of tissues. No other professor would have put up with all of the emotional breakdowns that you have and I cannot be more thankful that you did. You invited me to join your group and took a chance on me knowing that I didn't know anything (735). You have not only been my research advisor; you have also been a life coach, mentor, and friend. If I can say I know any chemistry at all, it is because you taught it to me. Thank you for everything and thank you for teaching me the most important question in life...so what?

## ABSTRACT

Aberrant glycosylation takes place during the onset of cancer and throughout the progression of the disease allowing certain glycans to serve as biomarkers<sup>1</sup>. Boronic acid functionalized synthetic lectins (SLs) have been shown to bind selectively and cross-reactively to certain glycoproteins through covalent yet reversible interactions between the boronic acids and 1,2- or 1,3-diols found on saccharides<sup>2</sup>. This project focuses on 1) data acquisition optimization of an existing SL array; 2) the identification, isolation, and characterization of new SLs that interact with proteins and glycoproteins targeting the prostate and prostate cancer; and 3) the utilization of new SLs in cross-reactive sensor arrays to distinguish between prostate cells with differing metastatic potential. This differentiation is based on the ability of the SLs to detect the subtle glycan changes in diseased cells. The results presented in this dissertation highlight the immense potential of SL arrays as tools in cancer diagnostics.



## TABLE OF CONTENTS

DEDICATION .....	iii
ACKNOWLEDGMENTS .....	iv
ABSTRACT .....	vii
LIST OF TABLES .....	xi
LIST OF FIGURES .....	xii
LIST OF ABBREVIATIONS .....	xviii
CHAPTER 1 : INTRODUCTION AND BACKGROUND.....	1
1.0 ABSTRACT .....	1
1.1 CANCER STATISTICS.....	1
1.2 CANCER DETECTION AND DIAGNOSIS .....	1
1.3 ABERRANT GLYOSYLATION .....	3
1.4 CANCER ASSOCIATED GLYCANS .....	4
1.5 DETECTION OF GLYCANS.....	6
1.6 BORONIC ACID-DIOL INTERACTION.....	6
1.7 BORONIC ACID FUNCTIONALIZED SENSORS .....	8
1.8 BORONIC ACID FUNCTIONALIZED SYNTHETIC LECTINS .....	11
1.9 THE NEED FOR NEW SLs.....	14
CHAPTER 2 : LIBRARY SCREENING.....	15
2.1 LIBRARY SYNTHESIS .....	15
2.2 MS/MS SEQUENCING.....	17

2.3 EDMAN DEGREDEATION.....	17
2.4 SCREENING AGAINST MEMBRANE EXTRACTED GLYCOPROTEINS .....	20
2.5 DUAL DYE SCREENING .....	25
2.6 RESYNTHESIZING SL10-14 .....	27
2.7 CONCLUSION .....	27
2.8 EXPERIMENTAL METHODS .....	28
CHAPTER 3 : PROSTATE SELECTED SYNTHETIC LECTINS .....	33
3.1 CROSS-REACTIVITY WITH PROSTATE CELL LINES .....	33
3.2 SL10-14 PROSTATE CELL LINE DISCRIMINATION .....	34
3.3 PROSTATE CANCER VS. COLON CANCER.....	37
3.4 THE UTILITY OF SL11 .....	37
3.5 TWO PBA VS. THREE PBA .....	38
3.6 CONCLUSION .....	39
3.7 EXPERIMENTAL METHODS .....	41
CHAPTER 4 : OPTIMIZATION OF DATA ACQUISITION .....	46
4.1 SL ARRAY DATA ACQUISITION.....	46
4.2 OBTAINING DATA FOR MATLAB .....	46
4.3 MATLAB .....	47
4.4 REGION OF INTEREST .....	48
4.5 COLOR SPACES .....	49
4.6 CONCLUSION .....	50
4.7 EXPERIMENTAL METHODS .....	51
CHAPTER 5 CONCLUSION .....	55

REFERENCES .....	57
APPENDIX A : SL IMAGING MATLAB CODE .....	60
APPENDIX B : EDMAN DEGRADATION SEQUENCING .....	76

## LIST OF TABLES

Table 1.1 Common overexpression patterns of cancer associated glycans on malignant tissues <sup>18</sup> .....	5
Table 1.2 Sequences of SLs screened against purified glycoproteins PSM and OVA. ....	12
Table 2.1 SL10-14 sequences, selectivity and the analyte it was screened against. ....	24
Table 3.1 Prostate cell lines with metastatic potential listed.....	33
Table 3.2 Prostate and colon cell lines with metastatic potential listed. ....	37
Table 3.3 Classification matrix for SL11 with prostate cell lines generated by LDA. ....	38
Table 4.1 SL sequences used in colon cancer array. ....	46
Table 4.2 Colon and breast cell lines with metastatic potential listed.....	46

## LIST OF FIGURES

- Figure 1.1 Aberrant Glycosylation. After the onset of cancer, normal glycan structures (the green chains on the left) are altered. The glycans can be over, under, or newly expressed, represented in the red and yellow chains. ....4
- Figure 1.2 Common cancer associated glycan structures. ....4
- Figure 1.3 Interaction between boronic acids and 1,2- or 1,3- cis diols. ....6
- Figure 1.4 The equilibrium between phenylboronic acid and a diol. ....7
- Figure 1.5 An aminomethyl group ortho to phenylboronic acid facilitates binding with diols at physiological pH. ....8
- Figure 1.6 Structures of boronic acid functionalized sensors used for sugar sensing. (a) first fluorescent probe, (b) Intramolecular charge transfer (ICT) sensors, (c-e) Photoinduced electron transfer (PET) sensors, (f-g) electrochemical sensors. ....9
- Figure 1.7 Cartoon 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. ....12
- Figure 1.8 SL cross-reactive sensor array discriminating seven colon cell types. A) The two-dimensional (2-D) LDA score plot of the SL array discriminating seven colon cell types. The green curved lines highlight the separation of healthy, cancerous/non-metastatic, and cancerous/metastatic cell types. B) Leave-one-out cross validation classification matrix for the discrimination of the seven colon cell types. Overall classification accuracy was 92% with a classification error of 8%. The expected chance accuracy of the SL array was 14.3%. ....13
- Figure 1.9 Discrimination of seven colon cell types by metastatic potential. A) 2-D LDA score plot of the SL cross-reactive array discriminating the seven colon cell types healthy, cancerous/non-metastatic, and cancerous/metastatic. B) 2-D LDA score plot of the SL cross-reactive array ungrouping the cancerous/non-metastatic cell types: HCT116, CT-26, and HT-29. C) 2-D LDA score plot of the SL cross-reactive array ungrouping the cancerous/metastatic cell types: CT-26-F1, CT-26-FL3, and LoVo. ....14
- Figure 2.1 Split and pool combinatorial methods used to generate SL peptide libraries. .16

Figure 2.2 Mechanism for Edman Degradation. The N-terminus amino group reacts with phenyl isothiocyanate under basic conditions. The resulting cyclic phenylthiocarbamoyl derivative is then cleaved under strong acidic conditions to produce an anilinothiazolinone derivative, which is selectively extracted with an organic solvent. The derivative is then converted to phenylthiohydrantoin (PTH) under strong acidic conditions. ....	18
Figure 2.3 Fmoc deprotection mechanism. Under basic conditions, the fluorenyl proton is deprotected to which produces the deprotected amino group, dibenzofulvene, and carbon dioxide. ....	19
Figure 2.4 Orbitrap LC chromatogram of SL11. ....	20
Figure 2.5 Orbitrap MS spectrum of SL11. The 1751 peak correlates with the total weight of SL11 with 3 PBA groups. ....	20
Figure 2.6 Workflow for screening a SL library to identify hits. Prostate cells are cultured and the membrane proteins and glycoproteins were extracted with Qiagen Plasma Membrane Protein Kit. After the extraction, the sample was labeled with fluorescein and then incubated with the SL library. The unbound fluorescent sample was washed away and library resin that was fluorescent were identified and isolated for sequencing. ....	22
Figure 2.7 SL10 hit located in the top left of the frame. This SL was screened against F-PC3. ....	23
Figure 2.8 SL13 hit located in the top right of the frame. This SL was screened against F-RWPE-1. ....	23
Figure 2.9 SL14 hit located in the top right of the frame. This SL was screened against R-PC3. ....	23
Figure 2.10 Dual Dye Screening Process. In order to select SLs that were selective, a dual dye screening process was used. In this process, the membrane glycoproteins from two different cell lines, one non-cancerous and the other cancerous-metastatic, were labeled with fluorescent dyes of different colors. In this figure, the non-cancerous glycoproteins were labeled with fluorescein and the cancerous-metastatic glycoproteins were labeled with rhodamine. ....	25
Figure 2.11 SL11 hit. This hit was found by dual dye screening with F-RWPE-1 and R-PC3. The hit is located in the bottom middle part of the right frame. ....	26
Figure 2.12 SL12 hit. This hit was found by dual dye screening with F-RWPE-1 and R-PC3. The hit is located in the middle part of the left frame. ....	27
Figure 3.1 Binding patterns/ “Fingerprint patterns” of SL10-14 with prostate cell lines. ....	34

Figure 3.2 LDA plot of SL10-14 classified by prostate cell line. ....	35
Figure 3.3 LDA of SL10-14 classified by metastatic potential. ....	36
Figure 3.4 LDA of SL1, 3, 4, 5 and SL10-14 Prostate by metastatic potential. ....	36
Figure 3.5 LDA plot of SL10 2BA-SL14 2BA classified by metastatic potential. ....	39
Figure 4.1 LDA plot for discrimination of three colon cancer cell lines by metastatic potential. This analysis was based on data acquisition without an ROI. ....	48
Figure 4.2 LDA plot for discrimination of three colon cancer cell lines by metastatic potential. This analysis was based on data acquisition with an ROI. ....	49
Figure 4.3 LDA plot for discrimination of three colon cancer cell lines by metastatic potential. This analysis was based on data acquisition using only bright numbers. .50	
Figure 4.4 LDA plot for discrimination of three colon cancer cell lines by metastatic potential. This analysis was based on data acquisition using all three color spaces. 50	
Figure 5.1 SL10 AA 8. The sequence of SL10 is R-L-D*-A-R-S-D*-G-B-B-R-M. R is a fixed position in the SL library. ....	77
Figure 5.2: SL10 AA 7. The sequence of SL10 is R-L-D*-A-R-S-D*-G-B-B-R-M. L is a random position in the SL library. ....	78
Figure 5.3: SL10 AA 6. The sequence of SL10 is R-L-D*-A-R-S-D*-G-B-B-R-M. D* is a fixed position in the library and is not recognized by Edman Degradation. The fixed position of this AA identifies where it falls in the SL sequence. It is also important to note that while L does show in the cycle of the sequence, the amount decreases from the cycle before. ....	79
Figure 5.4: SL10 AA 5. The sequence of SL10 is R-L-D*-A-R-S-D*-G-B-B-R-M. A is a random position in the library. ....	80
Figure 5.5: SL10 AA 4. The sequence of SL10 is R-L-D*-A-R-S-D*-G-B-B-R-M. R is a random position in the library. ....	81
Figure 5.6: SL10 AA 3. The sequence of SL10 is R-L-D*-A-R-S-D*-G-B-B-R-M. S is a random position in the library. ....	82
Figure 5.7: SL10 AA 2. The sequence of SL10 is R-L-D*-A-R-S-D*-G-B-B-R-M. D* is a fixed position in the library. The fixed position of this AA identifies where it falls in the SL sequence. ....	83

Figure 5.8: SL10 AA 1. The sequence of SL10 is R-L-D*-A-R-S-D*-G-B-B-R-M. G is a random position in the library. ....	84
Figure 5.9: SL11 AA 8. The sequence of SL11 is R-L-D*-Y-L-T-D*-R-B-B-R-M. R is a fixed position in library. ....	85
Figure 5.10: SL11 AA 7. The sequence of SL11 is R-L-D*-Y-L-T-D*-R-B-B-R-M. L is a random position in library. ....	86
Figure 5.11: SL11 AA 6. The sequence of SL11 is R-L-D*-Y-L-T-D*-R-B-B-R-M. D* is a fixed position in library. ....	87
Figure 5.12: SL11 AA 5. The sequence of SL11 is R-L-D*-Y-L-T-D*-R-B-B-R-M. Y is a random position in library. ....	88
Figure 5.13: SL11 AA 4. The sequence of SL11 is R-L-D*-Y-L-T-D*-R-B-B-R-M. L is a random position in library. ....	89
Figure 5.14: SL11 AA 3. The sequence of SL11 is R-L-D*-Y-L-T-D*-R-B-B-R-M. T is a random position in library. ....	90
Figure 5.15: SL11 AA 2. The sequence of SL11 is R-L-D*-Y-L-T-D*-R-B-B-R-M. D* is a random position in library. ....	91
Figure 5.16: SL11 AA 1. The sequence of SL11 is R-L-D*-Y-L-T-D*-R-B-B-R-M. R is a random position in library. ....	92
Figure 5.17: SL12 AA 8. The sequence of SL12 is R-L-D*-G-F-Y-D*-Q-B-B-R-M. R is a fixed position in library. ....	93
Figure 5.18: SL12 AA 7. The sequence of SL12 is R-L-D*-G-F-Y-D*-Q-B-B-R-M. L is a random position in library. ....	94
Figure 5.19: SL12 AA 6. The sequence of SL12 is R-L-D*-G-F-Y-D*-Q-B-B-R-M. D* is a fixed position in library. ....	95
Figure 5.20: SL12 AA 5. The sequence of SL12 is R-L-D*-G-F-Y-D*-Q-B-B-R-M. G is a random position in library. ....	96
Figure 5.21: SL12 AA 4. The sequence of SL12 is R-L-D*-G-F-Y-D*-Q-B-B-R-M. F is a random position in library. ....	97
Figure 5.22: SL12 AA 3. The sequence of SL12 is R-L-D*-G-F-Y-D*-Q-B-B-R-M. Y is a random position in library. ....	98



Figure 5.23: SL12 AA 2. The sequence of SL12 is R-L-D*-G-F-Y-D*-Q-B-B-R-M. D* is a fixed position in library. ....	99
Figure 5.24: SL12 AA 1. The sequence of SL12 is R-L-D*-G-F-Y-D*-Q-B-B-R-M. Q is a random position in library. ....	100
Figure 5.25: SL13 AA 8. The sequence of SL13 is R-T-D*-G-L-A-D*-V-B-B-R-M. R is a fixed position in library. ....	101
Figure 5.26: SL13 AA 7. The sequence of SL13 is R-T-D*-G-L-A-D*-V-B-B-R-M. T is a random position in library. ....	102
Figure 5.27: SL13 AA 6. The sequence of SL13 is R-T-D*-G-L-A-D*-V-B-B-R-M. D* is a fixed position in library. ....	103
Figure 5.28: SL13 AA 5. The sequence of SL13 is R-T-D*-G-L-A-D*-V-B-B-R-M. G is a random position in library. ....	104
Figure 5.29: SL13 AA 4. The sequence of SL13 is R-T-D*-G-L-A-D*-V-B-B-R-M. L is a random position in library. ....	105
Figure 5.30: SL13 AA 3. The sequence of SL13 is R-T-D*-G-L-A-D*-V-B-B-R-M. A is a random position in library. ....	106
Figure 5.31: SL13 AA 2. The sequence of SL13 is R-T-D*-G-L-A-D*-V-B-B-R-M. D* is a fixed position in library. ....	107
Figure 5.32: SL13 AA 1. The sequence of SL13 is R-T-D*-G-L-A-D*-V-B-B-R-M. V is a random position in library. ....	108
Figure 5.33: SL14 AA 8. The sequence of SL14 is R-Y-D*-R-A-S-D*-V-B-B-R-M. R is a fixed position in library. ....	109
Figure 5.34: SL14 AA 7. The sequence of SL14 is R-Y-D*-R-A-S-D*-V-B-B-R-M. Y is a random position in library. ....	110
Figure 5.35: SL14 AA 6. The sequence of SL14 is R-Y-D*-R-A-S-D*-V-B-B-R-M. D* is a fixed position in library. ....	111
Figure 5.36: SL14 AA 5. The sequence of SL14 is R-Y-D*-R-A-S-D*-V-B-B-R-M. R is a random position in library. ....	112
Figure 5.37: SL14 AA 4. The sequence of SL14 is R-Y-D*-R-A-S-D*-V-B-B-R-M. A is a random position in library. ....	113

Figure 5.38: SL14 AA 3. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. S is a random position in library. .... 114

Figure 5.39: SL14 AA 2. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. D\* is a fixed position in library. .... 115

Figure 5.40: SL14 AA 1. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. V is a random position in library. .... 116

## LIST OF ABBREVIATIONS

2-D	Two Dimensional
AA	Amino Acid
BSM	Bovine Submaxillary Mucin
DMF	N,N-Dimethylformamide
Dab	Diaminobutyric Acid
Fmoc	Fluorenylmethyloxycarbonyl
HBTU	O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate
H <sub>2</sub> O	Water
ivDde	1-(4,4-Dimethyl-2,6-dioxocyclo-hexylidene)-3-methylbutyl
MALDI	Matrix Assisted Laser Desorption/Ionization
MeOH	Methanol
MS/MS	Mass Spectrometry/Mass Spectrometry
OVA	Ovalbumin
PBA	Phenylboronic Acid
PSA	Prostate Specific Antigen
PSM	Porcine Submaxillary Mucin
SL	Synthetic Lectin
TIS	Triisopropylsilane

# CHAPTER 1:

## INTRODUCTION AND BACKGROUND

### 1.0 ABSTRACT

Cancer, the uncontrolled growth of abnormal cells in the body, is one of the leading causes of death not only in the United States but also worldwide. Cancer diagnostics that can detect the disease in early stages have become a necessary means of increasing a patient's survival rate. Changes in glycosylation take place during the onset of cancer and throughout the progression of the disease, which allows these glycans to serve as biomarkers. This chapter will describe how boronic acid functionalized synthetic lectins (SLs) have been developed as glycan sensors and how cross-reactive sensor arrays can be a useful tool in designing a novel cancer diagnostic.

### 1.1 CANCER STATISTICS

According to the National Cancer Institute, it is estimated that there will be 1,685,210 new cases of cancer and 595,690 people will die from the disease in 2016. In 2013, approximately 14,140,254 people were living with cancer of any site in the United States with the most common types of cancer being breast, lung and bronchus, prostate, colon and rectum, and bladder cancer<sup>3</sup>.

### 1.2 CANCER DETECTION AND DIAGNOSIS

Early detection and treatment of cancer has shown to increase survival rates in cancer patients and therefore makes the development of diagnostic tools imperative<sup>4</sup>.

Even though there are ways of diagnosing cancer, many of the procedures are associated with disadvantages.

One example of a “gold standard” cancer diagnostic is a colonoscopy, which is a procedure that uses a colonoscope to examine the colon for abnormal growths. While it has is the best way to visualize tumor growth, the test is expensive and invasive. Some studies have shown that the efficacy of the procedure depends where the cancer is located within the colon<sup>5-6</sup>. Colonoscopies can also cause potential health complications such as lacerations, bleeding, and infection in the colon<sup>7</sup>.

Most diagnostics are also associated with false positives and false negatives. False positives lead a patient to believe that they have cancer when they do not. The procedures that follow a primary diagnosis cause the patient emotional stress and more financial burden. For example, an exam that was previously used to diagnose prostate cancer, the prostate specific antigen (PSA) test, has a 75% false positive test result rate, meaning that only 25% of the men that undergo biopsies actually have cancer<sup>8</sup>. False negatives lead a patient to believe that they do not have cancer when they really do, preventing the patient from seeking medical attention sooner, which can lead to a more advanced cancer before treatment. Due to its inability to diagnose prostate cancer effectively, the PSA test is now being used as a means to monitor cancer progression in patients that have been diagnosed through systematic prostate biopsy under ultrasound guidance<sup>9</sup>.

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. Even though elevated levels of carcinoembryonic antigen (CEA) are

known to be associated with colon cancer, the CEA test is only 4% and 25% effective in the first and second stages of cancer, respectively<sup>10</sup>. This test was also discontinued as a screening test and utilized as a cancer progression monitor similar to the PSA test<sup>11</sup>. Relatively new tests such as the fecal immunochemical test (FIT) and the guaiac-based fecal occult blood test (gFOBT) have emerged as screening tools. These tests detect blood in stool that is present due to colon cancer. While these tests are less invasive than a colonoscopy, they also cannot detect new growth, which leads to false negatives<sup>12-14</sup>.

Clearly there is a dire need for new cancer screening tests. An ideal test would quick, inexpensive, and have a low amount of risk associated with it. The test would also need to have low false positive and false negative rates.

### 1.3 ABERRANT GLYOSYLATION

Proteins that have carbohydrate chains (glycans) attached to them via co- or post-translational glycosylation are glycoproteins. Over 70% of proteins are glycosylated<sup>15</sup>. Glycans are linked to proteins in two different ways: N-linked, where the glycan is attached through the side chain of an asparagine residue, or O-linked, where the glycan is attached through a serine or threonine residue<sup>16</sup>. During the onset of cancer, aberrant glycosylation takes place in which glycan structures are over, under, or neo-expressed (Figure 1.1). The expression of glycans not only changes on inter-membrane glycoproteins but also on glycoproteins that are secreted. These changes continue throughout the progression of the disease and therefore create specific biomarkers for non-cancerous and cancerous cells, which could be used to diagnose cancer<sup>17</sup>.

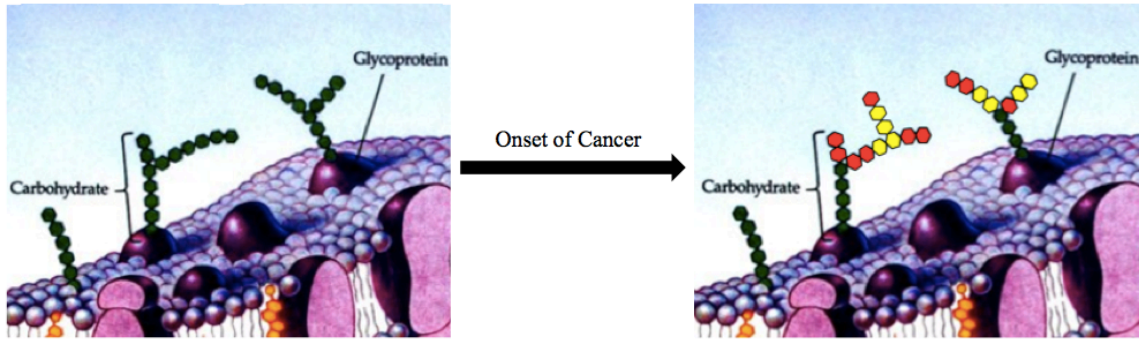


Figure 1.1 Aberrant Glycosylation. After the onset of cancer, normal glycan structures (the green chains on the left) are altered. The glycans can be over, under, or newly expressed, represented in the red and yellow chains.

#### 1.4 CANCER ASSOCIATED GLYCANS

The glycosylation changes that take place are due to changes in glycotransferase levels. These glycotransferase changes alter the composition of the glycan chains that are attached to proteins. Some common cancer associated glycans are found in Figure 1.2.

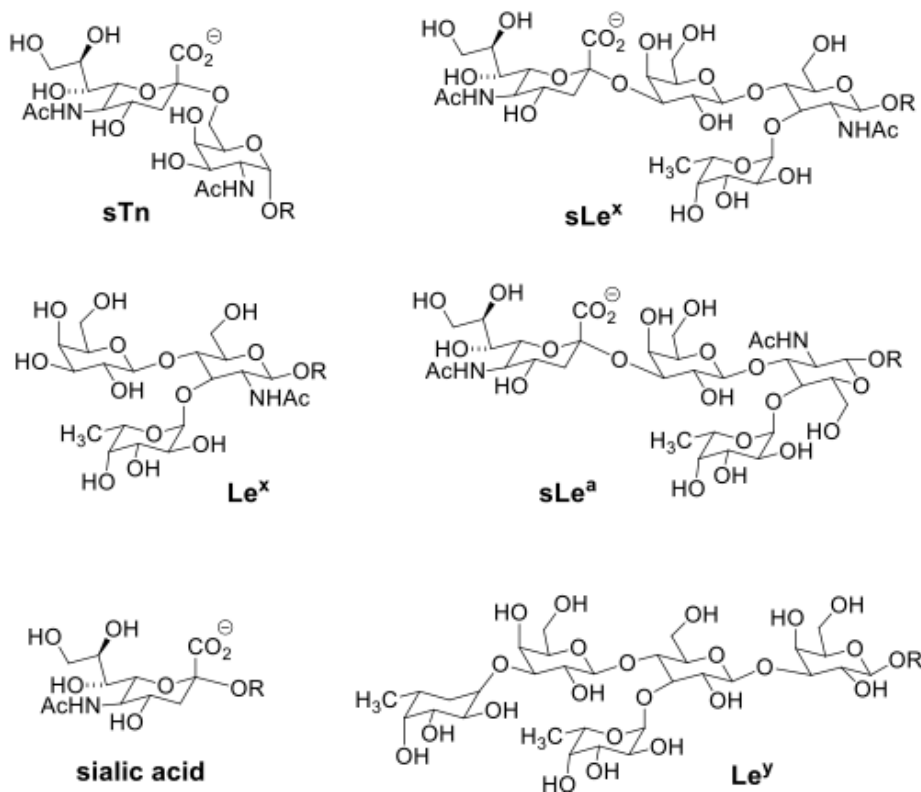


Figure 1.2 Common cancer associated glycan structures.

Many glycan structures have been found to be associated with certain cancer types (**Table 1.1**)<sup>18</sup>. It is important to note that not one type of cancer listed has the same combination of cancer-associated glycans, which creates a specific pattern for each cancer type. It is also important to add that some cancers have very similar patterns. For example breast and colon cancer have the same pattern except for the overexpression of polysialic acid in breast cancer. It is also interesting that breast and prostate cancer have very different patterns, differing by three cancer-associated glycans. The ability to detect the different patterns would provide a potential cancer diagnostic.

Table 1.1 Common overexpression patterns of cancer associated glycans on malignant tissues<sup>18</sup>.

Cancer glycan	Malignant tissue								
	Ovary	Pancreas	Blood	Breast	Colon	Brain	Prostate	Skin	Lung
sLe <sup>x</sup>		x		x	x				x
sLe <sup>a</sup>		x		x	x				x
sTn	x	x		x	x		x		x
TF	x			x	x		x		
Le <sup>y</sup>	x	x		x	x		x		x
GloboH	x	x		x	x		x		x
PSA		x	x	x		x			x
GD2			x			x		x	
GD3						x		x	
Fucosyl GM1									x
GM2	x	x	x	x	x	x	x	x	x

While many glycoproteins have been used to detect cancer, such as 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. These detection methods are



associated with high false positive and high false negative rates, therefore many of these diagnostics have become methods to monitor the disease after the disease has been confirmed<sup>10</sup>.

### 1.5 DETECTION OF GLYCANS

An ideal sensor that could detect the glycans and the changes occurring would have high specificity towards its target, stability, and low costs. Natural lectins are carbohydrate-binding proteins that have a high affinity to sugars and high specificity. These proteins could act as sensors in a cancer diagnostic but are costly and experimentally challenging<sup>19</sup>.

### 1.6 BORONIC ACID-DIOL INTERACTION

In order to circumvent the previously stated disadvantages of 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 can be found in Figure 1.3.

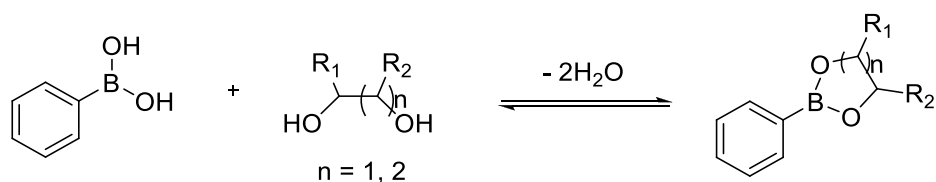


Figure 1.3 Interaction between boronic acids and 1,2- or 1,3- cis diols.

When using boronic acids in glycan sensors, it is important to understand what influences the boronic acid-diol interaction. The equilibrium between phenylboronic acid and a diol can be found in Figure 1.4.

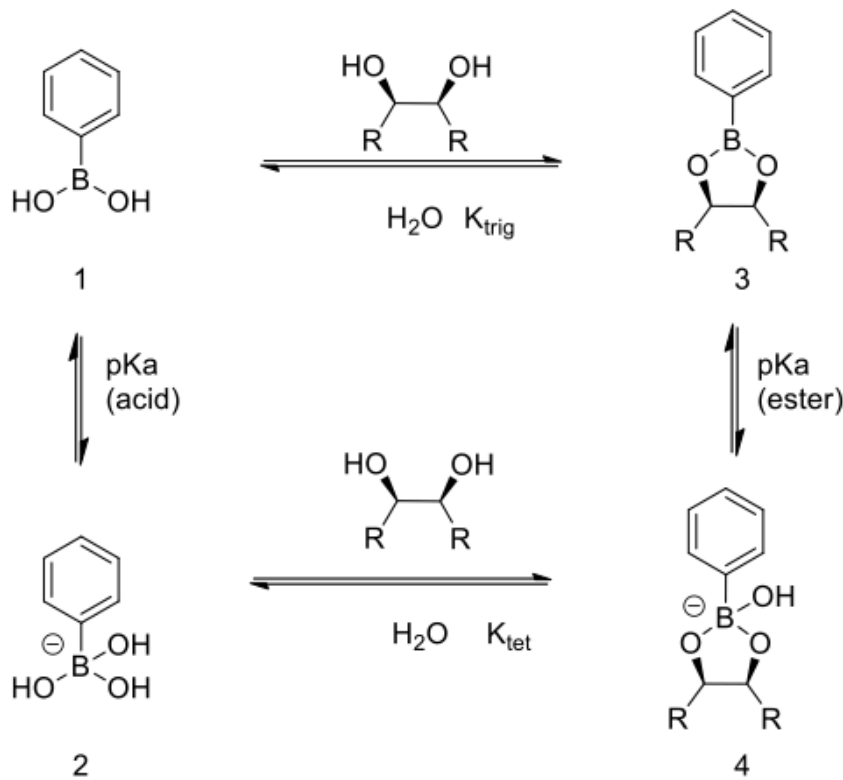


Figure 1.4 The equilibrium between phenylboronic acid and a diol.

The equilibrium is very dependent on pH. At higher pH, boronic acid becomes a boronate anion (in Figure 1.4, 1 → 2 or 3 → 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. This limits the use of boronic acid functionalized sensors at physiological pH<sup>20</sup>.

In order to use boronic acids on sensors at physiological pH, the Anslyn group investigated intramolecular boron-nitrogen interactions. They found that in neutral aqueous solutions, the incorporation of an ortho aminomethyl group to phenylboronic acid allowed boronic ester formation with diols<sup>21</sup>. The aminomethyl group interacts with the empty p orbital on the boron transforming the geometry into the favored tetrahedral conformation. The Anslyn group proposes two reasons for this boronate ester formation increase. First, the solvolysis of the boron is promoted because the nitrogen can act as a

Bronsted-Lowry base. Second, there is a weak dative bond between the nitrogen and the boron. Both of these explanations facilitate the boronate ester formation at physiological pH (Figure 1.5<sup>21</sup>).

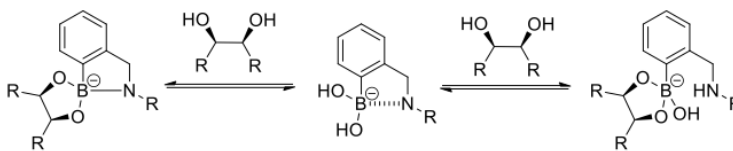


Figure 1.5 An aminomethyl group ortho to phenylboronic acid facilitates binding with diols at physiological pH.

### 1.7 BORONIC ACID FUNCTIONALIZED SENSORS

Many boronic acid functionalized sensors have been developed for different types of saccharide sensing with particular interest in D-glucose. As previously stated, sensing sugars can be valuable when trying to diagnose cancer and other diseases. The specific recognition of boronic acid for cis-diols provides selectivity in sensing. The structures of many different boronic acid functionalized sensors used in fluorescent, colorimetric, and electrochemical methods are in Figure 1.6.

Yoon and Czarnik used the first fluorescent probe using anthrylboronic acid (Figure 1.6a)<sup>22</sup>. When the boronic acid was bound to a sugar, the anionic boronate ester was formed and fluorescence intensity decreased by 40%. This decrease in fluorescence intensity was not extensively studied but it was thought that the boronate ester formation quenched the anthracene fluorescence. This sensor encouraged other researchers to work in this field of chemistry.

Intramolecular charge transfer (ICT) was also applied to develop fluorescent sensors. Substituted stilbenes containing boronic acid were synthesized by the Lakowicz group as sugar sensors (Figure 1.6b)<sup>23</sup>. The boron on the boronic acid acted as a Lewis

acid and the R group was an electron-donating group such as an amine or methoxy group. Therefore, electrons were transferred from an electron-donating group to the electron-accepting boron through a spacer fluorophore. 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.

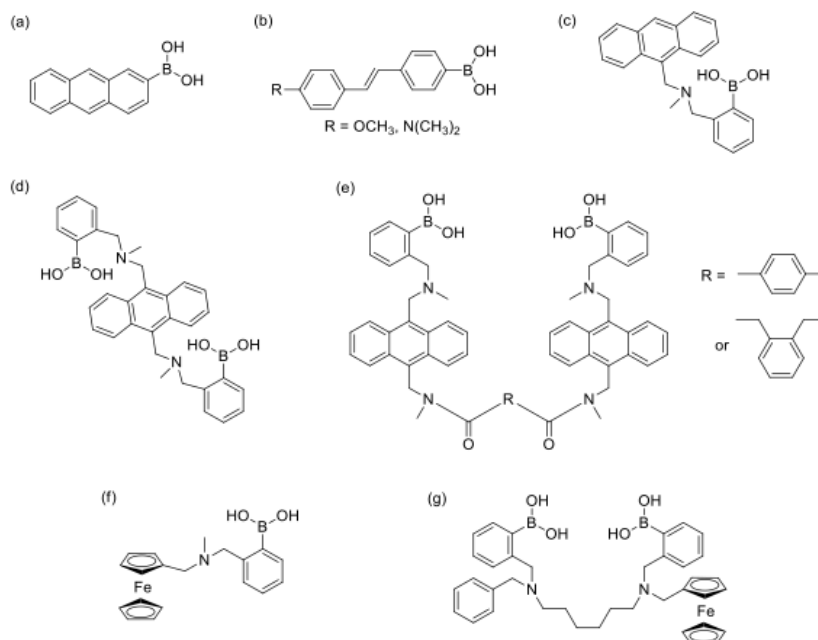


Figure 1.6 Structures of boronic acid functionalized sensors used for sugar sensing. (a) first fluorescent probe, (b) Intramolecular charge transfer (ICT) sensors, (c-e) Photoinduced electron transfer (PET) sensors, (f-g) electrochemical sensors.

Shinkai's group was the first to design PET sensors for sugars (Figure 1.6c)<sup>24</sup>. They used N-methyl-o-(aminomethyl)phenylboronic acid as the recognition motif and connected to a fluorophore anthracene through a methylene spacer. The amine in close proximity to the boronic acid allows for the interaction between the boronic acid and diols at neutral pH. Also, the amine-boron interaction dictates the molecule's ability to fluoresce. Without the sugar binding, the sensor is in an "off" state due to the nitrogen's lone pair. The lone pair quenches the fluorescence of anthracene. When binding the

sugar, the nitrogen coordinates with the boron and prevents fluorescence quenching by PET. These sensors exhibited some selectivity for sugars with it being the most selective for D-fructose. Unfortunately, the selectivity was altered when the amounts of boronic acid or the geometry around the boron was changed. Another example of an anthracene-based diboronic acid system is shown in Figure 1.6d<sup>25</sup>. This sensor had high selectivity for D-glucose over other sugars.

Wang's group developed a series of diboronic acid sensors using a similar strategy to the PET mechanism. This series had different linkers between two anthracene fluorophores (Figure 1.6e)<sup>26-27</sup>. They investigated these sensors' ability to selectively bind to cancer associated glycans. Wang's group found that the compounds with the para-benzene linker showed moderate selectivity for Sialyl Lewis X.

Other sensors that were developed were electrochemical sensors. James's group<sup>28</sup> attached electro-active ferrocene to monoboronic acid (Figure 1.6f) or diboronic acid (Figure 1.6g). Overall, the diboronic acid sensor was more selective for D-glucose and D-galactose compared to the monoboronic 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 differences in the sugar structures in glycans and glycoproteins in cancerous states create the need for sensors and therefore, boronic acids have been incorporated into many biosensors for their application in cancer diagnostics.

Boronolectins are the biosensors that have been developed in the cancer diagnostic area. The selectivity of these sensors has been greatly improved over the years.

One important feature of the sensors was the biocompatibility. In order to optimize these sensors, they needed to have good water solubility, low toxicity, and more stability than the other sugar 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 developed different sensors that contained quinolone and naphthalene. These structures improved water solubility and stability. When binding to glycans, the fluorescence of these sensors changed. The changes were specific to each glycan<sup>29-34</sup>.

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<sup>35</sup>.

## 1.8 BORONIC ACID FUNCTIONALIZED SYNTHETIC LECTINS

Previously, an array of boronic acid functionalized synthetic lectins (SLs) was isolated from a fixed-position library screening against purified glycoproteins PSM and OVA. The general sequence of the SLs was Ac-R-X-D\*-X-X-X-D\*-X-B-B-R-M-resin where X denotes a random AA position, D\* denotes Dab functionalized with PBA, and B denotes beta-alanine, which acts as a spacer between the resin and the peptide sequence. Four SLs were identified and sequenced using MS/MS (**Table 1.2**).

Table 1.2 Sequences of SLs screened against purified glycoproteins PSM and OVA.

SL Hit	Sequence
SL1	Ac-R-G-D*-V-T-F-D*-R-B-B-R-M-resin
SL3	Ac-R-S-D*-V-T-T-D*-R-B-B-R-M-resin
SL4	Ac-R-R-D*-T-Q-T-D*-Q-B-B-R-M-resin
SL5	Ac-R-A-D*-T-R-V-D*-V-B-B-R-M-resin

After the SLs were sequenced and resynthesized, the SLs were incubated with the purified glycoproteins to assess the selectivity/cross-reactivity of each SL. Unlike a “lock-and-key” approach, the SLs were not highly selective for one glycoprotein.

This is not a problem when using the SLs in an array because the SLs only need to have different affinities for a certain analyte, which produces a unique fingerprint pattern (Figure 1.7). Cross-reactive sensor arrays are similar to the mammalian sense of taste. It is not how one taste bud reacts with a type of food but how all of the taste buds react to form a pattern that the brain stores as a certain taste.

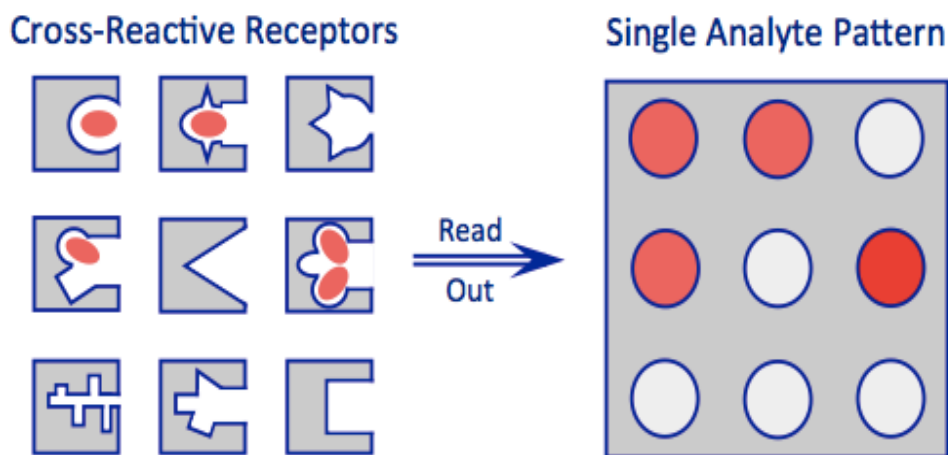


Figure 1.7 Cartoon of cross-reactive sensors/receptors arranged into a sensor array. In this figure, four of the receptors bind to a single analyte differently. The combined signals from the receptors create a unique pattern for the analyte.

Through linear discriminant analysis, previous lab members have used the four SL array to distinguish between seven colon cancer cell lines with 92% classification accuracy (Figure 1.8).

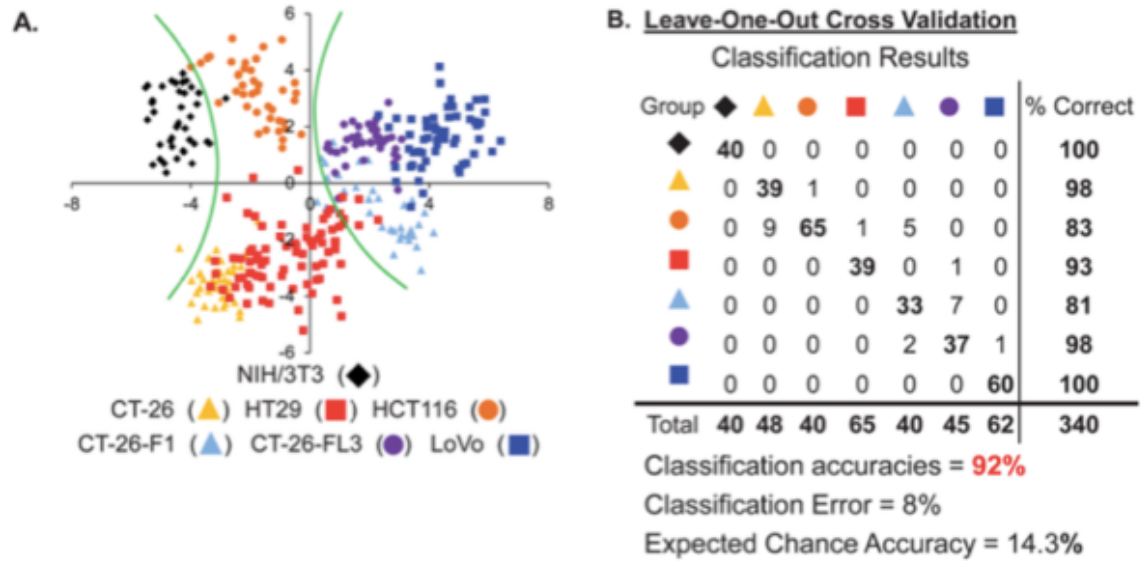


Figure 1.8 SL cross-reactive sensor array discriminating seven colon cell types. A) The two-dimensional (2-D) LDA score plot of the SL array discriminating seven colon cell types. The green curved lines highlight the separation of healthy, cancerous/non-metastatic, and cancerous/metastatic cell types. B) Leave-one-out cross validation classification matrix for the discrimination of the seven colon cell types. Overall classification accuracy was 92% with a classification error of 8%. The expected chance accuracy of the SL array was 14.3%.

While being able to distinguish cell lines is important, in thinking about a clinically relevant diagnostic there are two important initial questions: Does the patient have cancer? Has the cancer metastasized? With this in mind, the analysis was done to see if the array can differentiate based on metastatic potential (non-cancerous, cancerous low metastatic, and cancerous high metastatic). The array was also able to differentiate the seven cell lines by metastatic potential with 97% accuracy (Figure 1.9)<sup>2</sup>.



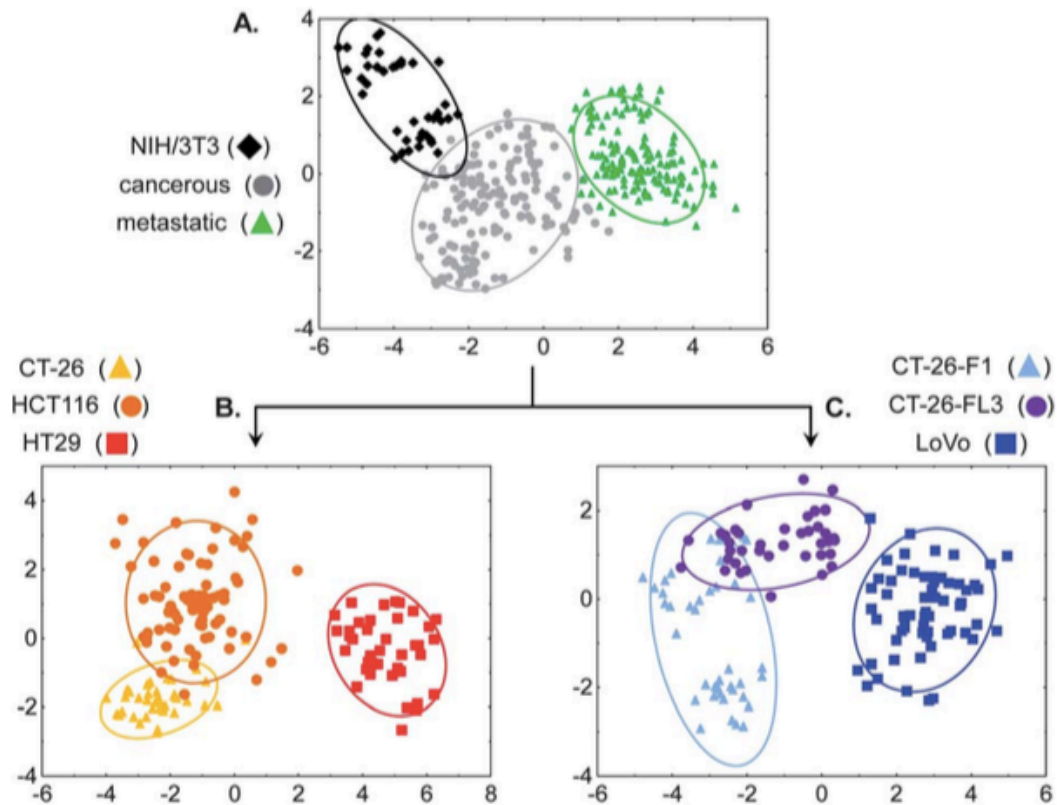


Figure 1.9 Discrimination of seven colon cell types by metastatic potential. A) 2-D LDA score plot of the SL cross-reactive array discriminating the seven colon cell types healthy, cancerous/non-metastatic, and cancerous/metastatic. B) 2-D LDA score plot of the SL cross-reactive array ungrouping the cancerous/non-metastatic cell types: HCT116, CT-26, and HT-29. C) 2-D LDA score plot of the SL cross-reactive array ungrouping the cancerous/metastatic cell types: CT-26-F1, CT-26-FL3, and LoVo.

### 1.9 THE NEED FOR NEW SLs

While the previous SLs show promise for a colon cancer diagnostic, there is a need for new screening methods that utilize proteins and glycoproteins that come from human sources. The SLs in the previous array were screened against purified glycoproteins that were isolated from animal sources. Also, the array's ability to discriminate other types of cancer had not been investigated.

## CHAPTER 2:

### LIBRARY SCREENING

#### 2.1 LIBRARY SYNTHESIS

Previous work was done to optimize the SL library screening against purified glycoproteins. The general peptide sequence that was found to have the most diversity with relatively easy sequencing ability when using Matrix Assisted Laser Desorption/Ionization (MALDI) was Ac-R-X-D\*-X-X-X-D\*-X-B-B-R-M-Resin, where the boronic acid moieties are fixed; D\* denotes phenylboronic acid (PBA) functionalized Diaminobutyric acid (Dab), B represents  $\beta$ -alanine, and X is a randomized amino acid. Dab is used in place of Lys because of its shorter side chain, which is thought to reduce background binding and increase interactions between the glycoproteins and the peptide backbone. O-aminomethyl PBA is incorporated because of its ability to bind to diols under physiological conditions. To create the peptide chain, standard Fluorenylmethyloxycarbonyl (Fmoc) protocols were used with O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate (HBTU) activation. The MRBB sequence was coupled to all beads. Methionine is the first AA because it provides a way to cleave the peptide from the resin using cyanogen bromide. Arginine was incorporated because it aids in MS/MS sequencing by integrating a positive charge and the  $\beta$ -alanine acts as spacers, which increases the distance between the remaining sequence and resin. PBA moieties were fixed to also aid in MS/MS sequencing. The randomized amino acids were coupled in a split-and-pool method.

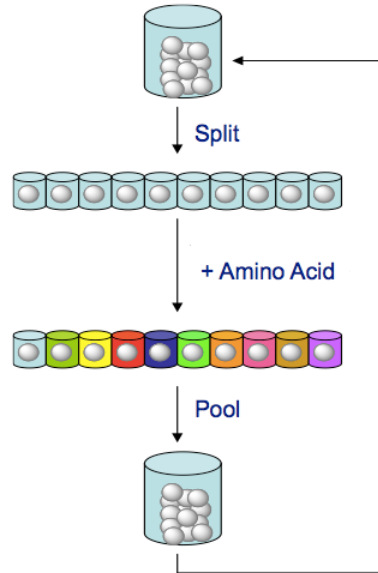


Figure 2.1 Split and pool combinatorial methods used to generate SL peptide libraries. For randomized AA positions in the SL library, the resin is separated equally into 11 tubes and a different activated AA is added to each tube. Once the AA is coupled, the resin from all the tubes is combined to deprotect the N-terminus Fmoc protecting group.

Of the 20 AAs, methionine, cysteine, isoleucine, proline, histidine, glutamic acid, aspartic acid, lysine, and tryptophan were excluded from the randomized positions. Methionine is used as the first amino acid in all of the SL sequences to enable peptide cleavage from the bead when treated with cyanogen bromide so methionine and cysteine are not used in the randomized amino acid positions to eliminate the chance of cutting the peptide sequence into pieces. Isoleucine is omitted to eliminate the chance of mistaking the amino acid for leucine when sequencing unknown SLs. Proline is excluded to avoid having SLs with turns. Histidine, glutamic acid, aspartic acid, and lysine were not used due to charged side chains. Tryptophan is not included because of its big and bulky side chain. The N-terminus arginine has the same benefits as the first arginine and the acetyl group is a cap that prevents reactions from taking place with the end of the sequence.

The PBA functional group was attached through a reductive amination reaction once the protecting group on the commercially available Dab was removed. The peptide was then treated with TFA to remove the remaining side-chain protecting groups.

## 2.2 MS/MS SEQUENCING

After the library was incubated with a fluorescently labeled analyte and hits were manually isolated, the SL was prepared for MS/MS sequencing. The analyte were washed away from the SL and the PBAs were removed as well. The SL peptide was then cleaved from the Tentagel resin.

Previous lab members used MALDI to sequence SLs. However, due to instrumentation complications and lab conditions, the hits isolated from screenings against the prostate analytes could not be sequenced. In attempts to still recover the sequence, another MS/MS technique, Orbitrap, was used. While the orbitrap has the ability to detect the small amount of peptide from one bead, it was still unable to sequence the cleave peptide from the hits.

## 2.3 EDMAN DEGRADATION

In order to sequence hits from library screenings against prostate analytes, an alternative method of peptide sequencing, Edman Degradation, was used. However, due to the experimental procedures, a few modifications were made to the SL library.

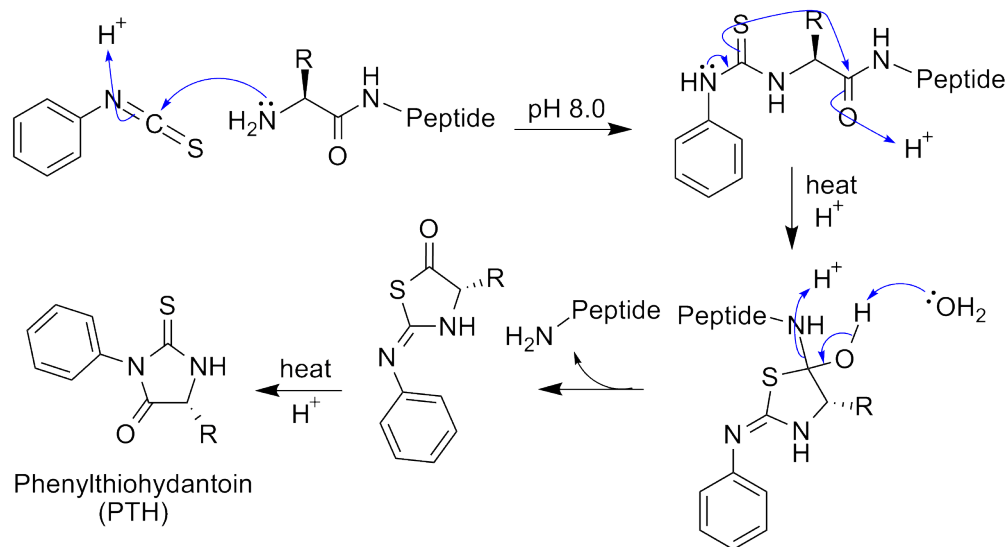


Figure 2.2 Mechanism for Edman Degradation. The N-terminus amino group reacts with phenyl isothiocyanate under basic conditions. The resulting cyclic phenylthiocarbamoyl derivative is then cleaved under strong acidic conditions to produce an anilinothiazolinone derivative, which is selectively extracted with an organic solvent. The derivative is then converted to phenylthiohydantoin (PTH) under strong acidic conditions.

In the first step of Edman Degradation, the N-terminus amine reacts with phenylisothiocyanate to form a phenylthiocarbamyl-peptide derivative. If the N-terminus is chemically modified with an acyl (Ac) group, the reaction cannot take place and therefore sequencing cannot take place. In the original peptide synthesizing procedure, the Ac protected the terminal amine from being functionalized with PBA. The synthetic method was modified so that after the last arginine was coupled to the resin, the Fmoc protecting group was not deprotected with 20% piperidine in DMF. This step was removed and the rest of the steps continued in the same order. The ivDde-protecting group on Dab was removed with 5% hydrazine in DMF and PBA was attached. The Fmoc-protecting group was then deprotected with 20% piperidine in DMF and acid labile groups were removed using the same methods stated above.

After screening the new library and sequencing hits against the prostate membrane extracted proteins and glycoproteins, the SLs were resynthesized using the same protocol for Edman Degradation sequencing. Orbitrap MS/MS was used to verify the peptide sequence of the resynthesized SL and the number of boronic acids attached to the peptide.

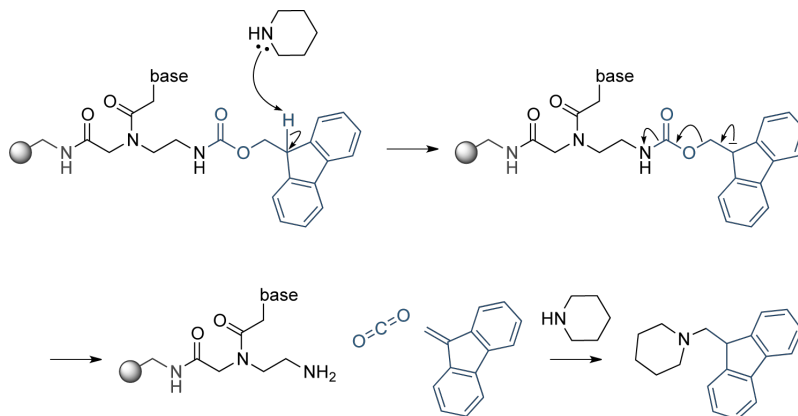


Figure 2.3 Fmoc deprotection mechanism. Under basic conditions, the fluorenyl proton is deprotected to which produces the deprotected amino group, dibenzofulvene, and carbon dioxide.

The Fmoc deprotection mechanism is shown in Figure 2.3. Because the  $\beta$ -hydrogen of the Fmoc group is so acidic, bases such as piperidine can drive the deprotection of Fmoc. Hydrazine, which has a slightly higher  $pK_b$  than piperidine, can also act as a base and deprotect the Fmoc as well. Orbitrap MS/MS concluded that the Fmoc group was deprotected by the 5% hydrazine in DMF solution that also deprotected the ivDde-protecting group on Dab. This means that when the boronic acid coupling occurred, 3 boronic acids were attached to the SL sequence; two located on the side chains of the Dab AAs and one located on the terminal amine of the peptide backbone.

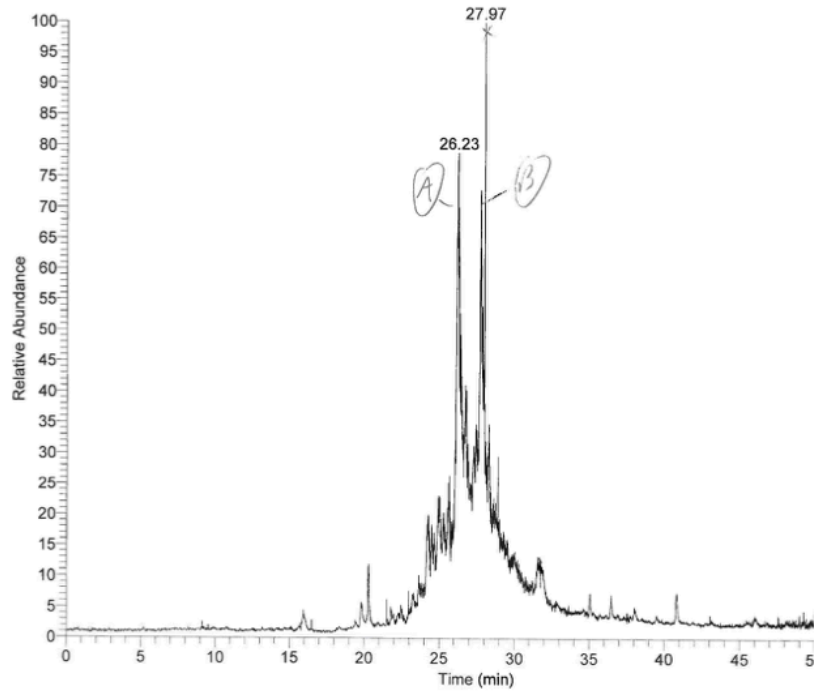


Figure 2.4 Orbitrap LC chromatogram of SL11.

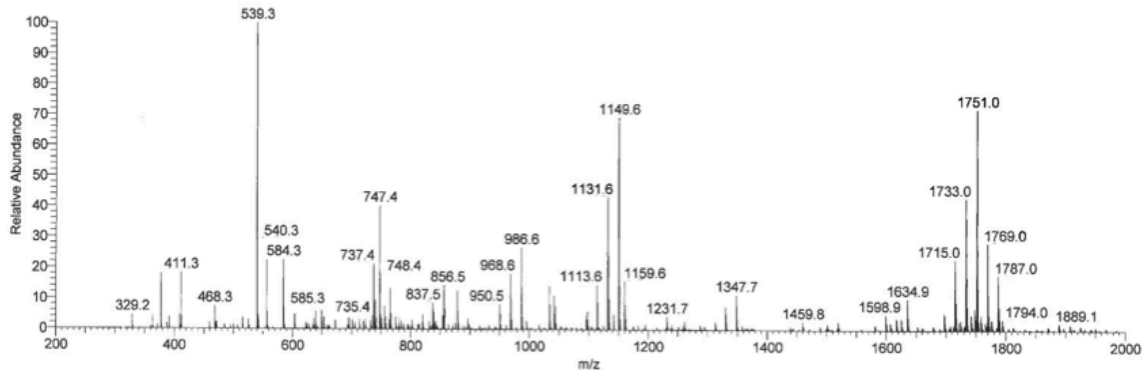


Figure 2.5 Orbitrap MS spectrum of SL11. The 1751 peak correlates with the total weight of SL11 with 3 PBA groups.

## 2.4 SCREENING AGAINST MEMBRANE EXTRACTED GLYCOPROTEINS

Previous lab members identified SLs that interacted with glycoproteins by screening the library against PSM, BSM, and OVA. These glycoproteins were chosen because they were easily accessible and they contain unique glycan structures associated with colon cancer such as sLe<sup>x</sup><sup>3, 36</sup>. The hits from the screening were isolated and

sequenced through MS/MS. When arranged in a cross-reactive sensor array, the 4 SLs discriminated 7 colon cancer cell lines with 92% classification accuracy determined by 3 different cross-validation techniques: Jackknife, Boot-strap, and Training/test set. Furthermore, when discriminating the 7 cell lines by metastatic potential, the array had 97% classification accuracy<sup>2</sup>.

When using the 4 SL cross-reactive sensor array to discriminate prostate cancer cell lines, the classification accuracy decreased. In order to discover SLs that discriminated prostate cancer cell lines with high classification accuracy and had potential to be used in a prostate cancer diagnostic, a solution containing membrane bound proteins and glycoproteins were used as a screening analyte. **Figure 2.6** is a general schematic for the screening procedure with membrane proteins and glycoproteins.

Two different prostate cell lines RWPE-1 and PC3, not cancerous and cancerous-metastatic, respectively, were chosen as analytes. The membrane proteins and glycoproteins were extracted using the membrane extraction kit that was used for the colon cancer cell line discrimination. By using the proteins and glycoproteins from prostate cell membranes, the hits isolated from the screening were more specific to prostate related glycosylation and the aberrant changes that are present.



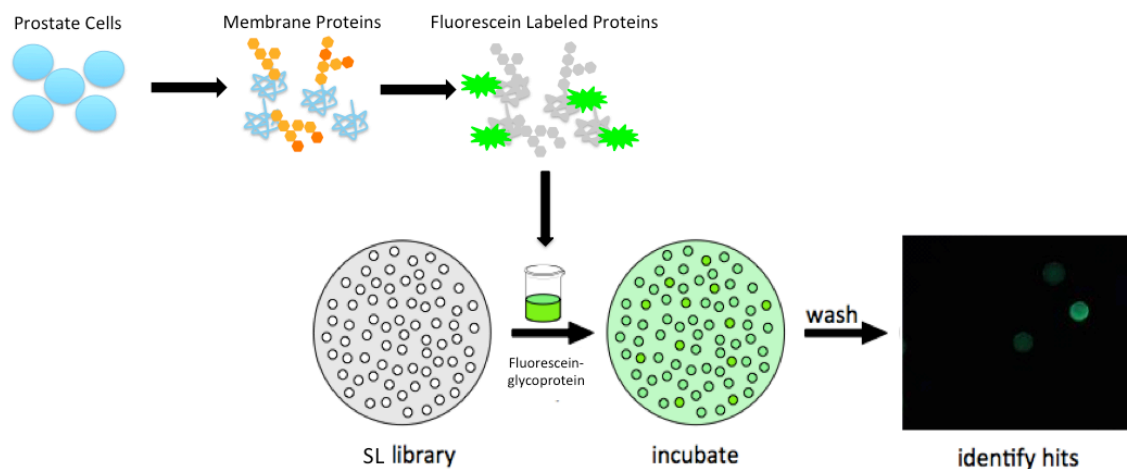


Figure 2.6 Workflow for screening a SL library to identify hits. Prostate cells are cultured and the membrane proteins and glycoproteins were extracted with Qiagen Plasma Membrane Protein Kit. After the extraction, the sample was labeled with fluorescein and then incubated with the SL library. The unbound fluorescent sample was washed away and library resin that was fluorescent were identified and isolated for sequencing.

Three hits were found when screening against fluorescent prostate cell membrane extracts. The first hit, SL10, was found when screening against F-PC3 (Figure 2.7). The second hit, SL13, was found when screening against F-RWPE-1 (Figure 2.8). The third hit, SL14, was found when screening against R-PC3 (Figure 2.9). For each hit, the SL that was more fluorescent than the surrounding SLs was manually removed. The fluorescent proteins and glycoproteins were removed from the SL by three alternating acid/base washes (70% TFA in H<sub>2</sub>O and 0.1 M NaOH) and then the individual SLs was sent to Tufts University Core Facility where the resin was sequencing by Edman Degredation.

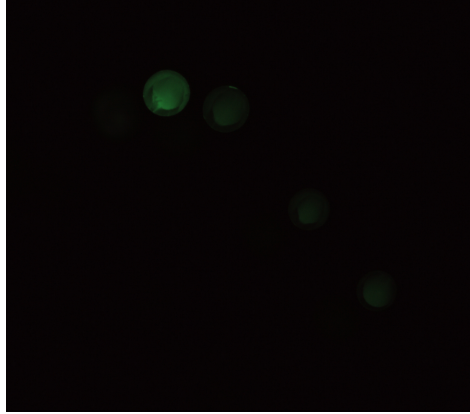


Figure 2.7 SL10 hit located in the top left of the frame. This SL was screened against F-PC3.

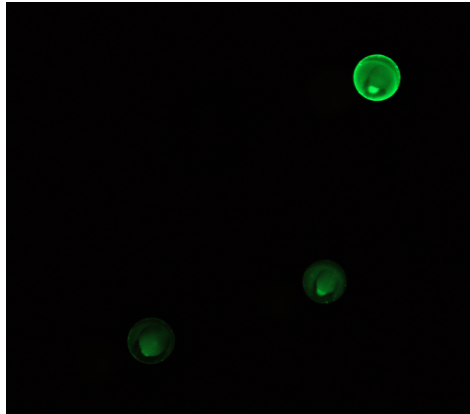


Figure 2.8 SL13 hit located in the top right of the frame. This SL was screened against F-RWPE-1.

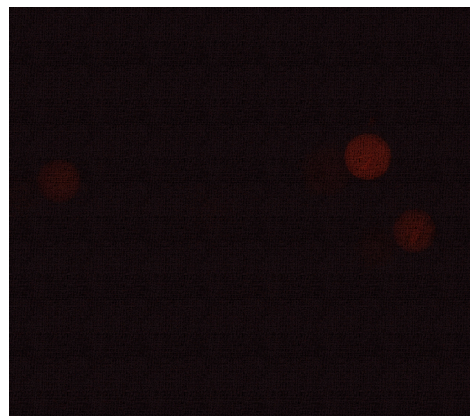


Figure 2.9 SL14 hit located in the top right of the frame. This SL was screened against R-PC3.

The sequences for SL10, SL13, and SL14 can be found in **Table 2.1** (for all Edman Degradation sequencing data, see Appendix A). Sequencing for B-B-R-M was not performed because the first AA was manually coupled and the next three AAs were coupled to the resin using a peptide synthesizer. When looking at the sequencing data, many of the AA positions are clear but some are difficult to distinguish.

Table 2.1 SL10-14 sequences, selectivity and the analyte it was screened against.

SL	Sequence	Screened Against	Selectivity
SINGLE ANALYTE SCREEN			
SL10	R-L-D*-A-R-S-D*-G-B-B-R-M	F PC3	
SL13	R-T-D*-G-L-A-D*-V-B-B-R-M	F RWPE-1	
SL14	R-Y-D*-R-A-S-D*-V-B-B-R-M	R PC3	
DUAL ANALYTE SCREEN			
SL11	R-L-D*-Y-L-T-D*-R-B-B-R-M	F RWPE-1/R PC3	PC3
SL12	R-L-D*-G-F-Y-D*-Q-B-B-R-M	F RWPE-1/R PC3	RWPE-1

In all of hits, the last AA, arginine, is not conclusive from the chromatogram but the last AA is a fixed position in the library. The arginine was manually coupled to the combined library at one time and the coupling was verified through a 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. Appendix B contains all chromatograms from the protein sequencing performed by Tufts University.

## 2.5 DUAL DYE SCREENING

As stated above, the purpose of the library screenings against prostate membrane extracts was to find SLs that interacted with relevant prostate proteins and glycoproteins. In order to identify SLs that could bind differently to non-cancerous samples versus cancerous samples, the prostate analytes that were used in the single screenings were combined to create a dual dye library screening.

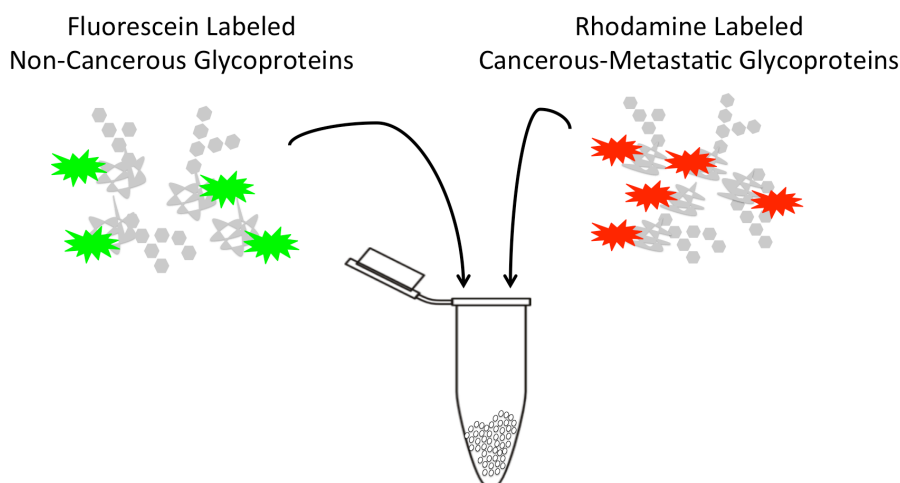


Figure 2.10 Dual Dye Screening Process. In order to select SLs that were selective, a dual dye screening process was used. In this process, the membrane glycoproteins from two different cell lines, one non-cancerous and the other cancerous-metastatic, were labeled with fluorescent dyes of different colors. In this figure, the non-cancerous glycoproteins were labeled with fluorescein and the cancerous-metastatic glycoproteins were labeled with rhodamine.

The general schematic for the dual dye screening is in Figure 2.10. The fluorescein-labeled non-cancerous RWPE-1 membrane proteins and glycoproteins were combined with rhodamine-labeled PC3 membrane proteins and glycoproteins and incubated with the library. After the resin was washed, the SLs that were fluorescent were identified. The reverse dual dye screening was also performed where RWPE-1 was labeled with rhodamine and PC3 was labeled with fluorescein. A bead that was more

fluorescent in the green channel than any of the other beads in that channel but was not more fluorescent than the other beads in the red channel (or vice versa) was selected.

Two hits were isolated from the dual dye screening. The sequences for both SLs can be found in **Table 2.1** and the sequencing data can be found in Appendix A. The first hit from dual dye screening was SL11. The SL was isolated from the library screening against F-RWPE-1 and R-PC3 combined. It was more fluorescent than the SLs surrounding it in the red channel but it was not more fluorescent than the same beads in the green channel (Figure 2.11). This means that SL11 bound more to PC3 membrane extract than the other SLs but it did not bind more to RWPE-1 than the other SLs.

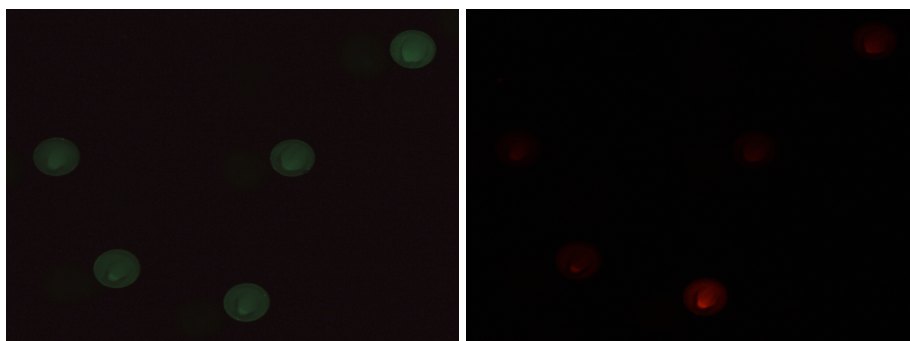


Figure 2.11 SL11 hit. This hit was found by dual dye screening with F-RWPE-1 and R-PC3. The hit is located in the bottom middle part of the right frame.

The second hit from the dual dye screening was SL12. It was also isolated from the library screening against F-RWPE-1 and R-PC3 but it was more fluorescent in the green channel versus the red channel meaning that it interacted more with RWPE-1 than PC3 (Figure 2.12).



Figure 2.12 SL12 hit. This hit was found by dual dye screening with F-RWPE-1 and R-PC3. The hit is located in the middle part of the left frame.

## 2.6 RESYNTHESIZING SL10-14

After the sequences of SL10-14 were identified, the SLs were resynthesized to confirm the structure of the overall SL. As stated previously, the new library synthesis protocol did not remove the Fmoc protecting group with 20% piperidine in DMF before deprotecting the ivDde protecting group on Dab. However, due to hydrazine's ability to act as a base like piperidine, hydrazine could remove the Fmoc protecting group and therefore allow for a third PBA group to be present.

To verify the structure, the SLs were first resynthesized using the same protocol as the new library. Once the peptide sequences had been coupled to resin, the ivDde groups were deprotected and PBAs were attached. The Fmoc group was then deprotected and the acid labile groups were removed. The SL was then cleaved from the bead using cyanogen bromide overnight and analyzed by Orbitrap (Figure 2.4 and Figure 2.5). The masses found for each SL showed that the SL had three PBA groups instead of two like the previous SL1, 3, 4, and 5.

## 2.7 CONCLUSION

In this chapter, library synthesis and screening against prostate specific analytes is discussed. Former lab members optimized the SL library general sequence (R-X-D\*-X-

X-X-D\*-X-B-B-R-M) and the general library screening procedure. While the SLs that were discovered through previous screenings were useful in discriminating colon cancer cell lines, they did not discriminate prostate cell lines with the same classification accuracy. Screening the library against prostate cell membrane extracts, which contained the membrane proteins and glycoproteins from prostate cells, identified new SLs (SL10-14). Once the SLs binding to the analytes had been isolated, the bound glycoproteins were removed and sequenced. Due to instrumentation complications, the MALDI, technique used by former lab members to sequence the SLs, could not be used and Edman Degradation was used instead. SL10, 13, and 14 were isolated from library screenings against single analytes such as RWPE-1 (non-cancerous) and PC3 (cancerous-metastatic). SL11 and SL12 were isolated from library screenings against the two single analytes combined together. In order to see the selectivity for one analyte versus another, the analytes were labeled with different fluorophores. Once the new SLs were resynthesized, Orbitrap was used to validate the overall structure of the SLs. Instead of just two PBA groups located on the fixed Dab positions, a third was also present on the N-terminus end of the peptide.

## 2.8 EXPERIMENTAL METHODS

### *Chemicals*

All AA and HBTU were purchased from EMD Millipore. Prostate Specific Antigen was purchased from Sigma-Aldrich. TentaGel resin (Cat. No. MB-300-002; loading level 0.25-0.3 mmol/g) was purchased from Rapp Polymere. All other chemicals were purchased from Acros Organics and used without further purification.

### *Microscope Usage*

A Leica MZ 16F microscope was used to evaluate the binding of SLs and analytes of interest. A GFP3 filter set with an excitation 450-490 nm and emission 500-550 nm. Images were taken using a QImaging MicroPublisher 5.0 RTV digital camera associated with the microscope. Images were analyzed by MATLAB, which provided SL fluorescence information.

### *Library Synthesis*

The general sequence of the library was R-X-D\*-X-X-X-D\*-X-B-B-R-M-Bead, which means that the first four amino acids were the same for all the beads in the library. The first AA, methionine, was attached manually by activating Fmoc-met-OH in a 5% N-methylmorpholine in DMF solution and then mixing the AA solution with Tentagel resin beads (300  $\mu$ m). After incubating the AA solution with the beads for roughly an hour, the AA solution was removed and a ninhydrin test was performed to confirm that the AA had coupled to the bead. If the ninhydrin test was positive, the beads were put onto the Protein Technologies Peptide Synthesizer (PS3) where the next 3 AAs (R, B, B) were coupled to the library of beads. After the beads were removed from the PS3, the random AA was coupled to the beads using a split-and-pool method. Briefly, the beads are split into 11 different reaction vessels. In each vessel, a different activated AA is coupled to the beads. The beads are then combined back into one vessel and deprotected. Because the next position is fixed with Dab, the beads are coupled to the AA and deprotected in the same vessel. The next 6 AAs are coupled depending on the fixed or random position of that particular AA.

Once all of the AAs are coupled, the ivDde protecting group of D\* is removed by using a 5% hydrazine in DMF solution. This deprotection solution also deprotects the last



FMOC protecting group on the final R of the peptide sequence. The PBA functional groups are added through reductive amination by incubating the beads with a 2% MeOH in DMF solution containing an 8 equivalent ratio of bead to 2-formylphenyl boronic acid. The remaining side-chain protecting groups are removed using a 95% TFA, 2.5% TIS, and 2.5% H<sub>2</sub>O.

#### *MS/MS Analysis*

To cleave the peptide from the resin, the beads were treated with 50 µl of 40 mg/ml CNBr in 0.1 M HCl for 16 hr in the dark. The beads were removed from the solution and the excess CNBR, TFA, and water were removed under reduced pressure by using an AES1010 SpeedVac system. The cleaved peptides were re-dissolved in 10 µl of 0.1% TFA in water. 1 µl of matrix containing saturated α-cyano-4-hydroxycinnamic acid (CHCA) in 50:50:0.1 acetonitrile/water/TFA was positioned on to a MALDI plate and mixed with 2 µl of the cleaved SL solution. Samples were then analyzed by MALDI-TOF/MS and MS/MS techniques.

#### *Cell Culture*

VWR was the supplier for all supplements and media for all cell lines. The media for RWPE-1 was K-SFM supplemented with recombinant human Epidermal Growth Factor (rhEGF) and Bovine Pituitary Extract (BPE). The media for PC3 was RPMI 1640 with 10% FBS. The cells were grown in T75 flasks at 37 degrees Celsius.

#### *PBS*

To make a 4L solution of PBS, sodium phosphate monobasic (monohydrate) (27.6 g, 200 mmol), sodium phosphate dibasic (anhydrous) (28.4 g, 200 mmol) and sodium chloride (35.06 g, 600 mmol) was dissolved in 3.5 L of DI water. The pH was adjusted to

7.2-7.3 using 3 M NaOH and then the solution was diluted to 4L by adding DI water.

#### *Fluorescent Labeling of Prostate Cell Membrane Extracts*

The Qiagen Plasma Membrane Protein Kit was used to extract plasma membrane proteins and glycoproteins. The extract was spun into carbonate buffer pH 9.8 and a BCA assay was performed to obtain the concentration of the extract. The amount of protein needed for library screening was added to carbonate buffer to a total volume of 600  $\mu$ L. Depending on which dye was being used to label, 0.07  $\mu$ mol NHS-fluorescein or NHS-rodamine was added and the solution was incubated at 37°C and tumbled for 1 hr. Amicon Pro Centrifugal Filter (10 kDa) were used to remove excess dye and to concentrate the proteins and glycoproteins after the 1 hr incubation.

#### *BCA Assay*

The Thermo Scientific BCA kit was purchased from VWR. A serial dilution of the BCA 2 mg/mL standard was used for the analysis. Standards and samples (10  $\mu$ L) were placed in a 96 well plate in triplicate and combined with 200  $\mu$ L of the reagent from the BCA kit. After the wells had been mixed thoroughly, plate was incubated in an oven for approximately 30 minutes at 37°C. The absorbance of the wells was read on a plate reader at 562 nm.

#### *Library Screening for Single Analytes*

Approximately 2 mg of library resin was rinsed in PBS twice and then pre-incubated with 1% BSA in PBS with 10% glycerol for 15 minutes. This pre-incubation was 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 1 mL of a 0.0002 mg/mL solution of the fluorescently-labeled prostate membrane extract in PBS was

incubated with the beads for 6 hrs at room temperature with gentle shaking. After incubation, the beads were rinsed with PBS three times and imaged with a fluorescent microscope, Leica MZ 16F.

#### *Library Screening Dual Dye Analytes*

The library screening for dual dye analytes compared to single analytes were almost completely identical. The only significant difference was the amount of each analyte that was used for the study. Instead of using 1 mL containing 0.0002 mg of both analytes to give a total concentration of 0.0004 mg/mL of protein, only 0.0001 mg was use of each to make a total concentration of 0.0002 mg/mL.

## CHAPTER 3:

### PROSTATE SELECTED SYNTHETIC LECTINS

#### 3.1 CROSS-REACTIVITY WITH PROSTATE CELL LINES

Previously, SLs that were isolated through library screenings against purified glycoproteins PSM and OVA were found to be cross-reactive to purified glycoproteins as well as cancer associated glycans and colon cancer cell lines. It is important to mention that the cross-reactivity of the SLs is what gives the SL array the ability to discriminate the different glycoproteins they interact with. In order to evaluate the selectivity or cross-reactivity of SL10-14 identified in Chapter 2, the SLs were resynthesized using the same general library synthesis methods as discussed in Chapter 2.

As seen in **Figure 3.1**, SL10-14 are cross-reactive to 6 prostate cell lines with varying metastatic potential. The prostate cell lines used can be found in Table 3.1

Table 3.1 Prostate cell lines with metastatic potential listed.

Cell Line	Tissue Type	Metastatic Potential
RWPE-1	Prostate	Non-Cancerous
WPE1-NA22	Prostate	Cancerous/Non-Metastatic
WPE1-NB14	Prostate	Cancerous/Non-Metastatic
LNCaP	Prostate	Cancerous/Metastatic
DU145	Prostate	Cancerous/Metastatic
PC3	Prostate	Cancerous/Metastatic

While some SLs interact more with one cell line over another (e.g. SL10 binding with WPE1-NB14 vs. its binding with DU145), not one is selective for one prostate cell line. Each SL binds cross-reactively to the prostate cell lines.

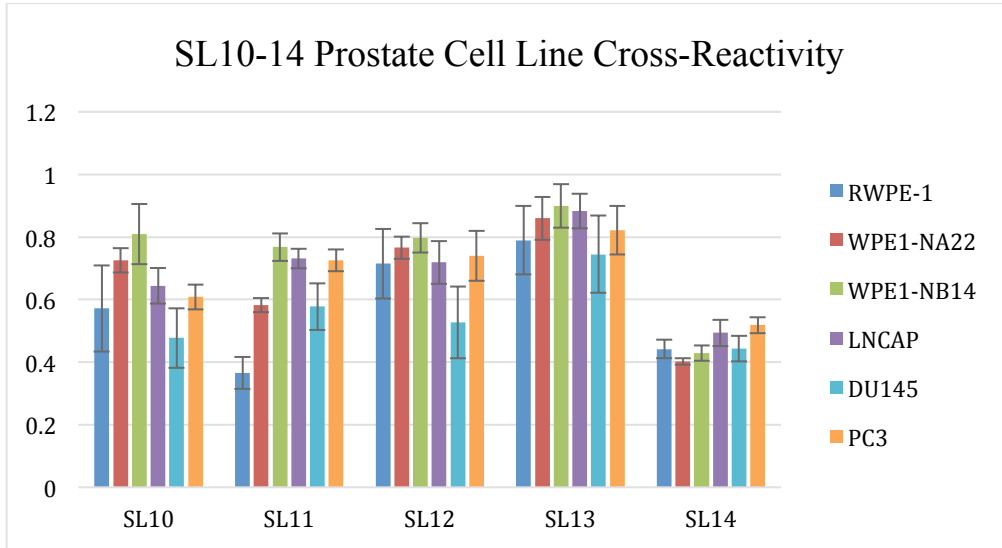


Figure 3.1 Binding patterns/ “Fingerprint patterns” of SL10-14 with prostate cell lines.

### 3.2 SL10-14 PROSTATE CELL LINE DISCRIMINATION

As mentioned before in Chapter 1, because it is difficult to see how the SLs are binding differently to cell lines, LDA is the statistical analysis used to better understand the similarities and differences in the data above. When using LDA, SL10-14 discriminated prostate cell lines with 83% classification accuracy (Figure 3.2).

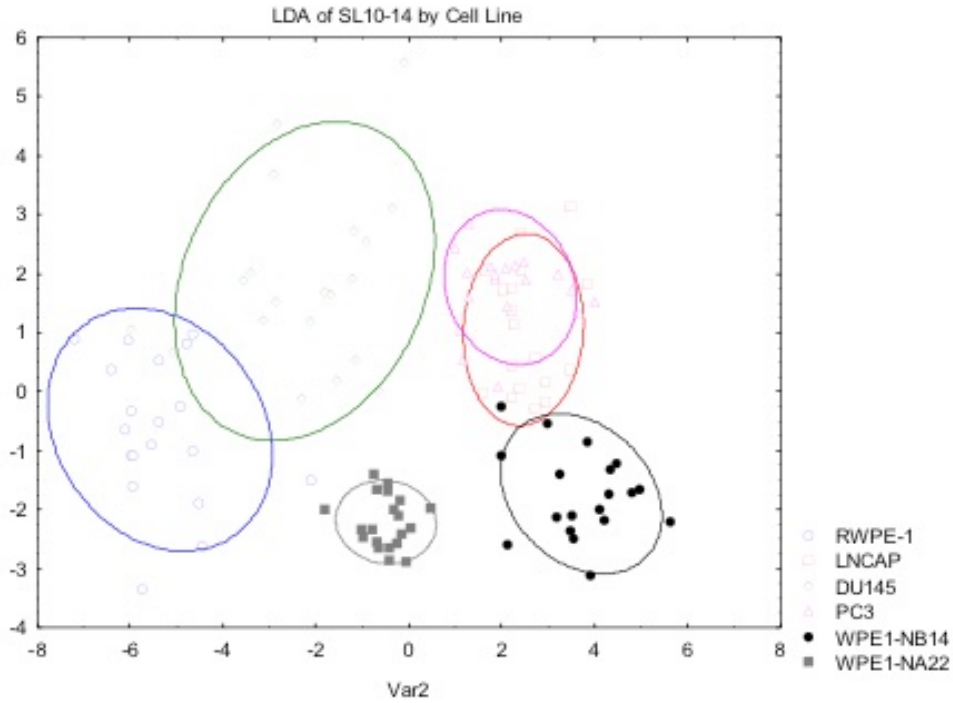


Figure 3.2 LDA plot of SL10-14 classified by prostate cell line.

While it is beneficial to have an array that can correctly identify the different cell lines present, it is more important to evaluate the array’s ability to distinguish the cell lines’ metastatic potential when developing a cancer screening array.

When discriminating the prostate cell lines by metastatic potential (**Table 3.2**) the classification accuracy was 92% (Figure 3.3). By incorporating the SLs in the former SL array, SL1, 3, 4, and 5, into an array with the new SLs, the classification accuracy increases to 100% (Figure 3.4).

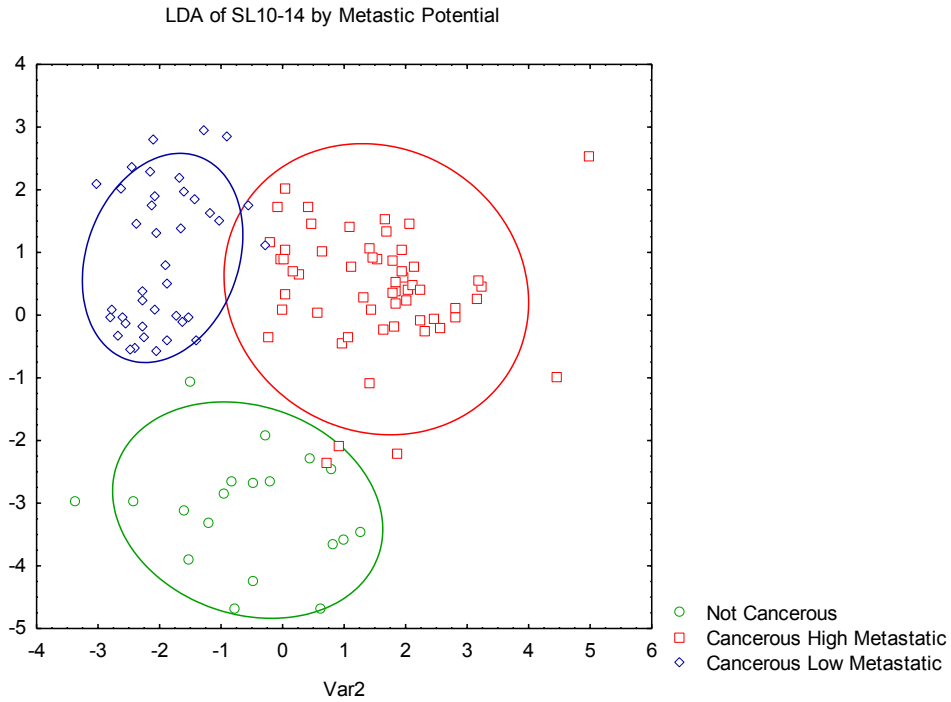


Figure 3.3 LDA of SL10-14 classified by metastatic potential.

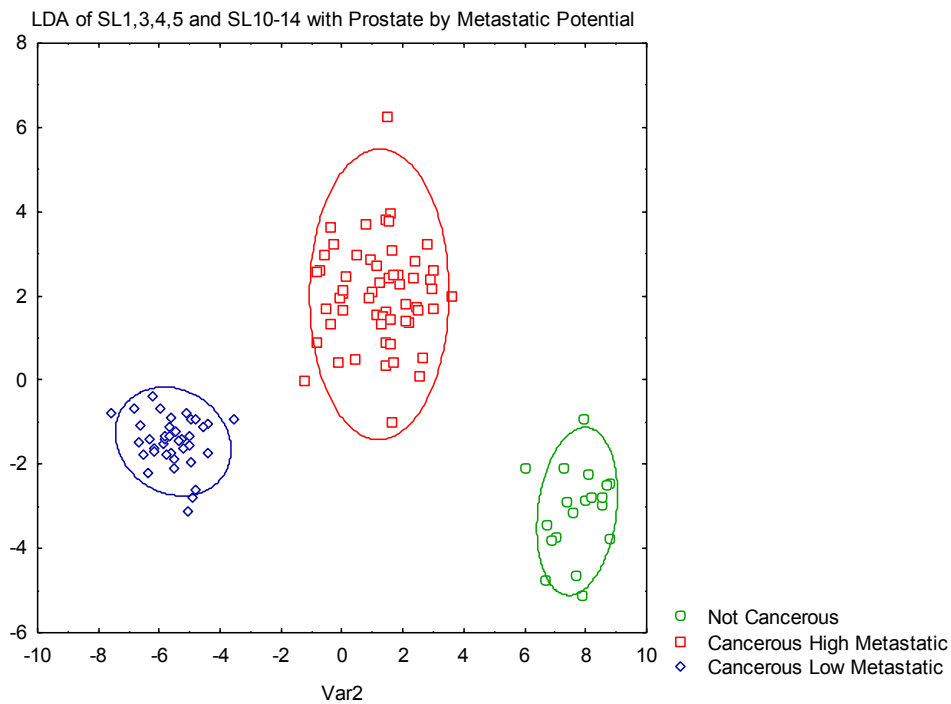


Figure 3.4 LDA of SL1, 3, 4, 5 and SL10-14 Prostate by metastatic potential.

### 3.3 PROSTATE CANCER VS. COLON CANCER

An ideal cancer diagnostic would not only be able to diagnose the severity of the disease but also what kind of cancer is present. Using the prostate and colon cell lines in **Table 3.2**, LDA was used to discriminate the two types of tissue. When using the SL array found by former lab members (SL1, 3, 4, and 5) the classification accuracy was 66%. When using the prostate selected SLs, the classification accuracy increased to 95%. The combined arrays were able to discriminate between prostate cell lines and colon cell lines with 97% classification accuracy. This information suggests that the prostate selected SLs bind to the glycoproteins from the two types of tissues in different ways.

Table 3.2 Prostate and colon cell lines with metastatic potential listed.

	<b>Type of Tissue</b>	<b>Metastatic Potential</b>
<b>RWPE-1</b>	Prostate	Non-Cancerous
<b>CCD 841 CoN</b>	Colon	Non-Cancerous
<b>WPE1-NA22</b>	Prostate	Lowly Metastatic
<b>WPE1-NB14</b>	Prostate	Lowly Metastatic
<b>HCT 116</b>	Colon	Lowly Metastatic
<b>HT-29</b>	Colon	Lowly Metastatic
<b>LNCAP</b>	Prostate	Highly Metastatic
<b>DU145</b>	Prostate	Highly Metastatic
<b>PC3</b>	Prostate	Highly Metastatic
<b>LoVo</b>	Colon	Highly Metastatic

### 3.4 THE UTILITY OF SL11

LDA provides more information than just classification accuracies. It can also information about which SL is most important for discrimination. It provides this information through an F-to-remove value, which is a value that describes the variables importance in the discrimination being performed<sup>37</sup>. If the variable with a large F-to-remove was removed from the analysis, the discrimination accuracy would decrease. In all of the analyses, SL11 always had a high F-to-remove value.



To see just how powerful SL11 is in discriminating the prostate cell lines, LDA was used to get classification accuracies. Each analysis was SL11 discriminating one cell line vs. a different cell line. The results are listed in **Table 3.3**.

SL11 can discriminate all of the cell lines with high classification accuracies except for distinguishing WPE1-NA22 from DU145 and LNCaP from PC3. What is the most important to note is that SL11 has high classification accuracies for distinguishing the non-cancerous cell line from all of the other cell lines regardless of their varying metastatic potential.

Table 3.3 Classification matrix for SL11 with prostate cell lines generated by LDA.

	<b>RWPE-1</b>	<b>WPE1-NA22</b>	<b>WPE1-NB14</b>	<b>LNCaP</b>	<b>DU145</b>	<b>PC3</b>
<b>RWPE-1</b>	NA	97	100	100	92	100
<b>WPE1-NA22</b>	97	NA	100	100	60	100
<b>WPE1-NB14</b>	100	100	NA	74	97	70
<b>LNCAP</b>	100	100	74	NA	97	57
<b>DU145</b>	92	60	97	97	NA	97
<b>PC3</b>	100	100	70	57	97	NA

### 3.5 TWO PBA VS. THREE PBA

In order to sequence hits, the synthesis of the library had to be altered from previous libraries. After resynthesizing the hits found from screening against prostate membrane proteins and glycoproteins, Orbitrap MS data confirmed there were three PBA groups attached to the peptide backbone instead of two PBA groups. Because the PBA groups interact with the sugars present on the glycoproteins, it is important to understand how the new SL structure plays a role in discriminating cell lines.

When using the prostate selected SLs with 2 boronic acid groups (SL10 2BA-SL14 2BA) to discriminate the prostate cell lines (Figure 3.5), the classification accuracy decreased from 92% (using SL10-14 with 3 boronic acid groups) to 86%. By

incorporating another boronic acid group to the SL sequence, the SLs have more possible interactions with the glycans present.

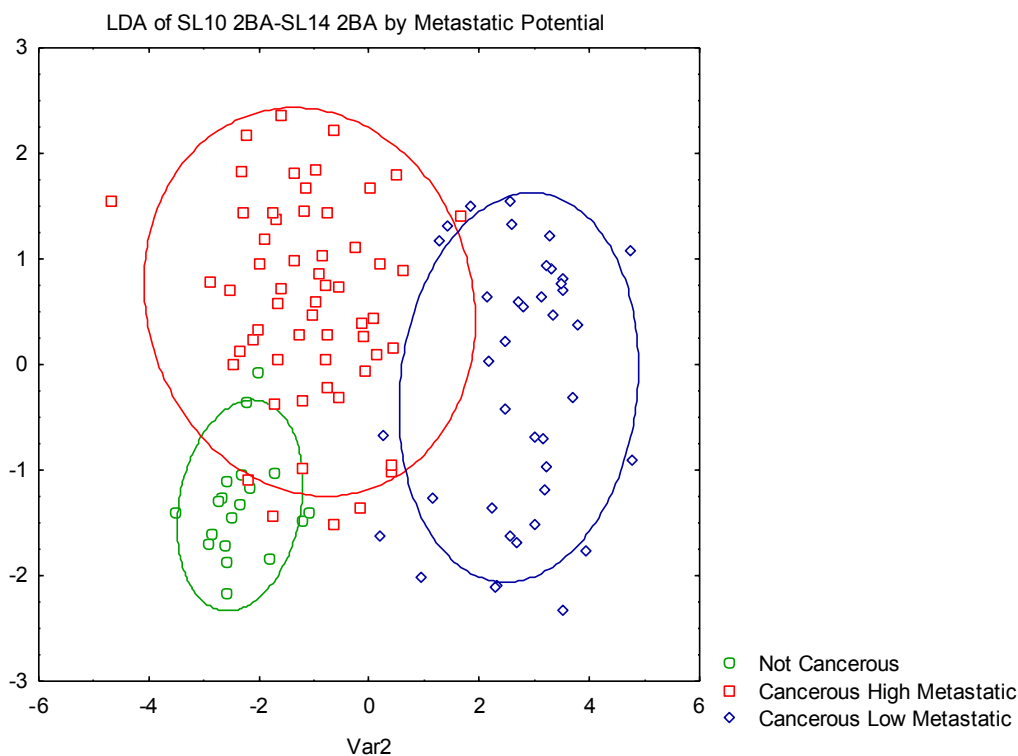


Figure 3.5 LDA plot of SL10 2BA-SL14 2BA classified by metastatic potential.

### 3.6 CONCLUSION

In this chapter, the prostate selected SLs that were isolated and characterized in Chapter 2 have been resynthesized and bound to analytes of interest. The prostate selected SL array (SL10-14) has the ability to discriminate 6 prostate cell lines with 83% classification accuracy. While it is important to be able to distinguish between the cell lines, it is more important to be able to discriminate the prostate cell lines by metastatic potential. The SL array had a 92% classification accuracy when discriminating the prostate cell lines by metastatic potential (non-cancerous, cancerous/non-metastatic, and cancerous/metastatic). By incorporating the SLs isolated by former lab members into the

prostate selected SL array, the classification accuracy of prostate cell lines by metastatic potential increased to 100%.

The prostate selected SL array also has the ability to distinguish between prostate cell lines and colon cell lines. This is a powerful tool when designing a diagnostic. Not only is it important to be able to tell a patient if they have cancer or not, but also to tell them what kind of cancer they have. When using the prostate selected SL array, the classification accuracy for prostate cell lines vs. colon cell lines was 95%. This was a significant increase from the previous SL array, which had a classification accuracy of 66%.

SL11 was a particularly important variable in all of the analyses discussed. It always had a high f-to-remove value, which suggests that if it were taken out of the analysis, the discrimination accuracy would decrease. SL11, when used by itself to discriminate between any two prostate cell lines, had high classification accuracies barring two analyses. In particular, SL11 had high classification accuracies when distinguishing the non-cancerous cell line RWPE-1 from any other prostate cell line. In further studies, it would be interesting to discover what glycans SL11 interacts with based on these results.

Lastly, due to sequencing issues discussed in Chapter 2, the SL hits isolated from screenings against prostate membrane proteins and glycoproteins contained 3 boronic acid functional groups instead of two boronic acid groups. The former SL array contained only two boronic acid groups. In order to see if the amount of boronic acid functional groups plays a role in discrimination, the two different arrays (SL10-14 with 3 boronic acids groups and SL10-14 with 2 boronic acids groups) were compared. When using the

prostate selected SL array, the classification accuracy of the prostate cell lines was 92%. When using the modified array, which contained only two boronic acids, the classification accuracy drops to 86%. These results suggest that amount of boronic acid functional groups are important for the discrimination of prostate cell lines.

### 3.7 EXPERIMENTAL METHODS

#### *Chemicals*

All AA and HBTU were purchased from EMD Millipore. Prostate Specific Antigen was purchased from Sigma-Aldrich. TentaGel resin (Cat. No. MB-300-002; loading level 0.25-0.3 mmol/g) was purchased from Rapp Polymere. All other chemicals were purchased from Acros Organics and used without further purification.

#### *Microscope Usage*

A Leica MZ 16F microscope was used to evaluate the binding of SLs and analytes of interest. A GFP3 filter set with an excitation 450-490 nm and emission 500-550 nm. Images were taken using a QImaging MicroPublisher 5.0 RTV digital camera associated with the microscope. Images were analyzed by MATLAB, which provided SL fluorescence information.

#### *Cell Culture*

VWR was the supplier for all supplements and media for all cell lines. All cells were grown in T75 flasks at 37 °C in 5% CO<sub>2</sub> and 95% air. The media for RWPE-1 was K-SFM supplemented with recombinant human Epidermal Growth Factor (rhEGF) and Bovine Pituitary Extract (BPE). The media for PC3 was RPMI 1640 with 10% FBS. LoVo human colorectal carcinoma cells were grown in RPMI 1640 with 10% FBS.

#### *General SL Synthesis Method*

TentaGel resin (200 mg, 0.05 mmol) was swollen in 10 mL of DMF for 10 min. The DMF was removed and Fmoc-Met-OH (4 eq, 74.3 mg, 0.2 mmol) preactivated with HBTU (4 eq, 75.9 mg, 0.2 mmol) for 10 min in 10 mL of 5% N-methylmorpholine in DMF, was then added. The resin was then tumbled for 1 hr. The amino acid solution was removed by vacuum filtration and the resin was then washed with DMF, MeOH, DMF, and MeOH. Ten mL of a 20% piperidine in DMF solution was then added to remove the Fmoc group. The same procedure was used to subsequently couple the remaining amino acids in the sequence.

#### *Two PBA SL Synthesis*

After the peptide backbone had been coupled to the resin and the last AA (arginine) was Fmoc deprotected, the peptide was the N-terminally acetylated by tumbling the resin in 10 mL of CH<sub>2</sub>Cl<sub>2</sub> containing 5% acetic anhydride and 5% pyridine.

#### *Three BA SL Synthesis*

Once the peptide backbone had been coupled to the resin and before the last AA (arginine), ivDde was deprotected using a 5% hydrazine in DMF solution. The Fmoc protecting group on the last AA was not deprotected with 20% piperidine in DMF before the ivDde deprotection.

#### *Deprotection of ivDde Protecting Group*

The ivDde protecting group of the Dab was removed by tumbling the resin in 10 mL of 5% hydrazine monohydrate in DMF (v/v) for 1 hr after the peptide backbone had been coupled to the resin.

### *Incorporation of PBA*

2-formylphenyl boronic acid (8 eq, 119.95 mg, 0.8 mmol) was dissolved in 10% MeOH in DMF. The boronic acid solution and activated 3Å molecular sieves was then added to the resin and was tumbled overnight at 37°C. NaBH<sub>4</sub> (8 eq, 30.26 mg, 0.8 mmol) was then added. Gas was evolved for 30 min and the resin was tumbled for 4 hr. Resin was then washed with DMF, MeOH, DMF, and MeOH. Molecular sieves removed and washed again.

### *Removal of Acid-labile Protecting Groups*

The resin was tumbled in 10 ml of 95% TFA, 2.5% water, and 2.5% TIS for 1 hr. The solution was removed by vacuum filtration and the resin was washed with DMF, MeOH, DMF, and MeOH.

### *MS/MS Analysis*

To cleave the peptide from the resin, the beads were treated with 50 µl of 40 mg/ml CNBr in 0.1 M HCl for 16 hr in the dark. The beads were removed from the solution and the excess CNBR, TFA, and water were removed under reduced pressure by using an AES1010 SpeedVac system. The cleaved peptides were re-dissolved in 10 µl of 0.1% TFA in water. 1 µl of matrix containing saturated α-cyano-4-hydroxycinnamic acid (CHCA) in 50:50:0.1 acetonitrile/water/TFA was positioned on to a MALDI plate and mixed with 2 µl of the cleaved SL solution. Samples were then analyzed by MALDI-TOF/MS and MS/MS techniques.

### *PBS*

To make a 4L solution of PBS, sodium phosphate monobasic (monohydrate) (27.6 g, 200 mmol), sodium phosphate dibasic (anhydrous) (28.4 g, 200 mmol) and sodium chloride

(35.06 g, 600 mmol) was dissolved in 3.5 L of DI water. The pH was adjusted to 7.2-7.3 using 3 M NaOH and then the solution was diluted to 4L by adding DI water.

#### *Fluorescent Labeling of Prostate Cell Membrane Extracts*

The Qiagen Plasma Membrane Protein Kit was used to extract plasma membrane proteins and glycoproteins. The extract was spun into carbonate buffer pH 9.8 and a BCA assay was performed to obtain the concentration of the extract. The amount of protein needed for library screening was added to carbonate buffer to a total volume of 600  $\mu$ L. Depending on which dye was being used to label, 0.07  $\mu$ mol NHS-fluorescein or NHS-rodamine was added and the solution was incubated at 37°C and tumbled for 1 hr. Amicon Pro Centrifugal Filter (10 kDa) were used to remove excess dye and to concentrate the proteins and glycoproteins after the 1 hr incubation.

#### *BCA Assay*

The Thermo Scientific BCA kit was purchased from VWR. A serial dilution of the BCA 2 mg/mL standard was used for the analysis. Standards and samples (10  $\mu$ L) were placed in a 96 well plate in triplicate and combined with 200  $\mu$ L of the reagent from the BCA kit. After the wells had been mixed thoroughly, plate was incubated in an oven for approximately 30 minutes at 37°C. The absorbance of the wells was read on a plate reader at 562nm.

#### *Binding Studies with Fluorescent-labeled Membrane Extracts*

Approximately 2 mg of SL was rinsed in PBS twice and then pre-incubated with 1% BSA in PBS with 10% glycerol for 15 minutes. This pre-incubation was 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 1 mL of a 0.0002 mg/mL solution

of the fluorescently-labeled prostate membrane extract in PBS was incubated with the beads for 6 hrs at room temperature with gentle shaking. After incubation, the beads were rinsed with PBS three times, imaged with a fluorescent microscope, Leica MZ 16F, and analyzed by Matlab R2012a. The output was analyzed by linear discriminant analysis using SYSTAT 11.00.01.



## CHAPTER 4:

### OPTIMIZATION OF DATA ACQUISITION

#### 4.1 SL ARRAY DATA ACQUISITION

As stated in Chapter 1, cancer diagnostics are not ideal and a novel, facile, and efficient procedure needs to be created to increase patient survival rates. The SL arrays that have been discussed throughout this dissertation have the potential to diagnose cancer with high accuracies and during the early stages of cancer, when it is most treatable. One important component of the SL array diagnostic is data acquisition.

#### 4.2 OBTAINING DATA FOR MATLAB

Through standard Fmoc protocols and HBTU activation, the four SLs (**Table 4.1**) used in the colon cancer array were re-synthesized and verified by matrix assisted laser desorption/ionization (MALDI).

Table 4.1 SL sequences used in colon cancer array.

SL Hit	Sequence
SL1	Ac-R-G-D*-V-T-F-D*-R-B-B-R-M-resin
SL3	Ac-R-S-D*-V-T-T-D*-R-B-B-R-M-resin
SL4	Ac-R-R-D*-T-Q-T-D*-Q-B-B-R-M-resin
SL5	Ac-R-A-D*-T-R-V-D*-V-B-B-R-M-resin

Three colon cell lines analyzed by previous lab members along with two new breast cell lines (Table 4.2) were grown in T75 flasks at 37°C in 5% CO<sub>2</sub> and 95% air.

Table 4.2 Colon and breast cell lines with metastatic potential listed.

Cell Line	Tissue Type	Metastatic Potential
3T3/NIH	Colon	Non-Cancerous
CT26	Colon	Cancerous/Non-Metastatic
LoVo	Colon	Cancerous/Metastatic
MCF10A	Breast	Non-Cancerous
MCF7	Breast	Cancerous/Non-Metastatic

After the cells had grown to 70-75% confluence, the cells were scraped and treated with the Qiagen Plasma Membrane Protein Kit to extract plasma membrane proteins and glycoproteins. The extracts were labeled with fluorescence isothiocyanate (FITC) and then dialyzed against screening buffer.

SL1, 3, 4, and 5 were first washed with phosphate buffered saline (PBS). The beads were then incubated with 1% bovine serum albumin (BSA) to decrease non-specific binding. After 15 minutes, the beads were then incubated with FITC labeled membrane extracts for 6 hours. The SLs were washed three times and imaged with a Leica MZ 16F microscope with a GFP filter set. Matlab™ R2012a analyzed the images and the data produced was analyzed by SYSTAT™.

#### 4.3 MATLAB

While Matlab™ eases the process of analyzing the images taken, there are many sources of error that are present. The beads are suspended on the slides in PBS so some beads could be floating while other beads might be resting on the microscope slide making some beads look out of focus and not as bright. The camera that is used may also have burned out pixels that the Matlab™ program could identify as a bead that is bright. Also, the lamp that is used is circular while the image that is taken is rectangular meaning

that the beads that fall in the outer area of the frame can appear to be less bright due to a decrease in binding with the glycoproteins but that is not the case. With all of these errors, there can be parameters adjusted in the Matlab™ program that enable the source of error to be eliminated.

#### 4.4 REGION OF INTEREST

The first parameter that was considered was the region of interest (ROI) feature. The ROI places a set area that Matlab™ picks the bright beads from which can allow the program to exclude the beads that fall in the outer areas of the frame that appear to be not as bright. The three colon cell lines (3T3/NIH, CT26, and LoVo) were analyzed twice by Matlab™, once with and once without the ROI. Classification accuracies increase by 19% (70% without an ROI, Figure 4.1, and 89% with an ROI, Figure 4.2) when the ROI is used on the images. These results show that using the ROI and excluding the beads that fall in the edges of the image are important for better classification accuracies.

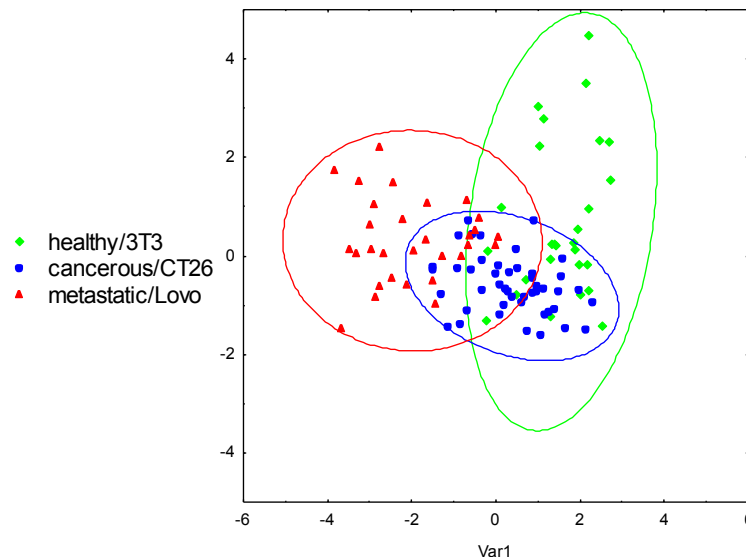


Figure 4.1 LDA plot for discrimination of three colon cancer cell lines by metastatic potential. This analysis was based on data acquisition without an ROI.

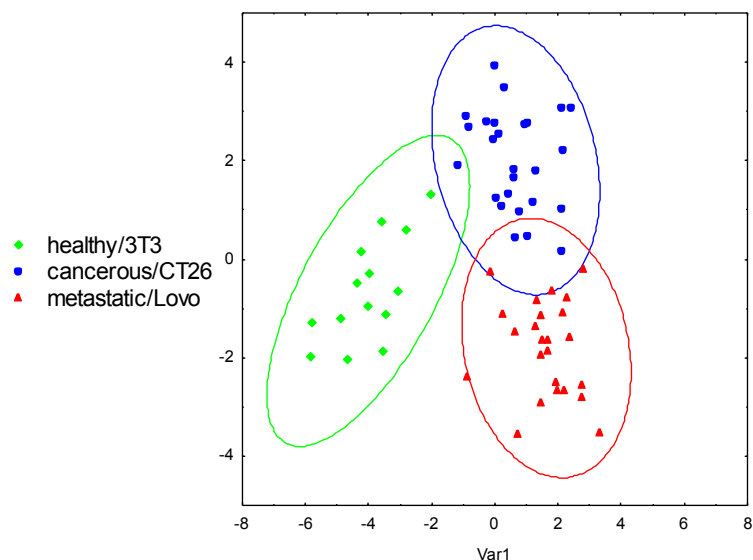


Figure 4.2 LDA plot for discrimination of three colon cancer cell lines by metastatic potential. This analysis was based on data acquisition with an ROI.

#### 4.5 COLOR SPACES

The second parameter investigated was the red, green, and blue color spaces in the color of the bead. Previous lab members did not have the Matlab program so they had to pick bright beads manually and record its luminosity using Adobe Photoshop™. Matlab™ picks the bright beads and records the luminosity as well as the intensities in the red, green, and blue color spaces. By using all of the red, green, and blue intensities, the number of independent variables increases which increases the SL array's ability to discriminate the cell lines with greater classification accuracy (a 14% increase) (Figure 4.3 and Figure 4.4). The variables are independent because of fluorescein's solvatochromism and the background contributions to the spectrum. These results show that using the red, green, and blue color spaces are important for higher classification accuracies.

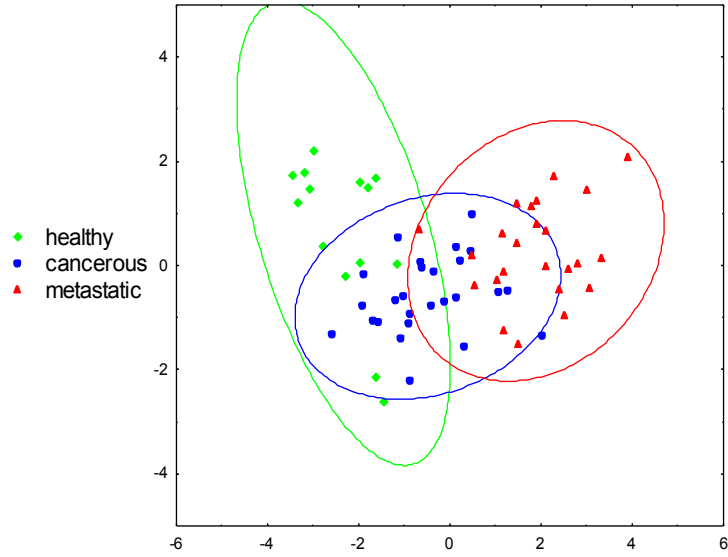


Figure 4.3 LDA plot for discrimination of three colon cancer cell lines by metastatic potential. This analysis was based on data acquisition using only bright numbers.

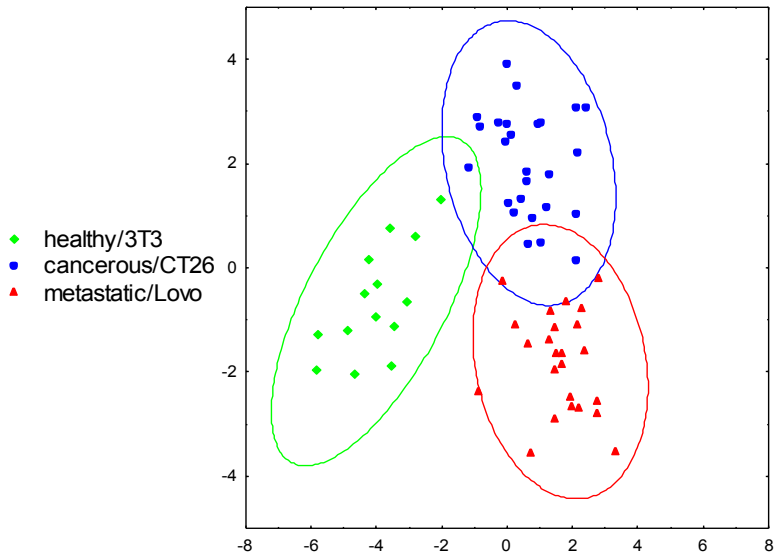


Figure 4.4 LDA plot for discrimination of three colon cancer cell lines by metastatic potential. This analysis was based on data acquisition using all three color spaces.

#### 4.6 CONCLUSION

Because the current SL array seems to discriminate better by metastatic potential, colon and breast cell lines were discriminated by metastatic potential to see if the array

has the ability to be used on other cell lines or if it is specific to colon cancer. The array had a classification accuracy of 88% when distinguishing the metastatic potential of five cell lines originating from different tissues and 97% accuracy when distinguishing the two breast cell lines. These results show that the array can discriminate colon and breast cell lines.

#### 4.7 EXPERIMENTAL METHODS

##### *Synthesis of Synthetic Lectins.*

For example, TentaGel resin (200 mg, 0.05 mmol) was swollen in 10 ml of Dimethylformamide (DMF) for 10 min. The DMF was removed and Fmoc-Met-OH (4 eq, 74.3 mg, 0.2 mmol) preactivated with HBTU (4 eq, 75.9 mg, 0.2 mmol) for 10 min in 10 ml of 5% N-methylmorpholine in DMF, was then added. The resin was then tumbled for 45 min at 37°C. The amino acid solution was then removed by vacuum filtration and the resin was then washed with DMF, methanol (MeOH), DMF, and MeOH. 10 ml of a 20% piperidine in DMF solution was then added to remove the Fmoc group. The same procedure was used to subsequently couple the remaining amino acids in the sequence. The peptide was the N-terminally acetylated by tumbling the resin in 10 ml of CH<sub>2</sub>Cl<sub>2</sub> containing 5% acetic anhydride and 5% pyridine.

##### *Deprotection of ivDde Protecting Group.*

The ivDde protecting group of the diamino acid Dab was removed by tumbling the resin in 10 ml of 2% hydrazine monohydrate in DMF (v/v) for 1 hr.

##### *Incorporation of Phenyl Boronic Acid (PBA).*

2-formylphenyl boronic acid (8 eq, 119.95 mg, 0.8 mmol) was dissolved in 10% MeOH in DMF. The boronic acid solution and activated 3Å molecular sieves was then added to

the resin and was tumbled overnight at 37°C. NaBH<sub>4</sub> (8 eq, 30.26 mg, 0.8 mmol) was then added. Gas was evolved for 30 min and the resin was tumbled for 4 hr. Resin was then washed with DMF, MeOH, DMF, and MeOH. Molecular sieves removed and washed again.

#### *Removal of Acid-labile Protecting Groups.*

The resin was tumbled in 10 ml of 95% trifluoroacetic acid (TFA), 2.5% water, and 2.5% triisopropylsilane (TIS) for 1 hr. The solution was removed by vacuum filtration and the resin was washed with DMF, MeOH, DMF, and MeOH.

#### *MS/MS analysis.*

To cleave the peptide from the resin, the beads were treated with 50 µl of 40 mg/ml CNBr in 0.1 M HCl for 16 hr in the dark. The beads were removed from the solution and the excess CNBR, TFA, and water were removed under reduced pressure by using an AES1010 SpeedVac system. The cleaved peptides were re-dissolved in 10 µl of 0.1% TFA in water. 1 µl of matrix containing saturated α-cyano-4-hydroxycinnamic acid (CHCA) in 50:50:0.1 acetonitrile/water/TFA was positioned on to a MALDI plate and mixed with 2 µl of the cleaved SL solution. Samples were then analyzed by MALDI-TOF/MS and MS/MS techniques.

#### *Cell Culture.*

All cells were grown in T75 flasks at 37°C in 5% CO<sub>2</sub> and 95% air. CT-26 colon carcinoma cells and MCF7 human breast cancer cells were grown in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS) and 1% antibiotic/antimycotic. NIH/3T3 murine fibroblast cells were grown in DMEM with 10% neonatal calf serum and 1% antibiotic/antimycotic. LoVo human colorectal carcinoma

cells were grown in Roswell Park Memorial Institute (RPMI) media with 10% FBS and 1% antibiotic/antimycotic. MCF10A human breast cells were grown in DMEM/Nutrient F-12 Ham (DMEM/F12) with 5% horse serum, 10 µg/ml of insulin, 20 ng/ml of epidermal growth factor (EGF), 100 ng/ml of cholera toxin, and 0.5 µg/ml hydrocortisone.

*FITC-labeling Cell Membrane Extracts.*

After cells were grown to 80% confluence, the corresponding media was added and cells were scraped. The cells were then collected in 50 ml centrifuge tubes from four T75 flasks and centrifuged. Media was removed and PBS buffer was used to wash the cell pellets two times. The Qiagen Plasma Membrane Protein Kit was used to extract plasma membrane proteins and glycoproteins. The pH of the 2 ml extraction solution was adjusted to 8.8. FITC (.07 µmol, 28 µg) in DMF (5.3 µl) was incubated at 37°C for 1 hr. The solution was then transferred to a 3.5 kDa Slide-A-Lyzer cassette and dialyzed against 500 ml of screening buffer (100 mM NaH<sub>2</sub>PO<sub>4</sub>, 150 mM NaCl, 10% glycerol, pH 7.2) at 4°C. Screening buffer was changed every 6-8 hr 3 times. FITC-labeled extracts were collected and concentrations were obtained through Absorbance and fluorescence analysis.

*General procedure for the SL array.*

2 mg of each SL were weighed in 1.5 ml microcentrifuge tubes and rinsed with 1 ml of PBS buffer for 10 min. The PBS buffer was then removed and the beads were incubated in 1% BSA for 15 min to reduce non-specific binding. The beads were then treated with 1 ml of 20 fold dilutions of FITC-labeled cell membrane extracts and tumbled for 6 hr at room temperature. Beads were then washed with PBS three times,



imaged with a Leica MZ 16F microscope with a GFP filter set, and analyzed by Matlab R2012a. The output was analyzed by linear discriminant analysis using SYSTAT 11.00.01.

## CHAPTER 5

### CONCLUSION

Aberrant glycosylation takes place during the onset of cancer and throughout the progression of the disease allowing certain glycans to serve as biomarkers<sup>1</sup>. Boronic acid functionalized synthetic lectins (SLs) have been shown to bind selectively and cross-reactively to certain glycoproteins through covalent yet reversible interactions between the boronic acids and 1,2- or 1,3-diols found on saccharides<sup>2</sup>.

In chapter 2, an SL library was screened against fluorescently-labeled prostate membrane glycoproteins and proteins in order to create a prostate specific array. This was achieved by screening with fluorescently labeled prostate membrane proteins and glycoproteins. The hits selected using the aforementioned screening method, were resynthesized and incubated with prostate cell lines with varying metastatic potentials as well as colon cell lines with varying metastatic potentials, as seen in Chapter 3. The results indicated that when using the prostate selected SLs classification accuracies were high for the discrimination of prostate cell lines by metastatic potential. Classification accuracy improved when including these prostate selected SLs with the previous SL array with colon and prostate cell lines.

The last piece of this project involved investigating data acquisition parameters and how these could be adjusted in order to obtain the most accurate representation of the SL-glycoprotein interaction. This was achieved through including a region of interest

(ROI) as well as using all color spaces when collecting data from fluorescent images captured.

There are many different directions that the SL array can move towards in the future. The arrays should be applied to more types of cancers to assess the versatility of the arrays' differentiation abilities. It is important to evaluate the current SLs' binding affinities for different cancer associated glycans in order to optimize the SL array. Also, the new library screening discussed in this dissertation can be used to find SLs with greater selectivity. The results presented in this dissertation highlight the immense potential of SL arrays as tools in cancer diagnostics.

## REFERENCES

1. Hollingsworth, M. A.; Swanson, B. J., Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer* **2004**, *4* (1), 45-60.
2. Bicker, K. L.; Sun, J.; Harrell, M.; Zhang, Y.; Pena, M. M.; Thompson, P. R.; Lavigne, J. J., Synthetic lectin arrays for the detection and discrimination of cancer associated glycans and cell lines. *Chemical Science* **2012**, *3* (4), 1147-1156.
3. Howlader N, N. A., Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA, SEER Cancer Statistics Review 2013.
4. Choi, Y.-E.; Kwak, J.-W.; Park, J. W., Nanotechnology for Early Cancer Detection. *Sensors* **2010**, *10* (1).
5. Brenner, H.; Hoffmeister, M.; Arndt, V.; Stegmaier, C.; Altenhofen, L.; Haug, U., Protection From Right- and Left-Sided Colorectal Neoplasms After Colonoscopy: Population-Based Study. *Journal of the National Cancer Institute* **2010**, *102* (2), 89-95.
6. Singh, H.; Nugent, 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* *139* (4), 1128-1137.
7. Gupta, A. K.; Brenner, D. E.; Turgeon, D. K., Early Detection of Colon Cancer. *Molecular Diagnosis & Therapy* **2008**, *12* (2), 77-85.
8. Barry, M. J., Prostate-Specific-Antigen Testing for Early Diagnosis of Prostate Cancer. *New England Journal of Medicine* **2001**, *344* (18), 1373-1377.
9. Heidenreich, A.; Bellmunt, J.; Bolla, M.; Joniau, S.; Mason, M.; Matveev, V.; Mottet, N.; Schmid, H.-P.; van der Kwast, T.; Wiegand, T.; Zattoni, F., EAU Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Treatment of Clinically Localised Disease. *European Urology* **2011**, *59* (1), 61-71.
10. Fakih, M. G., Carcinoembryonic Antigen Monitoring in Metastatic Colorectal Cancer: Words of Caution. *Journal of Clinical Oncology* **2008**, *26* (34), e7-e7.
11. Bast, R. C.; Ravdin, P.; Hayes, D. F.; Bates, S.; Fritsche, H.; 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\*. *Journal of Clinical Oncology* **2001**, *19* (6), 1865-1878.
12. Levi, Z.; Birkenfeld, 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 Journal of Cancer* **2011**, *128* (10), 2415-2424.
13. van Rossum, L. G.; van Rijn, A. F.; Laheij, R. J.; van Oijen, M. G.; Fockens, P.; van Krieken, H. H.; Verbeek, A. L.; Jansen, J. B.; Dekker, E., Random Comparison of

Guaiac and Immunochemical Fecal Occult Blood Tests for Colorectal Cancer in a Screening Population. *Gastroenterology* **135** (1), 82-90.

14. Hol, L.; Wilschut, J. A.; van Ballegooijen, M.; van Vuuren, A. J.; van der Valk, H.; Reijerink, J. C. I. Y.; van der Togt, A. C. M.; Kuipers, E. J.; Habbema, J. D. F.; van Leerdam, M. E., Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. *British Journal of Cancer* **2009**, *100* (7), 1103-1110.

15. 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*, 14.

16. Lodish H, B. A., Zipursky SL, et al., Section 17.7 Protein Glycosylation in the ER and Golgi Complex. In *Molecular Cell Biology. 4th edition.*, W. H. Freeman: New York, 2000.

17. Adamczyk, B.; Tharmalingam, T.; Rudd, P. M., Glycans as cancer biomarkers. *Biochim Biophys Acta* **2012**, *1820* (9), 1347-53.

18. Dube, D. H.; Bertozzi, C. R., Glycans in cancer and inflammation [mdash] potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **2005**, *4* (6), 477-488.

19. Bicker, K. L.; Sun, J.; Lavigne, J. J.; Thompson, P. R., Boronic acid functionalized peptidyl synthetic lectins: Combinatorial library design, peptide sequencing, and selective glycoprotein recognition. *ACS combinatorial science* **2011**, *13* (3), 232-243.

20. Oshovsky, G. V.; Reinhoudt, D. N.; Verboom, W., Supramolecular Chemistry in Water. *Angewandte Chemie International Edition* **2007**, *46* (14), 2366-2393.

21. 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. *The Journal of Organic Chemistry* **2009**, *74* (11), 4055-4060.

22. 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. *Journal of the American Chemical Society* **1992**, *114* (14), 5874-5875.

23. 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.

24. 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.

25. 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.

26. 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* **11** (4), 439-448.

27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. Zou, Y.; Broughton, D. L.; Bicker, K. 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.
37. Klecka, W. R., *Discriminant Analysis*. SAGE Publications: 1980.

## APPENDIX A:

### SL IMAGING MATLAB CODE

This Appendix is the MATLAB code written by Dr. Greytak that was used to analyze images taken of the SLs. He incorporated many different features that could be used for different types of analyses. Even though the complete code is below, only certain sections are used for SL image analysis. The black lines separate the sections. Sections that are not used in the analysis are in light blue and should be skipped when running the MATLAB program.

```
% For analysis images of silica particles
% 7/11/2012



---


%% setup

% Define colormaps (all length 256)
redmap=zeros(256,3);
redmap(:,1)=linspace(0,1,256);
greenmap=zeros(256,3);
greenmap(:,2)=linspace(0,1,256);
bluemap=zeros(256,3);
bluemap(:,3)=linspace(0,1,256);
ratiomap=zeros(256,3);
ratiomap(1:128,1)=linspace(0,1,128);
ratiomap(129:256,1)=1;
ratiomap(1:128,2)=1;
ratiomap(129:256,2)=linspace(1,0,128);
ratiomap(1,:)=0;

ratio_min=0.0;
ratio_max=1.0;

ratiolevels=linspace(ratio_min,ratio_max,101);
ratiolevels=(ratiolevels(2:end) + ratiolevels(1:end-1))./2;

% sets the bins used to track pixel or average brightness
levels=0:255;
```

```

% levels=(levels(1:end-1)+levels(2:end))./2; % gets midpoints between
level centers

% set thresholding values to exclude dark and/or saturated regions
red_squelch=0;
red_limit=+Inf;
% green_squelch=0;
% green_limit=+Inf;
green_squelch=0;
green_limit=+Inf; % 8-bit image: 255 would definitely be clipping
blue_squelch=0;
blue_limit=+Inf;

% allows you to blur incoming images if they are pixellated
h=1;
% blur_radius=1;
% h=fspecial('gaussian',21,blur_radius);
%

labelstyle='particle'; % 'grid', 'particle', or 'single'



---


%% define dataset

indir='Z:\Huang, Rong\RH 082415 pictures\Pictures After Conversion/';
clear outfile
% comment out below to suppress all output files
outdir=['Z:\Huang, Rong\RH Matlab\082515 RH PC3 rwpe_1 SL11, Mut1 mut2
mut3/'];
outfile='test';

% set up a list of files corresponding to a single (or each) experiment

% experiment 1

files=dir([indir '*.jpg']);
for(i=1:length(files))
    inputfile(i).rgbfile=files(i).name;
end

% initialize the (manual) region of interest
clear red merged
roix=[];
roiy=[];
merged=imread([indir inputfile(1).rgbfile]);
red=merged(:,:,1);
myroi=logical(ones(size(red)));

clear red merged files



---


%% Correction factors for inhomogenous illumination/detection
% from 6-12B solution

```



```

%
ff_redfile='\andrew\aaa_svn\z_old\research\work\imaging\80801_rox_analy
sis\ff_red_6_12b.mat';
%
ff_greenfile='\andrew\aaa_svn\z_old\research\work\imaging\80511_mouse_p
H_sensor\ff_red_6_12b.mat';
%
% % load correction factors
% load(ff_redfile);
% load(ff_greenfile);

% create correction factors
% ff_red=mean(red(myroi))./double(red);
% ff_green=mean(green(myroi))./double(green);
%
% % save
% save(ff_redfile,'ff_red');
% save(ff_greenfile,'ff_green');

%% index slices required

% this would let us choose a manual region of interest to use for all
% images
% for(thelocation=1:length(location))
%     location(thelocation).myroi=[];
%     location(thelocation).roix=[];
%     location(thelocation).roiy=[];
% end

% initialize -- set histograms for pixel intensities and label
(particle
% average) intensities to zero
for(theinputfile=1:length(inputfile))
    inputfile(theinputfile).redhist ...
        =zeros(size(levels));
    inputfile(theinputfile).greenhist ...
        =zeros(size(levels));
    inputfile(theinputfile).bluehist ...
        =zeros(size(levels));
    inputfile(theinputfile).brighthist ...
        =zeros(size(levels));
    inputfile(theinputfile).validlabels ...
        =[];
end

clear newslice thelocation theslice thefile

%% show an image
theinputfile=1;
% myroi=location(currentlocation).myroi;
% roix=location(currentlocation).roix;
% roiy=location(currentlocation).roiy;

titlestr=['Input file ' num2str(theinputfile)];

```

```

% aggregate images
clear red green blue merged
merged=imread([indir inputfile(theinputfile).rgbfile]);
red=merged(:,:,1);
green=merged(:,:,2);
blue=merged(:,:,3);

% merged
figure(1);clf;
%imshow(merged(:,:,:).*(2^16/(double(max(merged(myroi))))))
imshow(merged)
title(titlestr);
text(0.02*max(xlim),0.02*max(ylim),titlestr,'color','w')

% % red slice
figure(2);clf;
%imshow(red,[0 max(stretchlim(red,0.00001))]*2^16)
imshow(red,[0 max(red(myroi))])
colormap(redmap); colorbar;
title(titlestr);
%
% % green slice
figure(3);clf;
%imshow(green,[0 max(stretchlim(green,0.00001))]*2^16)
imshow(green,[0 max(green(myroi))])
colormap(greenmap); colorbar;
title(titlestr)
%
% % blue slice
figure(4);clf;
%imshow(blue,[0 max(stretchlim(blue,0.00001))]*2^16)
imshow(blue,[0 max(blue(myroi))])
colormap(blue); colorbar;
title(titlestr)

% display ROI outline
if(exist('roix','var') && length(roix)>2)
    for(i=find(ishandle(1:4)))
        figure(i);
        line(roix,roiy,'color','w','linestyle','-','linewidth',2);
        line(roix,roiy,'color','k','linestyle':'','linewidth',2);
    end
end

%% Pick ROI
myroi=true(size(red)); % initialize all-on, pick subregion only if we
want
roix=[];
roiy=[];
% uncomment if you want to pick a subregion
delete(findall(gcf,'type','line'));
[myroi roix roiy]=roipoly;
% display ROI outline
if(length(roix)>2)

```

```

        line(roix,roiy,'color','w','linestyle','-','linewidth',2);
        line(roix,roiy,'color','k','linestyle',':','linewidth',2);
end

%% Reset ROI
myroi=true(size(red)); % initialize all-on, pick subregion only if we
want
roix=[];
roiy=[];

%% Save an ROI
save annas_roi myroi roix roiy

%% Load an ROI
load 'Z:\Veldkamp\Data\2012 Matlab\myroi.mat'
% load myroi ... load Hannah's ROI
% display ROI outline
if(exist('roix','var') && length(roix)>2)
    for(i=find(ishandle(1:4)))
        figure(i);
        line(roix,roiy,'color','w','linestyle','-','linewidth',2);
        line(roix,roiy,'color','k','linestyle',':','linewidth',2);
    end
end

%% compute statistics for each slice
clear red green blue displaymask
myinputfiles=1:276;
% myinputfiles=1:length(inputfile); % all files
showimages=0; % set to 1 if you want it to display each time
saveimages=0; % write labeled images out to output directory ...
doesn't work yet

if(exist('outfile','var'))
    if(~exist(outdir,'dir'))
        mkdir(outdir);
    end
end
particletable=fopen([outdir outfile '_particles.txt'],'w');

% now gather values for selected slices
for(theinputfile=myinputfiles)
    % print to command line the index of the file we are currently
studying
    fprintf(1,'\n Inputfile %d',theinputfile);

    % read data from the current input image of interest
clear red green blue merged label
merged=imread([indir inputfile(theinputfile).rgbfile]);
red=merged(:,:,1);
green=merged(:,:,2);
blue=merged(:,:,3);

    % Gaussian filter -- blur images
%     red=imfilter(red,h,'replicate');
%     green=imfilter(green,h,'replicate');

```

```

% blue=imfilter(blue,h,'replicate');

% Apply correction factors
if(exist('ff_red','var'))
    red=uint16(double(red).*ff_red);
end
if(exist('ff_green','var'))
    green=uint16(double(green).*ff_green);
end

% Background subtraction
% not used here... ideally uses background image

if(showimages==1)
    figure(1);clf
    imshow(merged,stretchlim(merged))
    titlestr=['Merged: Inputfile ' num2str(theinputfile) ': '
inputfile(theinputfile).rgbfile];
    text(0.02*max(xlim),0.02*max(ylim),titlestr,'color','w')
    if(exist('roix','var') && length(roix)>2)
        line(roix,roiy,'color','w','linestyle','-','linewidth',2);
        line(roix,roiy,'color','k','linestyle',':','linewidth',2);
    end
end
% Thresholding
% full auto
% mymask = im2bw(merged,graythresh(merged));
mymask = im2bw(green,graythresh(green));

% threshold based on some multiple of background intensity
bg_factor = 3; % background is anything less than this times the
average
bg_radius = 108; % radius by which we grow the particle mask to
find bg
bg_img=rgb2gray(merged); % grayscale image, could use a channel
bg_mask = ~imdilate(mymask,strel('octagon',bg_radius));
bg_level = mean(bg_img(bg_mask));
mymask = bg_img>(bg_level.*bg_factor);
% manual
mymask = mymask & ...
    (red <= red_limit) & (red >= red_squelch) & ...
    (green <= green_limit) & (green >= green_squelch) & ...
    (blue <= blue_limit) & (blue >= blue_squelch);

% include a manually-chosen region of interest if desired ... if it
% doesn't exist, we create one that includes all of the image
if(~exist('myroi','var'))
    myroi=true(size(red)); roix=[]; roiy=[];
end

if(showimages == 1) % display threshold mask and ROI
    figure(2);
    imshow(mymask & myroi,[]);
    titlestr=['mymask & myroi: Inputfile ' num2str(theinputfile)
': ' inputfile(theinputfile).rgbfile];
    text(0.02*max(xlim),0.02*max(ylim),titlestr,'color','w')

```

```

% draw ROI box
if(exist('roix','var') && length(roix)>2)
    line(roix,roi,'color','w','linestyle','-','linewidth',2);
    line(roix,roi,'color','k','linestyle',':','linewidth',2);
end
end

% Break image into labeled regions
mylabel=uint16(zeros(size(red)));
switch labelstyle
case 'grid'
    % impose 4x4 grid at image center
    clear x y
    x=floor((0:size(red,2)-1)/(size(red,2)/6));
    y=floor((0:size(red,1)-1)/(size(red,1)/6));
    x(x>4)=0;
    y(y>4)=0;
    [x y]=meshgrid(x,y);
    x(y==0)=0;
    y(x==0)=0;
    y=(y-1)*4;
    y(y<0)=0;
    mylabel=x+y;
    clear x y
    mylabel(~mymask)=0;
    mylabel(~myroi)=0;
case 'particle'
    % break image into labeled thresholded islands (e.g.,
cells)
    mylabel=bwlabel(mymask);
    % we don't kill outside roi for particles: can be more
clever
    % below
otherwise
    % default (trivial case): single tile for entire image
    mylabel(:)=1;
    mylabel(~mymask)=0;
    mylabel(~myroi)=0;
end

% Display all labeled regions for current slice
if(showimages==1)
    figure(3);
    titlestr=['All labels: Inputfile ' num2str(theinputfile) ': '
inputfile(theinputfile).rgbfile];
    imshow(mylabel.*mymask,[]);colormap('colorcube')
    text(0.02*max(xlim),0.02*max(ylim),titlestr,'color','w')
    % draw ROI box
    if(exist('roix','var') && length(roix)>2)
        line(roix,roi,'color','w','linestyle','-','linewidth',2);
        line(roix,roi,'color','k','linestyle',':','linewidth',2);
    end
end

% Compute statistics by pixel (actually just histograms here;
% averages and variances can be determined later for display)

```

```

redhist=hist(red(mymask & myroi),levels);
greenhist=hist(green(mymask & myroi),levels);
bluehist=hist(blue(mymask & myroi),levels);
brightimg=rgb2gray(merged);
brighthist=hist(brightimg(mymask & myroi),levels);

        % compute ratio
%         ratio=zeros(size(red));
%         ratio(mymask) = ...
%             double(red(mymask)) ./ (double(red(mymask)) +
double(green(mymask)));
%         ratiohist=hist(ratio(mymask),ratiolevels);

% Compute statistics for all color channels, by label
validlabels=false(1,max(mylabel(:)));
for(j=1:max(mylabel(:)))
    if(~(mylabel==j)) % mostly for grid-based labels: skip if no
pixels have this label
        continue
    end
    % get particle properties -- characterize entire particle even
if
    % part is OUTSIDE of ROI
    % mean intensity
    label(j).red=mean(red(mylabel==j));
    label(j).green=mean(green(mylabel==j));
    label(j).blue=mean(blue(mylabel==j));
    label(j).bright=mean(brightimg(mylabel==j));
    % std dev within each label
    label(j).redstd=std(double(red(mylabel==j)));
    label(j).greenstd=std(double(green(mylabel==j)));
    label(j).bluestd=std(double(blue(mylabel==j)));
    label(j).brightstd=std(double(brightimg(mylabel==j)));
    % get percentile values within each label
    tolerance=[0.2 0.8];
    label(j).redhi=255*max(stretchlim(red(mylabel==j),tolerance));

label(j).greenhi=255*max(stretchlim(green(mylabel==j),tolerance));

label(j).bluehi=255*max(stretchlim(blue(mylabel==j),tolerance));

label(j).brighthi=255*max(stretchlim(brightimg(mylabel==j),tolerance));

    labelprops=regionprops((mylabel==j) &
mymask), 'Area', 'Perimeter', 'Centroid');
%     label(j).area=labelprops.Area;
    label(j).area=sum(mylabel(:)==j); % more reliable: works if
disconnected
    label(j).perimeter=labelprops.Perimeter;
    label(j).circ=4*pi()* (label(j).area/(label(j).perimeter^2));
    label(j).centroid=cat(1,labelprops.Centroid);

    validlabels(j)=true; % innocent until proven guilty
    % disqualify labels that extend outside ROI borders
    if(find(mylabel==j & ~myroi))

```

```

        validlabels(j)=false;
    end
    % disqualify labels that touch ROI edge (including edge of
    % image)
    if(find(mylabel==j & bwperim(myroi)))
        validlabels(j)=false;
    end
    % disqualify ONLY labels that lie ENTIRELY outside ROI
    if(find(mylabel==j & ~myroi))
        validlabels(j)=false;
    if(find(mylabel==j & myroi))
        validlabels(j)=true;
    end
    end
    % disqualify particles that we don't like
    if(label(j).area < 5000)
        validlabels(j)=false;
    end
    % disqualify particles with saturated regions
    if(label(j).greenhi >=250)
        validlabels(j)=false;
    end
    % disqualify particles on circularity
    if(label(j).circ < 0.85)
        validlabels(j)=false;
    end
end

if(showimages==1 || saveimages==1) % what is this?
myvalidlabel=mylabel;
for(j=1:length(validlabels))
    if(validlabels(j) == 0)
        myvalidlabel(myvalidlabel==j)=0;
    end
end
end
if(showimages==1)
    % make a copy of labeled indexed color image, and delete
invalid
    % particles
    figure(4);
    imshow(myvalidlabel,[]);colormap('colorcube')
    titlestr=['Valid labels: Inputfile ' num2str(theinputfile) ':
' inputfile(theinputfile).rgbfile];
    text(0.02*max(xlim),0.02*max(ylim),titlestr,'color','w')
    % draw ROI box
    if(exist('roix','var') && length(roix)>2)
        line(roix,roi,'color','w','linestyle','-','linewidth',2);
        line(roix,roi,'color','k','linestyle',':','linewidth',2);
    end
end
if(saveimages==1)
    % This doesn't work right now.
    %
    % imwrite(ind2rgb(myvalidlabel,'colorcube'),[outdir
outfile '_' num2str(theinputfile,'%3d') '.png'],'png');
end

```

```

end

% Copy values for current image file to inputfile structure
inputfile(theinputfile).redhist ...
    =redhist;
inputfile(theinputfile).greenhist ...
    =greenhist;
inputfile(theinputfile).bluehist ...
    =bluehist;
inputfile(theinputfile).brighthist ...
    =brighthist;
%
    inputfile(theinputfile).ratiohist ...
%
    =ratiohist;
inputfile(theinputfile).validlabels ...
    =validlabels;
inputfile(theinputfile).label ...
    =label;

% print particle statistics to logfile
labelfields=fieldnames(label);

% print header row, if desired, if it hasn't been printed already
if(ftell(particletable)==0) % tell us if we're at the beginning
of the file
    % file name
    fprintf(particletable,'Inputfile
',inputfile(theinputfile).rgbfile);
    % print file number, particle number, valid yes/no
    fprintf(particletable,'FileNo. ParticleNo. Valid ');
    % print label data fields
    for(k=1:length(labelfields))
        fprintf(particletable,'%s ',labelfields{k});
    end
    % end of line -- formally \n will do but \r\n makes readable in
Notepad
    fprintf(particletable,'\r\n');
end

% generate a line for each particle
for(j=1:length(validlabels))
    % print input filename, if desired
    fprintf(particletable,'%s ',inputfile(theinputfile).rgbfile);
    % print file number, particle number, valid yes/no
    fprintf(particletable,'%d %d %d ', ...
        theinputfile,j,validlabels(j));
    % print label data fields
    for(k=1:length(labelfields))
        fprintf(particletable,'%d ',label(j).(labelfields{k}));
    end
    % end of line -- formally \n will do but \r\n makes readable in
Notepad
    fprintf(particletable,'\r\n');
end

% Build up the union of the masks of (valid) labels for all images,
for

```



```

% diagnostic purposes
if( ~exist('displaymask','var') )
    displaymask=zeros(size(red));
end
if(exist('displaymask','var') )
    % all masks
    displaymask=displaymask | (mymask & myroi);
    % only valid
    for(j=find(validlabels))
        displaymask=displaymask | (mylabel==j);
    end
end

end

fclose(particletable);
save([outdir outfile '_data.mat'],'inputfile');

if(exist('displaymask','var'))
    figure(5);clf
    imshow(displaymask)
    titlestr=['All labels, all files'];
    text(0.02*max(xlim),0.02*max(ylim),titlestr,'color','w')
    % draw ROI box
    if(exist('roix','var') && length(roix)>2)
        line(roix,roiy,'color','w','linestyle','-','linewidth',2);
        line(roix,roiy,'color','k','linestyle':','linewidth',2);
    end
end

clear red green blue redhist greenhist bluehist ratiohist
clear validlabels labelratio labelred labelblue labelgreen
clear labelredhist labelgreenhist labelratiohist
clear i j
clear mymask particletable



---


%% Load saved data set -- from current directory, not outdir
load('test_data.mat');



---


%% Display aggregated statistics for a set of slices
% pick slices
myinputfiles=1:4;

titlestr=['Inputfiles          '          num2str(myinputfiles(1))          '- '
num2str(myinputfiles(end))];

% initialize totals
redhist=zeros(size(levels));
greenhist=zeros(size(levels));
bluehist=zeros(size(levels));
brighthist=zeros(size(levels));
% ratiohist=zeros(size(ratiolevels));
labelredhist=zeros(size(levels));
labelgreenhist=zeros(size(levels));
labelbluehist=zeros(size(levels));
labelbrighthist=zeros(size(levels));

```

```

% labelratiohist=zeros(size(ratiolevels));

% aggregate data from all slices
for(theinputfile=myinputfiles)
    redhist=redhist ...
        +inputfile(theinputfile).redhist;
    greenhist=greenhist ...
        +inputfile(theinputfile).greenhist;
    bluehist=bluehist ...
        +inputfile(theinputfile).bluehist;
    brighthist=brighthist ...
        +inputfile(theinputfile).brighthist;

    validlabels=inputfile(theinputfile).validlabels;
    % opportunity to adjust choice of valid labels
    validlabels([inputfile(theinputfile).label(:).area] >
1000)=true;
    labelredhist=labelredhist ...

+hist([inputfile(theinputfile).label(validlabels).red],levels);
    labelgreenhist=labelgreenhist ...

+hist([inputfile(theinputfile).label(validlabels).green],levels);
    labelbluehist=labelbluehist ...

+hist([inputfile(theinputfile).label(validlabels).blue],levels);
    labelbrighthist=labelbrighthist ...

+hist([inputfile(theinputfile).label(validlabels).bright],levels);

%         labelratiohist=labelratiohist ...
%
+location(thelocation).time(thetime).slices(theslice).labelratiohist;
    end

% compute mean and standard deviation for each metric
redmean=sum(levels.*redhist) / sum(redhist);
redstd= sqrt( sum( redhist.*((levels - redmean).^2) ) ...
    / sum(redhist));
greenmean=sum(levels.*greenhist) / sum(greenhist);
greenstd= sqrt( sum( greenhist.*((levels - greenmean).^2) ) ...
    / sum(greenhist));
% ratiomean=sum(ratiolevels.*ratiohist) / sum(ratiohist);
% ratiostd= sqrt( sum( ratiohist.*((ratiolevels - ratiomean).^2) ) ...
%     / sum(ratiohist));
labelredmean=sum(levels.*labelredhist) / sum(labelredhist);
labelredstd= sqrt( sum( labelredhist.*((levels - labelredmean).^2) )
...
    / sum(labelredhist));
labelgreenmean=sum(levels.*labelgreenhist) / sum(labelgreenhist);
labelgreenstd= sqrt( sum( labelgreenhist.*((levels -
labelgreenmean).^2) ) ...
    / sum(labelgreenhist));
%         labelratiomean=sum(ratiolevels.*labelratiohist) /
sum(labelratiohist);

```

```

% labelratiostd= sqrt( sum( labelratiohist.*((ratiolevels -
labelratiomean).^2) ) ...
% / sum(labelratiohist));

% plot data
figure(6);clf;

% xlims=[0 max(levels([find(redhist,1,'last')
find(greenhist,1,'last')]))];
% xlims=[0 500*ceil(max(xlims)/500)];
xlims=[0 256];

% pixel data
subplot(2,2,1);
handles=bar(levels,redhist);
set(handles,'FaceColor','r','EdgeColor','none','BarWidth',1.0);
xlim(xlims);
line([redmean redmean],ylim,'color','k');

subplot(2,2,3);
handles=bar(levels,greenhist);
set(handles,'FaceColor',[0 0.7 0],'EdgeColor','none','BarWidth',1.0);
xlim(xlims);
line([greenmean greenmean],ylim,'color','k');

% subplot(3,2,5);
% handles=bar(ratiolevels,ratiohist);
% set(handles,'FaceColor','k','EdgeColor','none','BarWidth',1.0);
% line([ratiomean ratiomean],ylim,'color','r');

% label data
subplot(2,2,2);
handles=bar(levels,labelredhist);
set(handles,'FaceColor','r','EdgeColor','none','BarWidth',1.0);
xlim(xlims);
line([labelredmean labelredmean],ylim,'color','k');

subplot(2,2,4);
handles=bar(levels,labelgreenhist);
set(handles,'FaceColor',[0 0.7 0],'EdgeColor','none','BarWidth',1.0);
xlim(xlims);
line([labelgreenmean labelgreenmean],ylim,'color','k');

% subplot(3,2,6);
% handles=bar(ratiolevels,labelratiohist);
% set(handles,'FaceColor','k','EdgeColor','none','BarWidth',1.0);
% line([labelratiomean labelratiomean],ylim,'color','r');

set(axes(),'Visible','off')
set(title(titlestr),'visible','on')

%% Scatter plot
myinputfiles=1:4;

```

```

titlestr=['Inputfiles      '      num2str(myinputfiles(1))      '- '
num2str(myinputfiles(end))];

xfield='blue';
yfield='green';
titlestr=['Inputfiles      '      num2str(myinputfiles(1))      '- '
num2str(myinputfiles(end)) ...
': ' yfield ' vs. ' xfield];

xvalues=[];
yvalues=[];
for(theinputfile=myinputfiles)
    validlabels=inputfile(theinputfile).validlabels;
    xvalues=[xvalues
[inputfile(theinputfile).label(validlabels).(xfield)]];
    yvalues=[yvalues
[inputfile(theinputfile).label(validlabels).(yfield)]];
end

figure(7);clf
plot(xvalues,yvalues,'bx');
% ylim([0 max(ylim)])
ylim([0 255])
% xlim([0 max(xlim)])
xlim([0 255])
title(titlestr);

%% Scatter plot -- overlay
myinputfiles=1:4;

titlestr=['Inputfiles      '      num2str(myinputfiles(1))      '- '
num2str(myinputfiles(end))];

xfield='area';
yfield='greenhi';
titlestr=['Inputfiles      '      num2str(myinputfiles(1))      '- '
num2str(myinputfiles(end)) ...
': ' yfield ' vs. ' xfield];

xvalues=[];
yvalues=[];
for(theinputfile=myinputfiles)
    validlabels=inputfile(theinputfile).validlabels;
    xvalues=[xvalues
[inputfile(theinputfile).label(validlabels).(xfield)]];
    yvalues=[yvalues
[inputfile(theinputfile).label(validlabels).(yfield)]];
end

figure(7);
hold on
plot(xvalues,yvalues,'ro');
% ylim([0 max(ylim)])
% ylim([0 255])
% xlim([0 max(xlim)])

```

```

% xlim([0 20000])
title(titlestr);

%% output settings data

if(exist('outfile','var'))
    if(~exist(outdir,'dir'))
        mkdir(outdir);
    end

    if(~isempty(roix))
        imwrite(myroi,[outdir outfile
'_roimask.tif'],'tif','compression','none');
    end

    % % save current channel and merged IMAGES
    % imwrite(red,[outdir outfile
'_red.tif'],'tif','Compression','none');
    % imwrite(green,[outdir outfile
'_green.tif'],'tif','Compression','none');
    % imwrite(blue,[outdir outfile
'_blue.tif'],'tif','Compression','none');
    % imwrite(merged,[outdir outfile
'_merged.tif'],'tif','Compression','none');
    %
    % save current channel and merged FIGURES
    % saveas(1,[outdir outfile '_merged.emf'],'emf');
    % saveas(2,[outdir outfile '_red.emf'],'emf');
    % saveas(3,[outdir outfile '_green.emf'],'emf');
    % saveas(4,[outdir outfile '_blue.emf'],'emf');

    % figure of collected statistics (fig 6)
    % saveas(6,[outdir outfileloc '_levels.emf'],'emf');

    % spit out a log file containing parameters
    logfile=fopen([outdir outfile '_log.txt'],'w');
    fprintf(logfile,'Input dir: %s \r\n',indir);
    fprintf(logfile,'Red squelch: %d \r\n',red_squelch);
    fprintf(logfile,'Red limit: %d \r\n',red_limit);
    fprintf(logfile,'Green squelch: %d \r\n',green_squelch);
    fprintf(logfile,'Green limit: %d \r\n',green_limit);
    fprintf(logfile,'Blue squelch: %d \r\n',green_squelch);
    fprintf(logfile,'Blue limit: %d \r\n',green_limit);
    if(length(h)>1)
        fprintf(logfile,'Blur radius: %f \r\n',blur_radius);
    end
    if(exist('ff_redfile','var'))
        fprintf(logfile,'Red correction file: %s \r\n',ff_redfile);
    end
    if(exist('ff_greenfile','var'))
        fprintf(logfile,'Green correction file: %s \r\n',ff_greenfile);
    end
    fprintf(logfile,'labelstyle: %s \r\n',labelstyle);

```

```
if(~isempty(roix))
    fprintf(logfile,'ROI: enabled \r\n');
else
    fprintf(logfile,'ROI: none selected \r\n');
end

fclose(logfile);
clear logfile

end
```

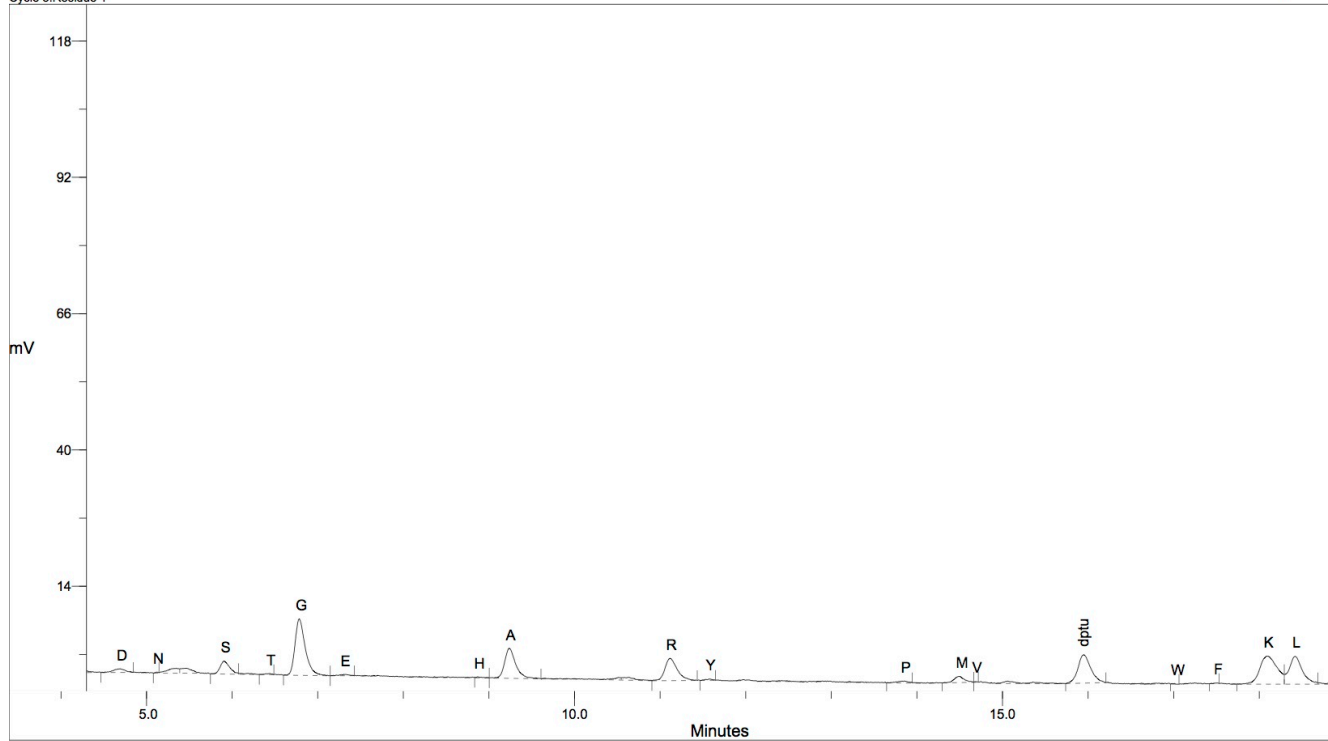
**APPENDIX B:**  
**EDMAN DEGREDDATION SEQUENCING**

The following figures are the Edman Degredation chromatograms received from Tufts University. Michael Berne, the director of the Core Facility, evaluated the sequences of the 5 new SLs (**Table B.1**).

Table B.1 SL10-14 sequences, selectivity and the analyte it was screened against.

<b>SL</b>	<b>Sequence</b>	<b>Screened Against</b>	<b>Selectivity</b>
<b>SINGLE ANALYTE SCREEN</b>			
SL10	R-L-D*-A-R-S-D*-G-B-B-R-M	F PC3	
SL13	R-T-D*-G-L-A-D*-V-B-B-R-M	F RWPE-1	
SL14	R-Y-D*-R-A-S-D*-V-B-B-R-M	R PC3	
<b>DUAL ANALYTE SCREEN</b>			
SL11	R-L-D*-Y-L-T-D*-R-B-B-R-M	F RWPE-1/R PC3	PC3
SL12	R-L-D*-G-F-Y-D*-Q-B-B-R-M	F RWPE-1/R PC3	RWPE-1

Cycle 3:Residue 1



PEAK ID	R.TIME (mins)	C.TIME (mins)	HEIGHT (mV)	PMOL HT	PEAK ID	R.TIME (mins)	C.TIME (mins)	HEIGHT (mV)	PMOL HT
D	4.69	4.59	0.674	1.851	P	13.85	13.80	0.373	1.393
N	5.12	5.13	0.185	0.513	M	14.50	14.48	1.174	3.766
S	5.90	5.87	2.461	9.619	V	14.69	14.72	0.129	0.420
T	6.44	6.38	0.202	1.030	dptu	15.95	15.93	5.411	22.244
G	6.78	6.75	10.781	40.418	W	17.01	16.94	0.106	0.353
E	7.31	7.28	0.254	0.750	F	17.51	17.50	0.103	0.337
H	8.87	8.85	0.170	0.734	K	18.09	18.18	5.359	13.752
A	9.24	9.21	5.700	19.312	L	18.43	18.41	5.298	18.579
R	11.11	11.11	4.233	33.055					
Y	11.57	11.54	0.254	0.865					

Friday, July 18, 2014 08:43:06

6FPC3-4 - Page 3 of 11

Figure 5.1 SL10 AA 8. The sequence of SL10 is R-L-D\*-A-R-S-D\*-G-B-B-R-M. R is a fixed position in the SL library.



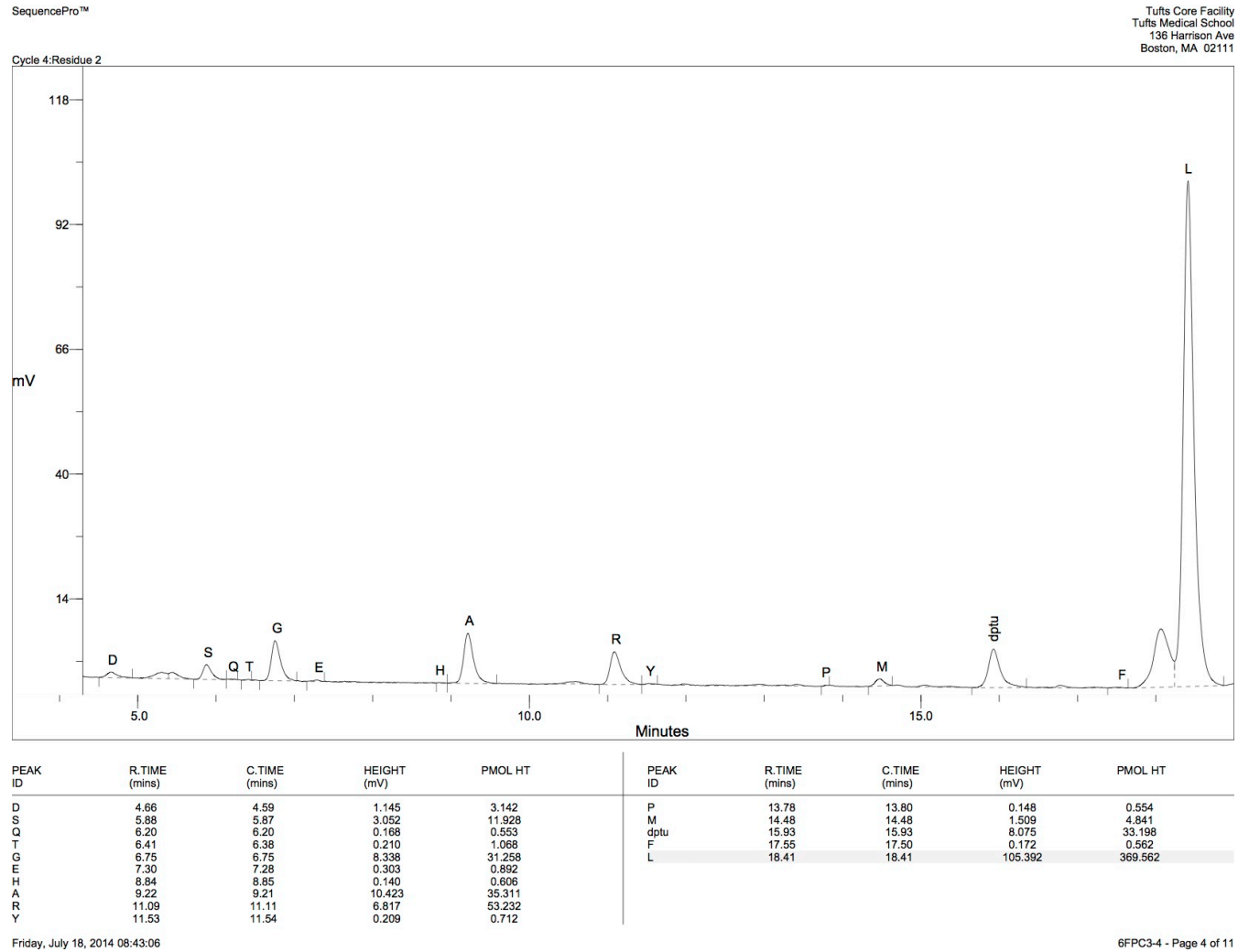


Figure 5.2: SL10 AA 7. The sequence of SL10 is R-L-D\*-A-R-S-D\*-G-B-B-R-M. L is a random position in the SL library.

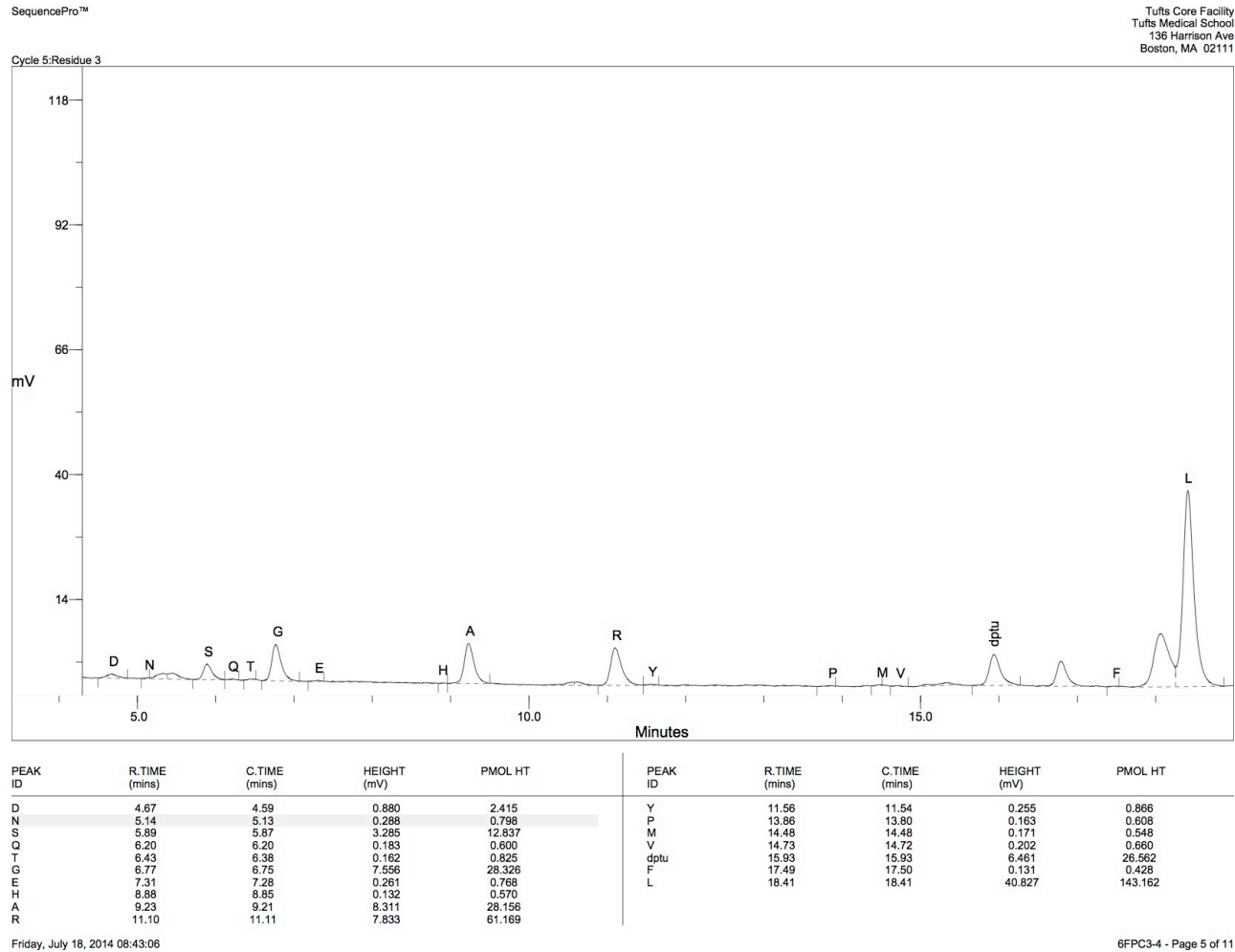


Figure 5.3: SL10 AA 6. The sequence of SL10 is R-L-D\*-A-R-S-D\*-G-B-B-R-M. D\* is a fixed position in the library and is not recognized by Edman Degradation. The fixed position of this AA identifies where it falls in the SL sequence. It is also important to note that while L does show in the cycle of the sequence, the amount decreases from the cycle before.

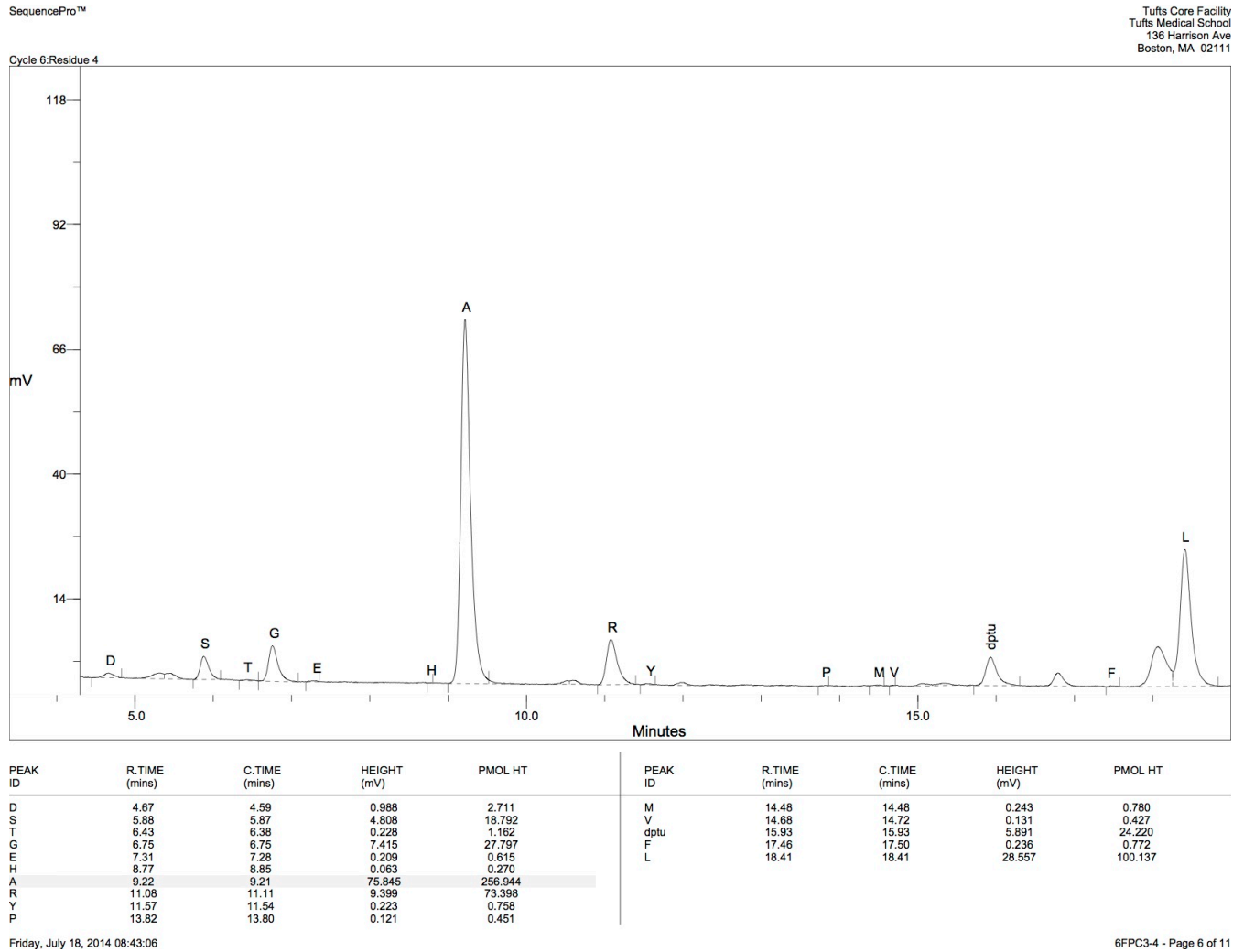


Figure 5.4: SL10 AA 5. The sequence of SL10 is R-L-D\*-A-R-S-D\*-G-B-B-R-M. A is a random position in the library.

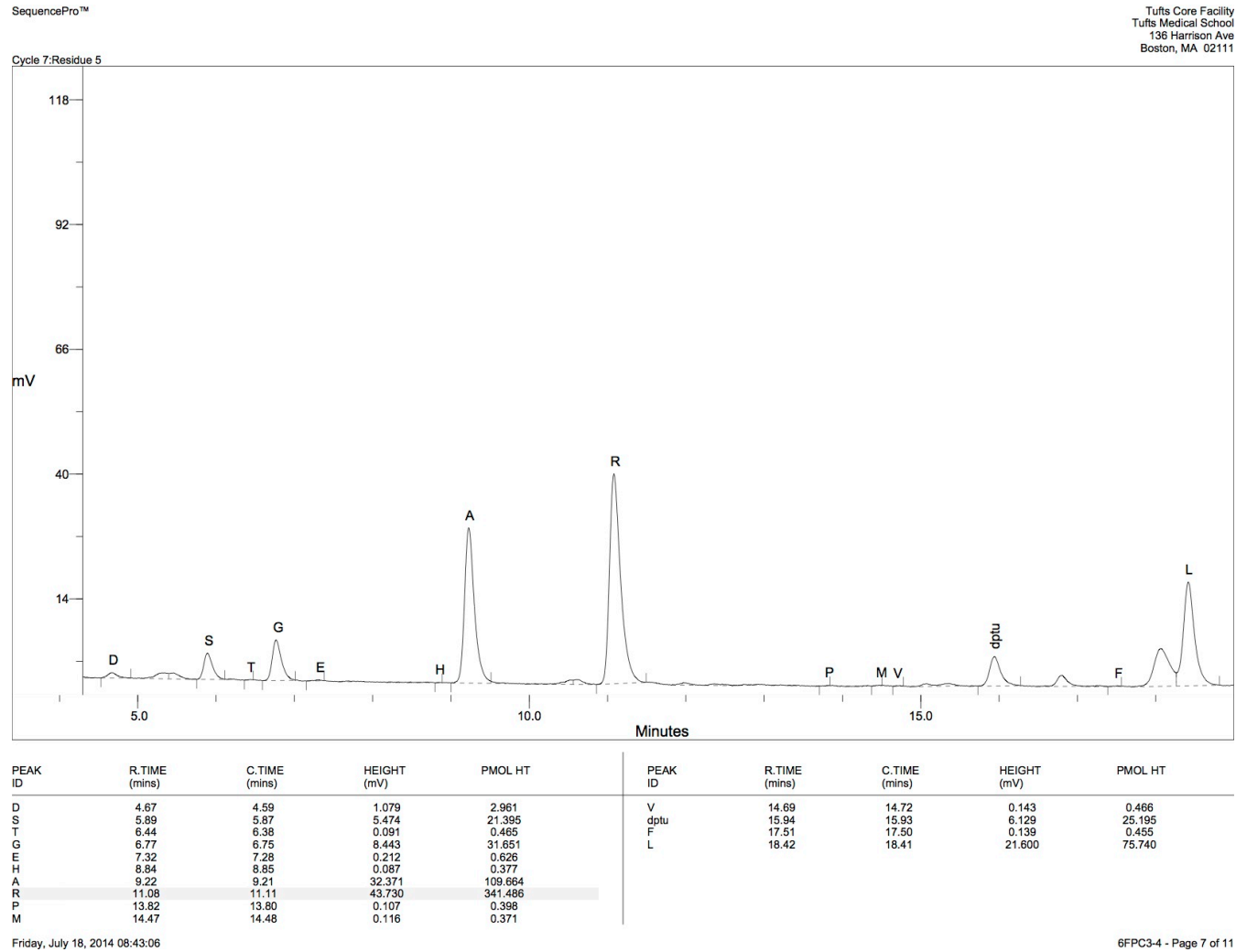


Figure 5.5: SL10 AA 4. The sequence of SL10 is R-L-D\*-A-R-S-D\*-G-B-B-R-M. R is a random position in the library.

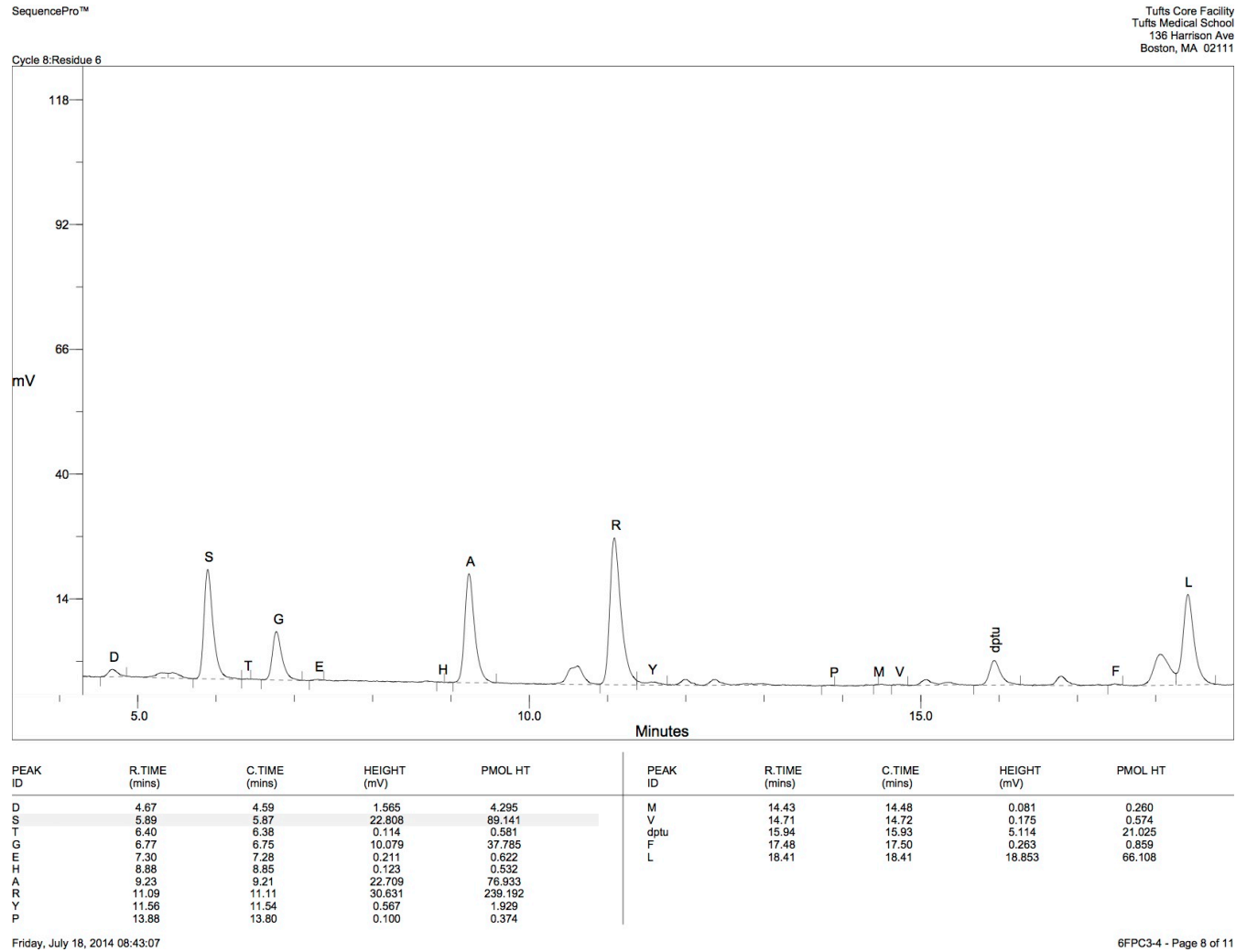


Figure 5.6: SL10 AA 3. The sequence of SL10 is R-L-D\*-A-R-S-D\*-G-B-B-R-M. S is a random position in the library.

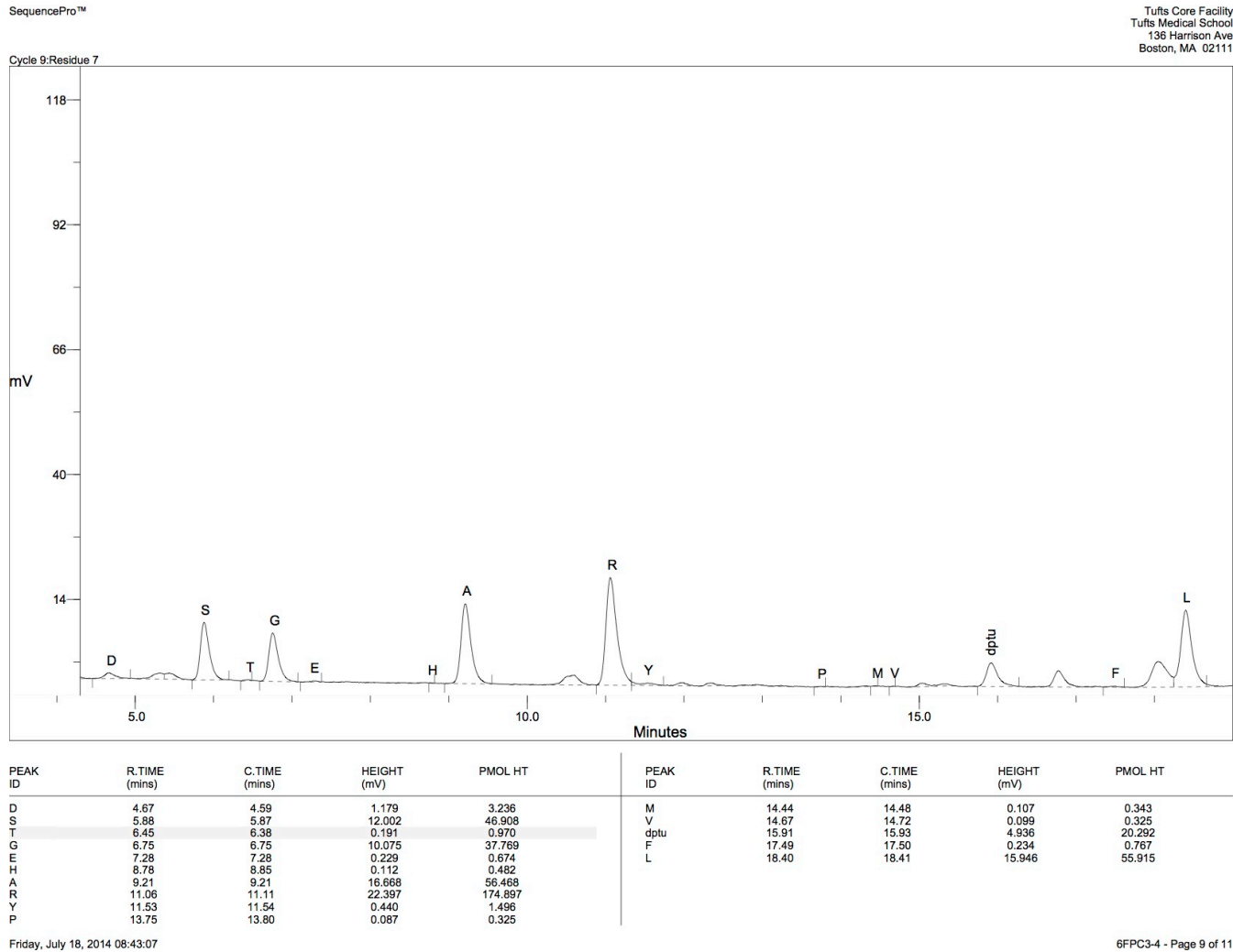


Figure 5.7: SL10 AA 2. The sequence of SL10 is R-L-D\*-A-R-S-D\*-G-B-B-R-M. D\* is a fixed position in the library. The fixed position of this AA identifies where it falls in the SL sequence.

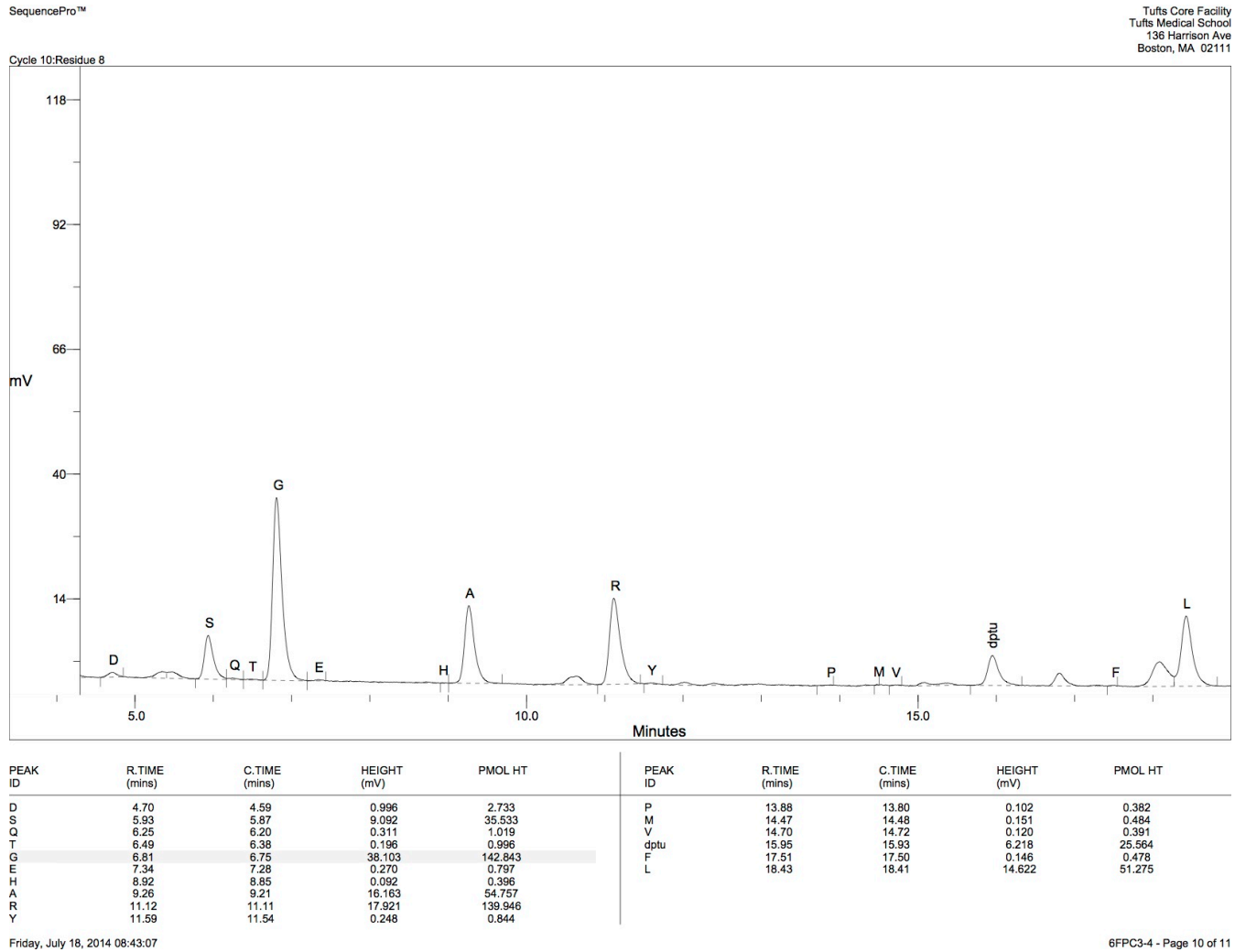


Figure 5.8: SL10 AA 1. The sequence of SL10 is R-L-D\*-A-R-S-D\*-G-B-B-R-M. G is a random position in the library.

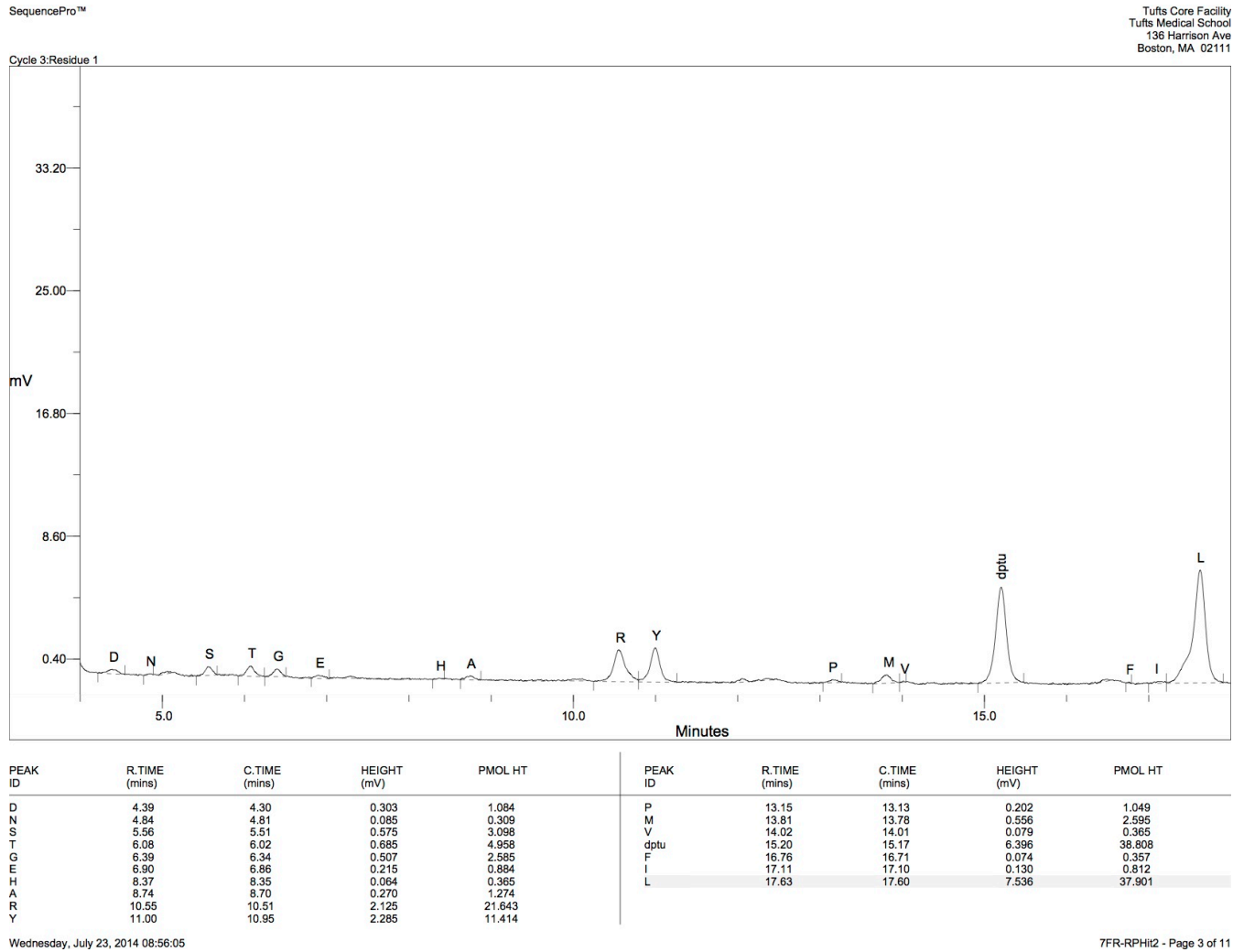


Figure 5.9: SL11 AA 8. The sequence of SL11 is R-L-D\*-Y-L-T-D\*-R-B-B-R-M. R is a fixed position in library.



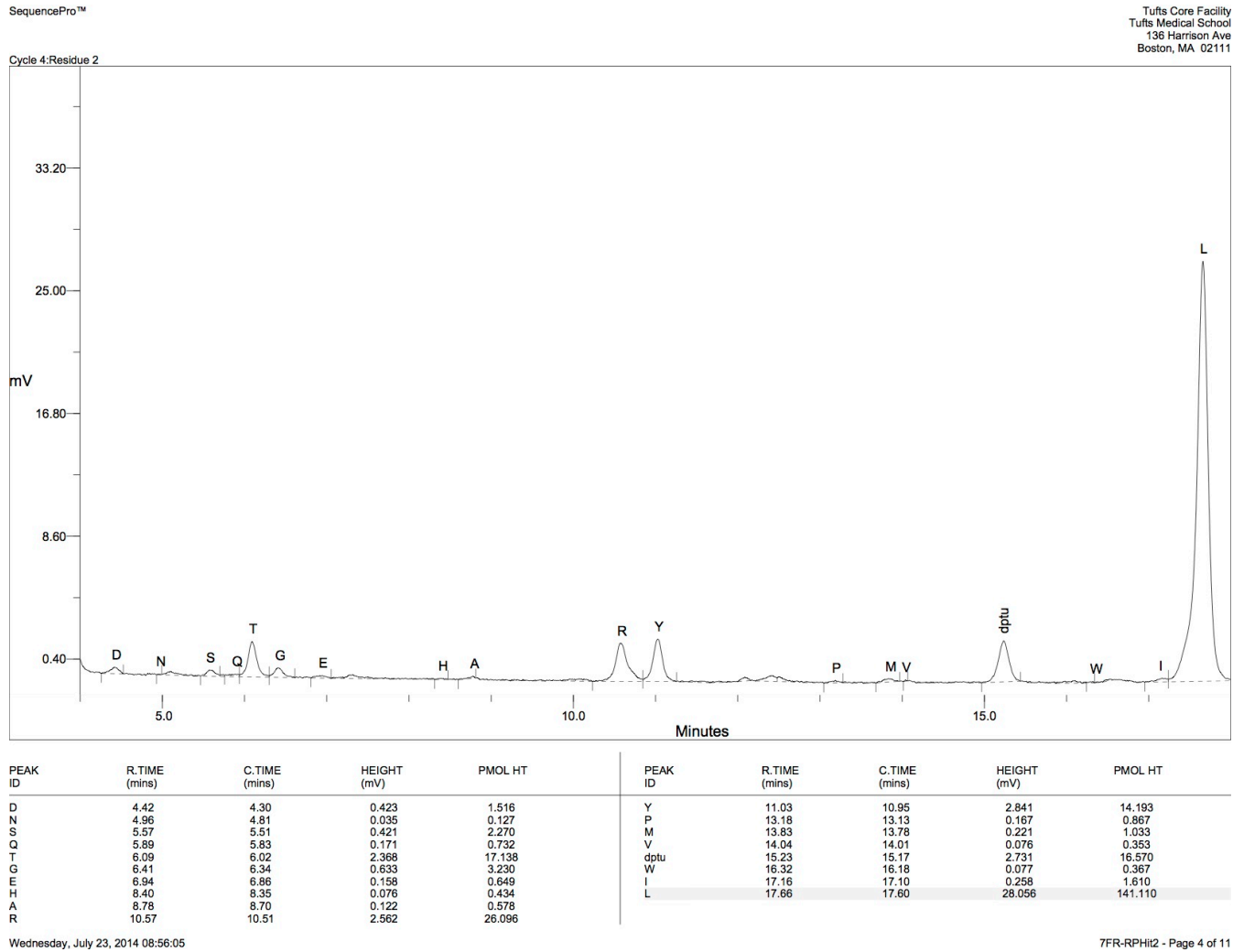


Figure 5.10: SL11 AA 7. The sequence of SL11 is R-L-D\*-Y-L-T-D\*-R-B-B-R-M. L is a random position in library.

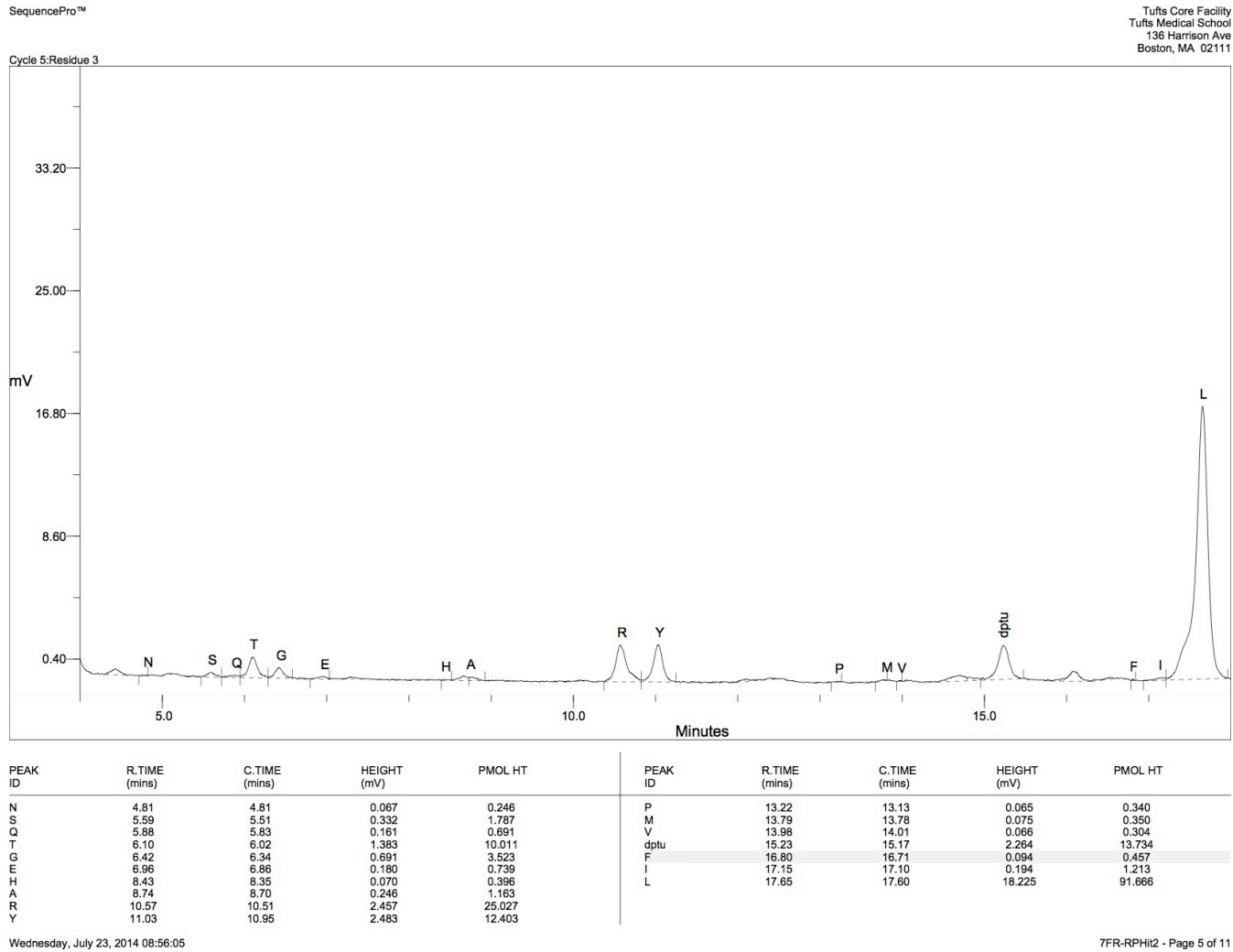
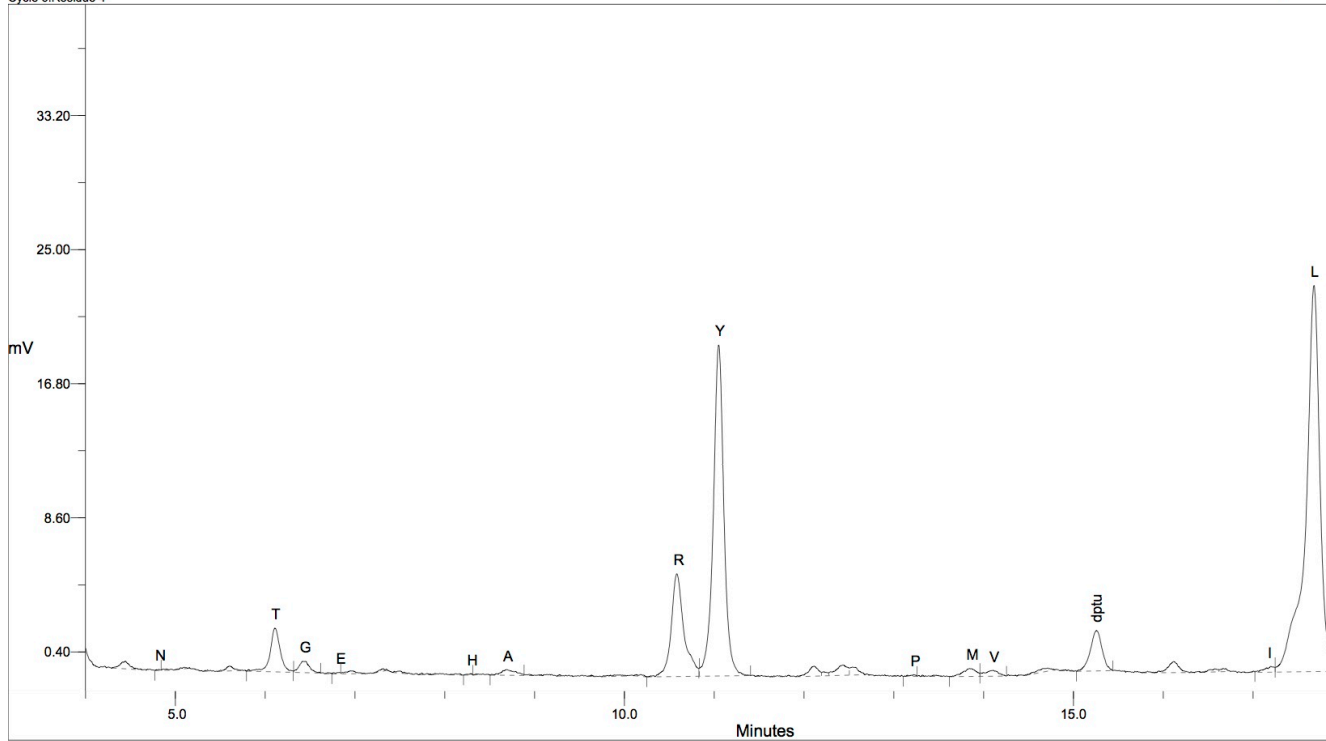


Figure 5.11: SL11 AA 6. The sequence of SL11 is R-L-D\*-Y-L-T-D\*-R-B-B-R-M. D\* is a fixed position in library.

Cycle 6:Residue 4



PEAK ID	R.TIME (mins)	C.TIME (mins)	HEIGHT (mV)	PMOL HT	PEAK ID	R.TIME (mins)	C.TIME (mins)	HEIGHT (mV)	PMOL HT
N	4.81	4.81	0.061	0.224	V	14.10	14.01	0.359	1.658
T	6.11	6.02	2.676	19.366	dptu	15.26	15.17	2.461	14.934
G	6.42	6.34	0.690	3.522	I	17.20	17.10	0.355	2.214
E	6.83	6.86	0.090	0.370	L	17.68	17.60	23.622	118.812
H	8.29	8.35	0.094	0.533					
A	8.69	8.70	0.346	1.636					
R	10.58	10.51	6.287	64.038					
Y	11.05	10.95	20.253	101.185					
P	13.23	13.13	0.099	0.516					
M	13.85	13.78	0.481	2.241					

Wednesday, July 23, 2014 08:56:05

7FR-RPHit2 - Page 6 of 11

Figure 5.12: SL11 AA 5. The sequence of SL11 is R-L-D\*-Y-L-T-D\*-R-B-B-R-M. Y is a random position in library.

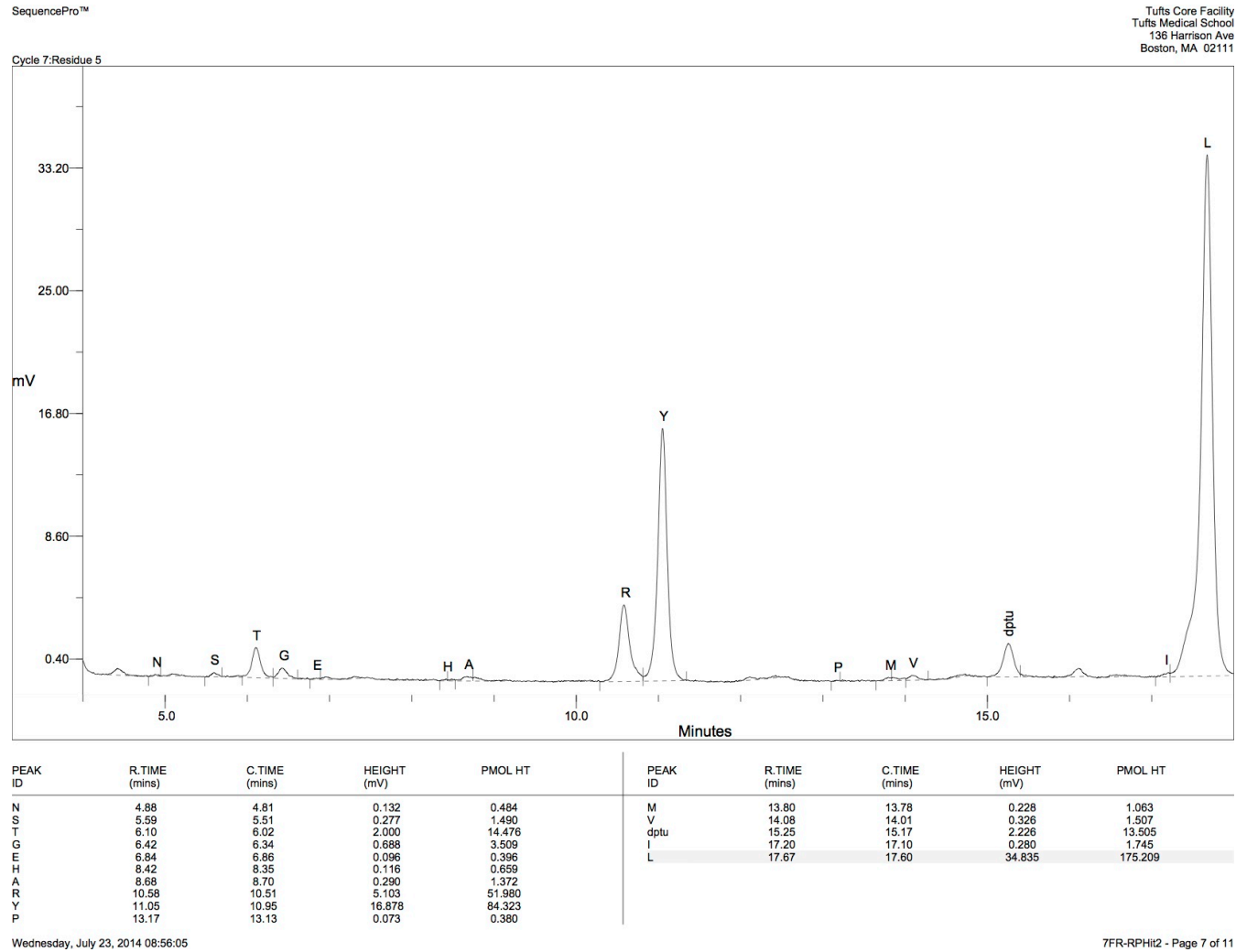


Figure 5.13: SL11 AA 4. The sequence of SL11 is R-L-D\*-Y-L-T-D\*-R-B-B-R-M. L is a random position in library.

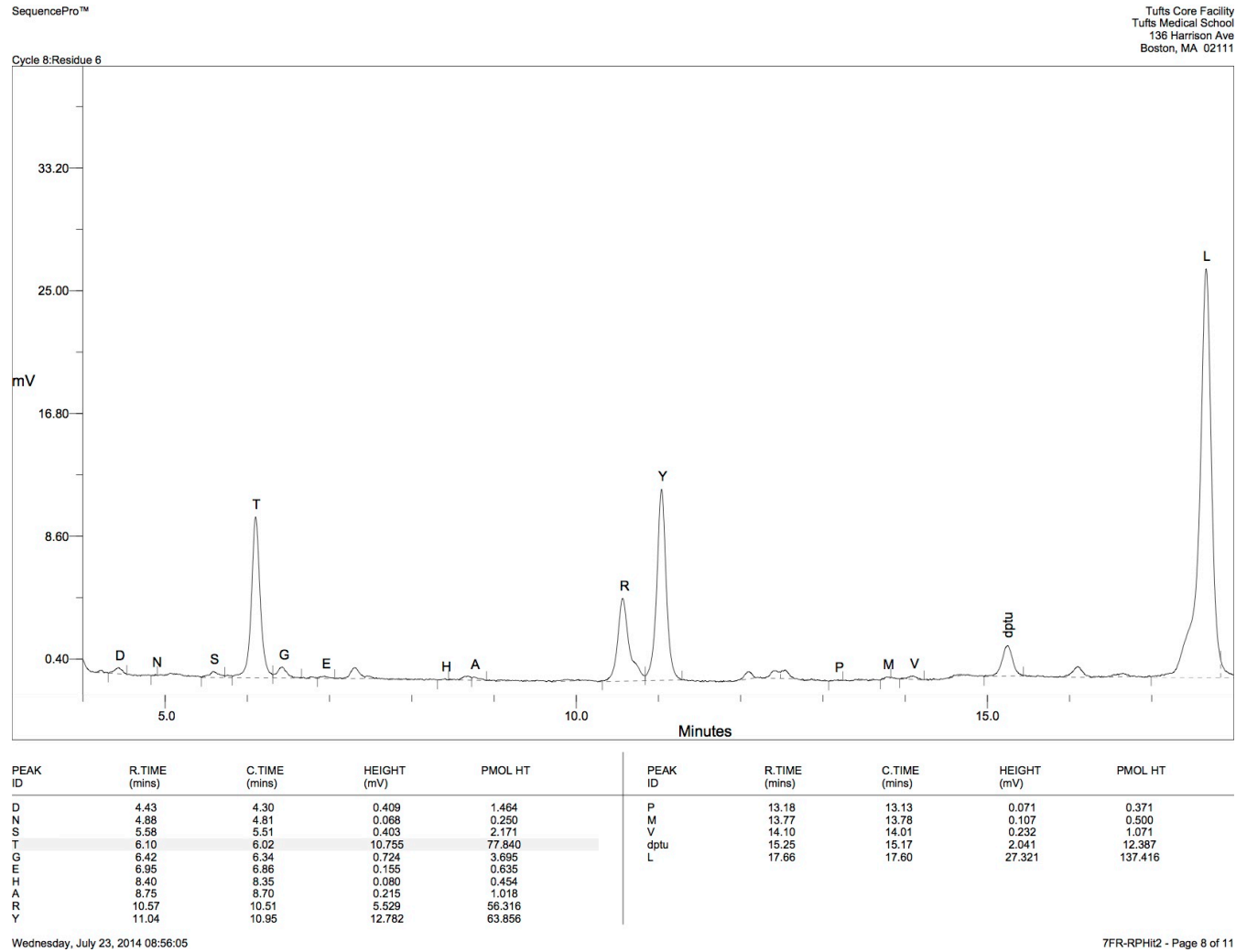


Figure 5.14: SL11 AA 3. The sequence of SL11 is R-L-D\*-Y-L-T-D\*-R-B-B-R-M. T is a random position in library.

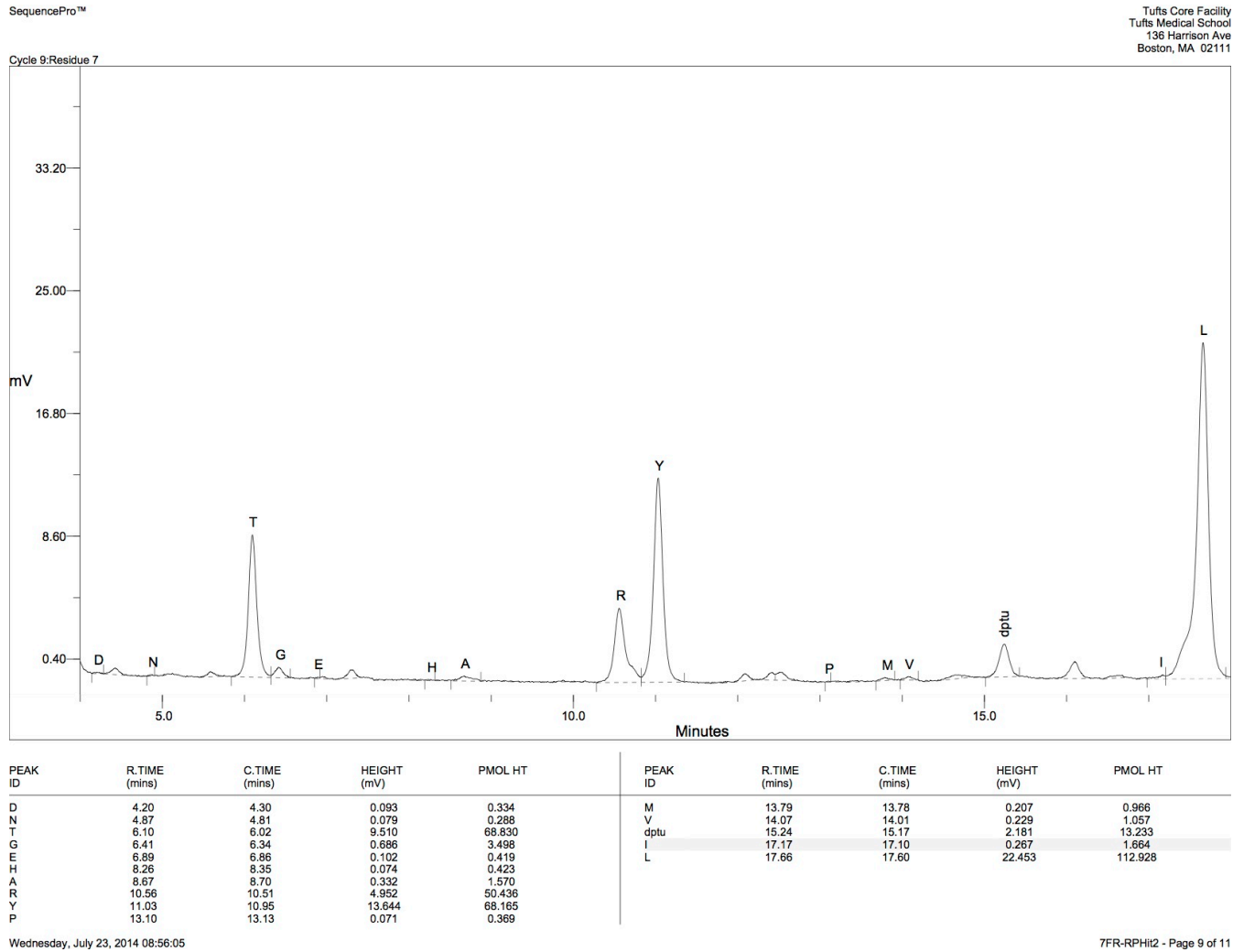


Figure 5.15: SL11 AA 2. The sequence of SL11 is R-L-D\*-Y-L-T-D\*-R-B-B-R-M. D\* is a random position in library.

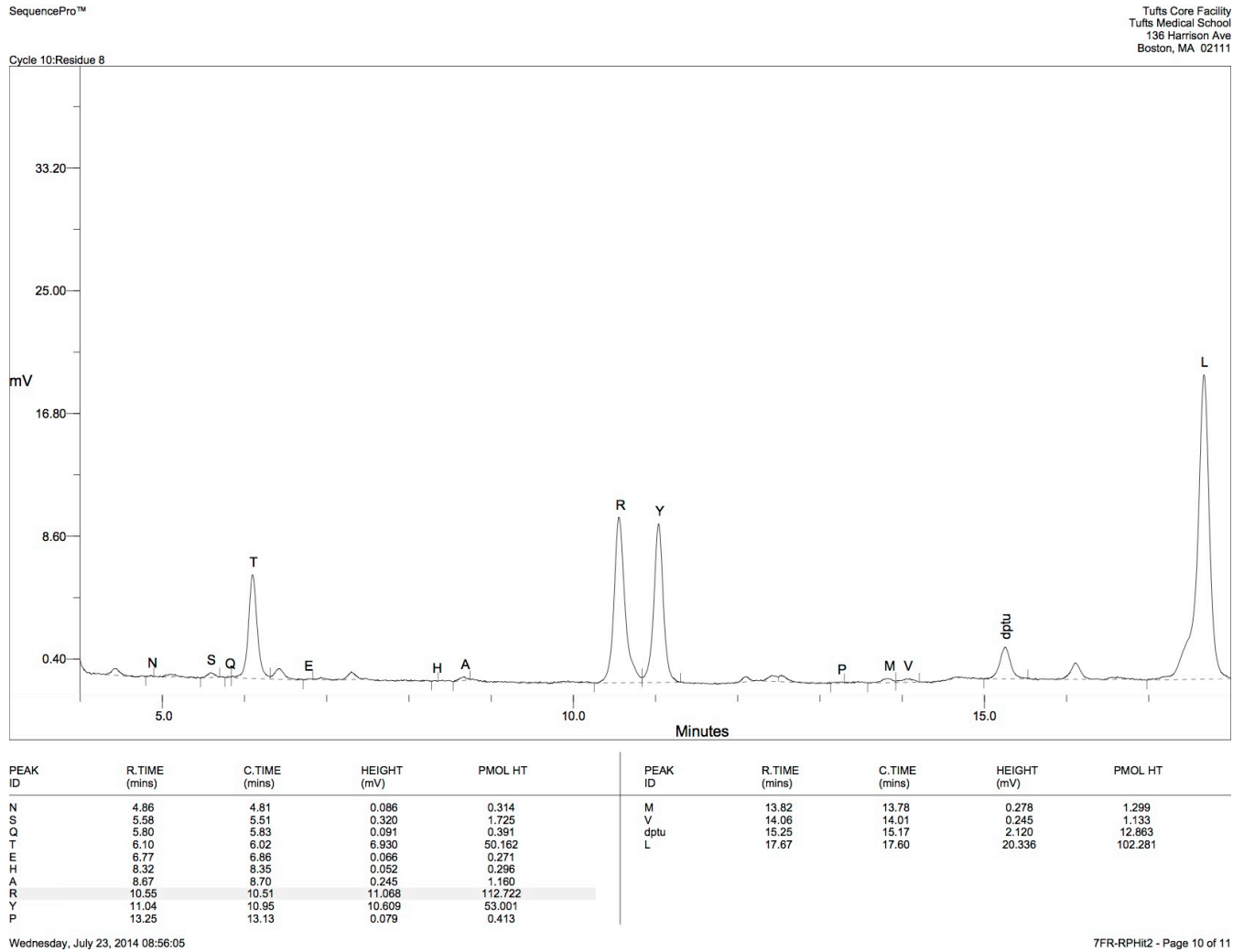


Figure 5.16: SL11 AA 1. The sequence of SL11 is R-L-D\*-Y-L-T-D\*-R-B-B-R-M. R is a random position in library.

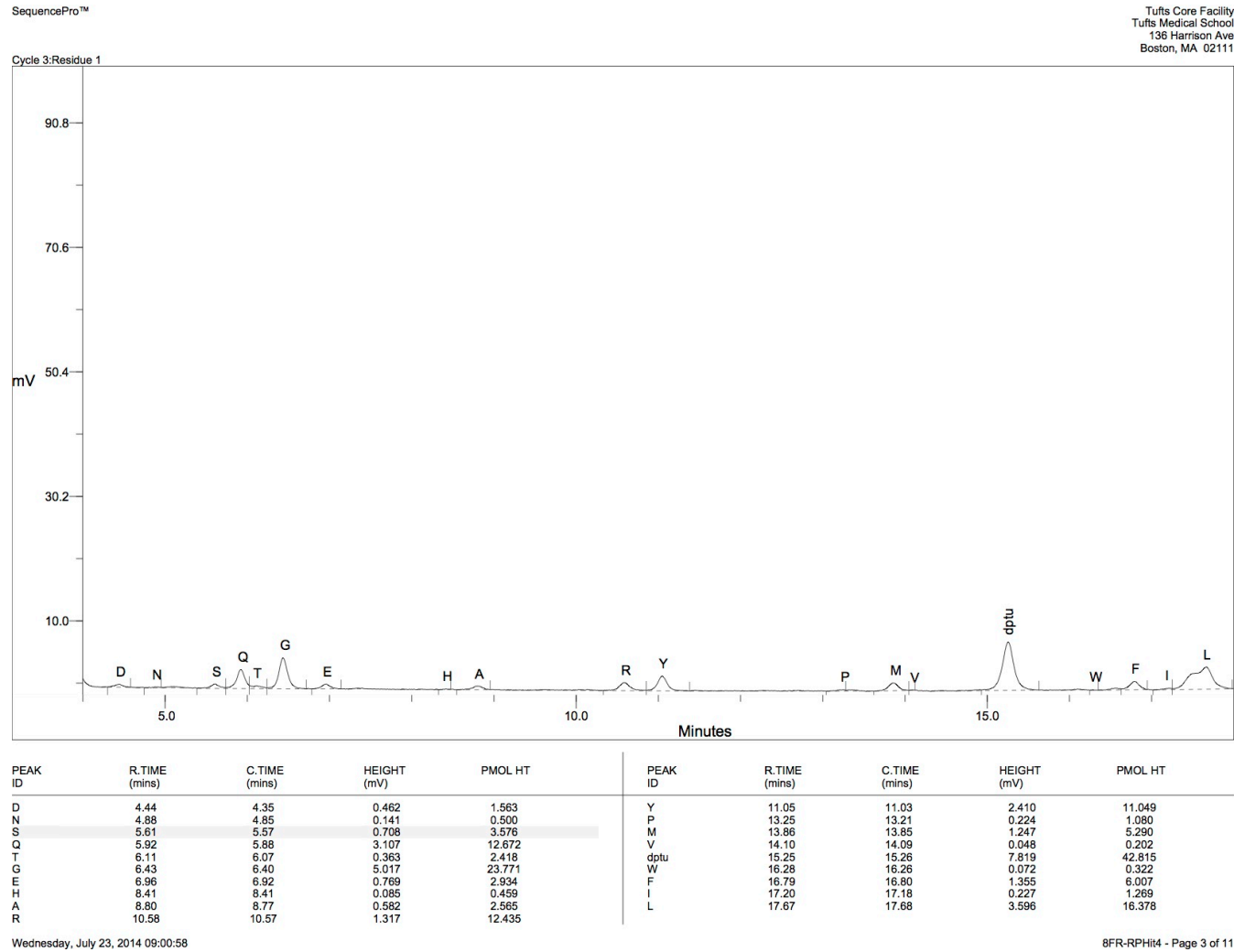


Figure 5.17: SL12 AA 8. The sequence of SL12 is R-L-D\*-G-F-Y-D\*-Q-B-B-R-M. R is a fixed position in library.



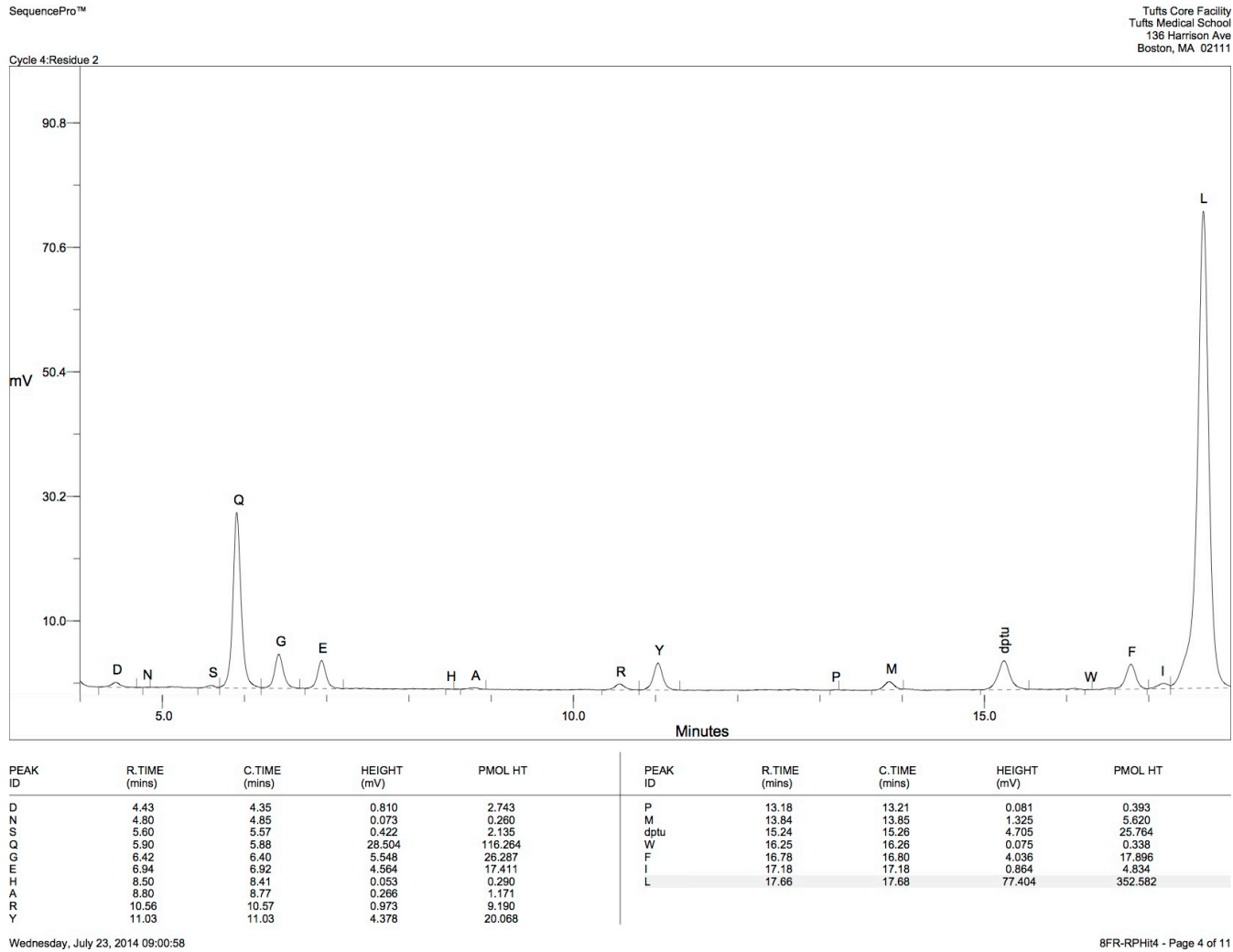


Figure 5.18: SL12 AA 7. The sequence of SL12 is R-L-D\*-G-F-Y-D\*-Q-B-B-R-M. L is a random position in library.

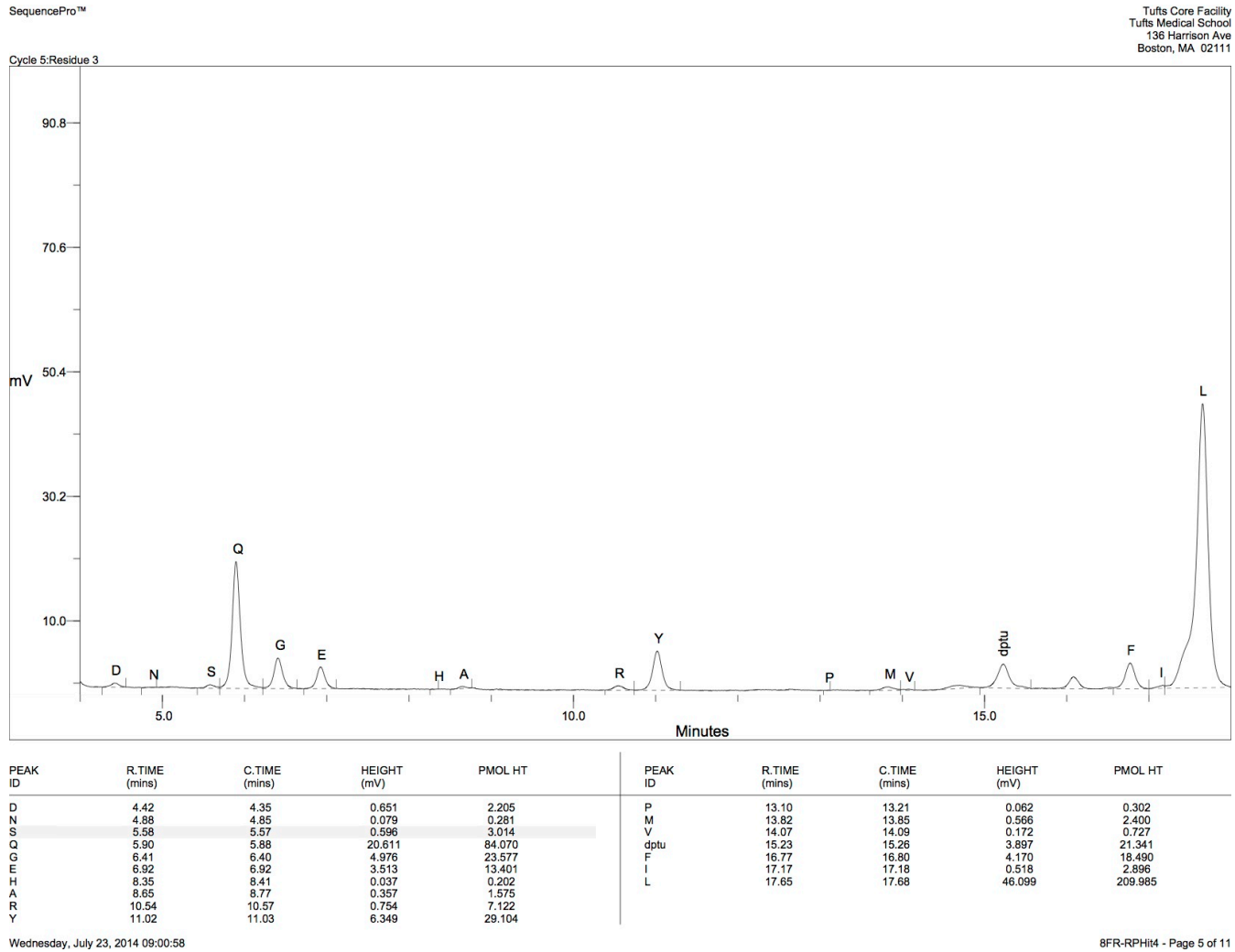


Figure 5.19: SL12 AA 6. The sequence of SL12 is R-L-D\*-G-F-Y-D\*-Q-B-B-R-M. D\* is a fixed position in library.

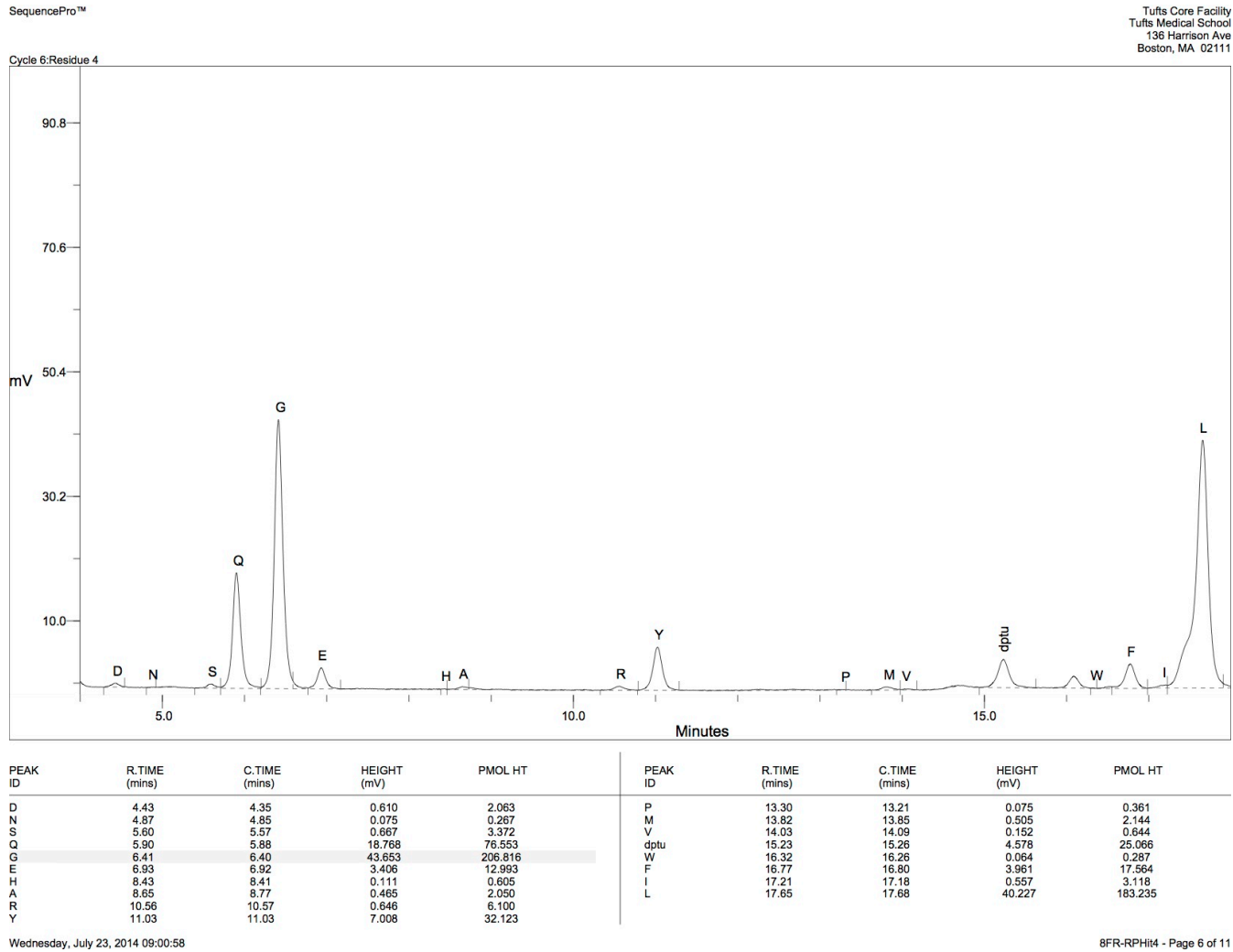


Figure 5.20: SL12 AA 5. The sequence of SL12 is R-L-D\*-G-F-Y-D\*-Q-B-B-R-M. G is a random position in library.

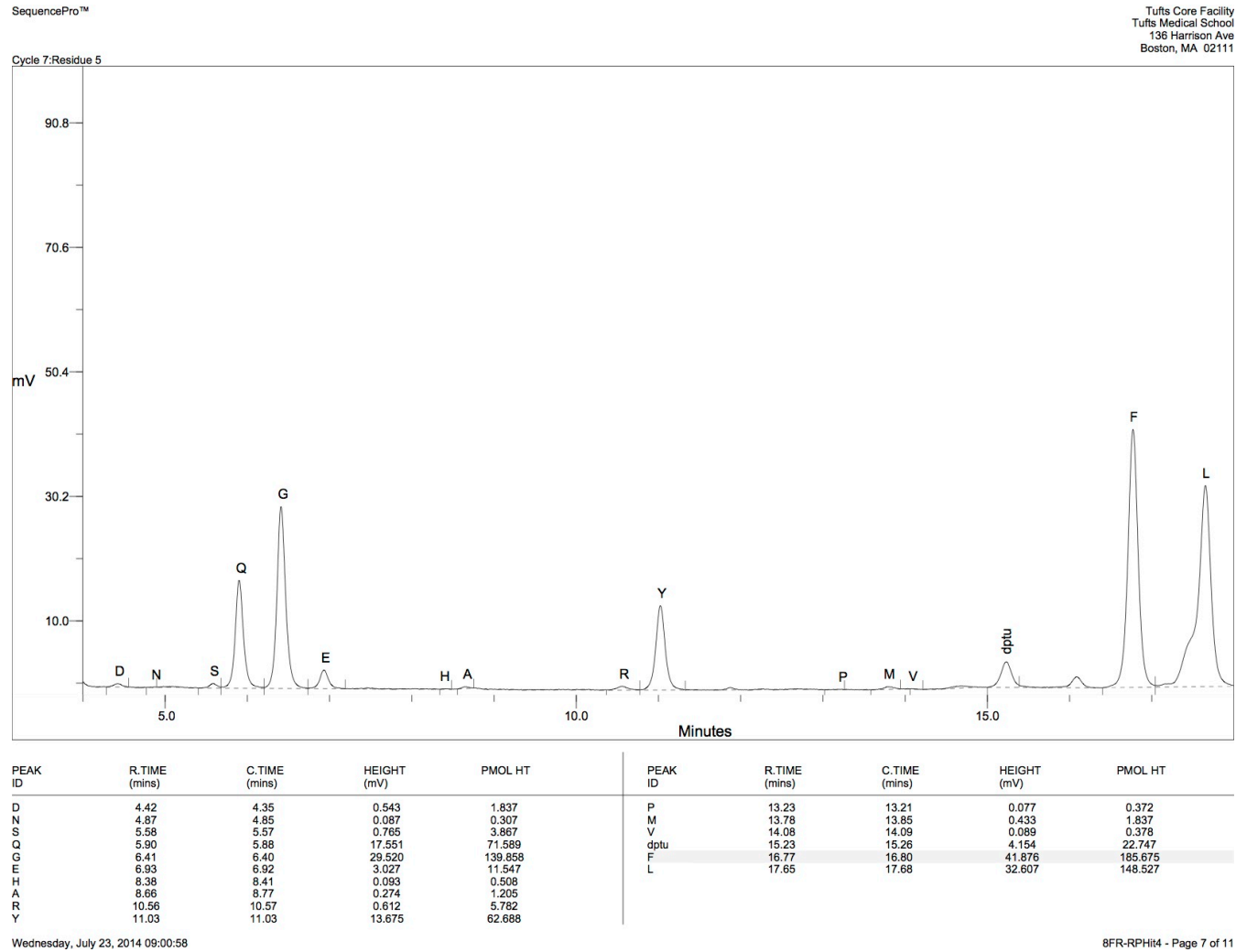


Figure 5.21: SL12 AA 4. The sequence of SL12 is R-L-D\*-G-F-Y-D\*-Q-B-B-R-M. F is a random position in library.

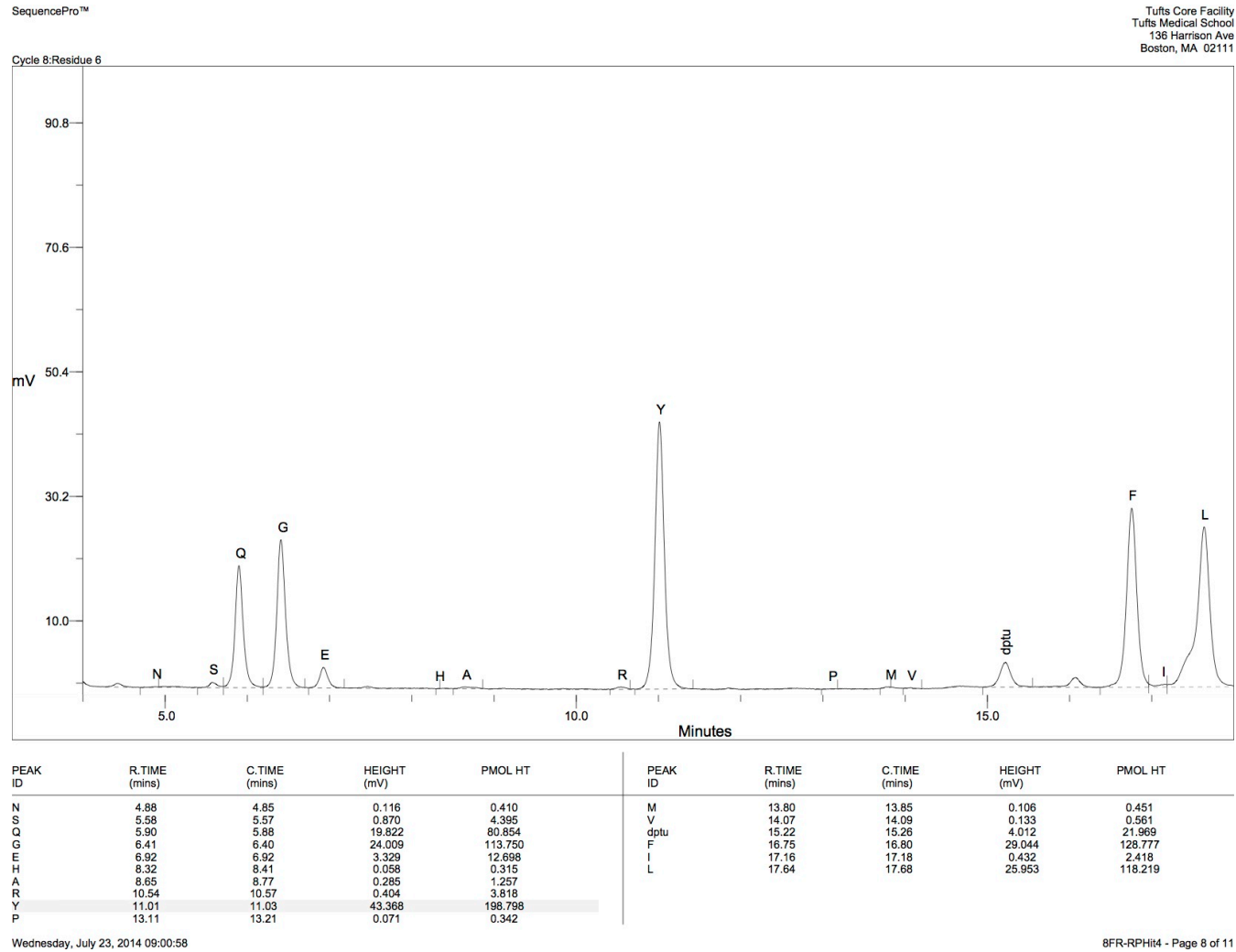


Figure 5.22: SL12 AA 3. The sequence of SL12 is R-L-D\*-G-F-Y-D\*-Q-B-B-R-M. Y is a random position in library.

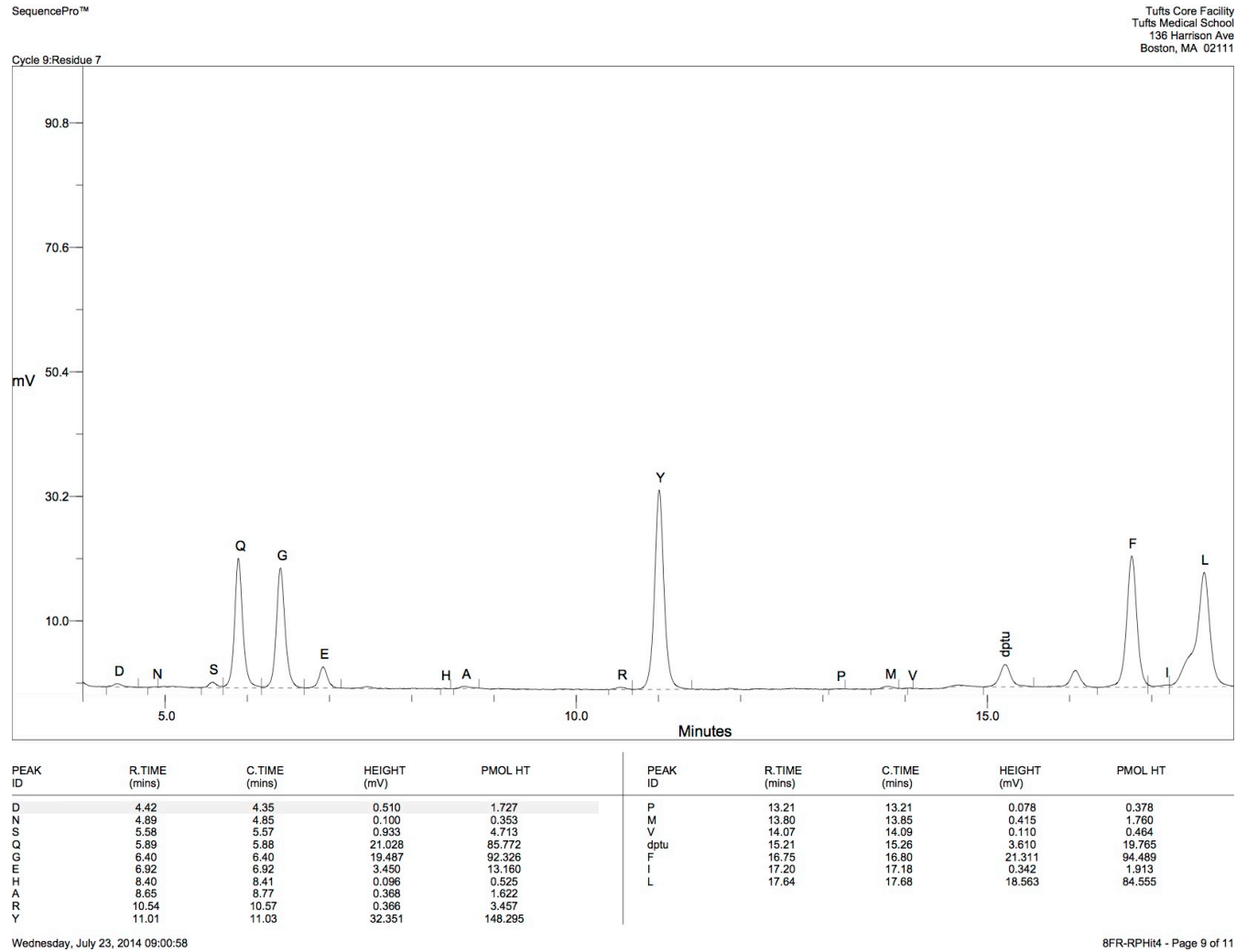


Figure 5.23: SL12 AA 2. The sequence of SL12 is R-L-D\*-G-F-Y-D\*-Q-B-B-R-M. D\* is a fixed position in library.

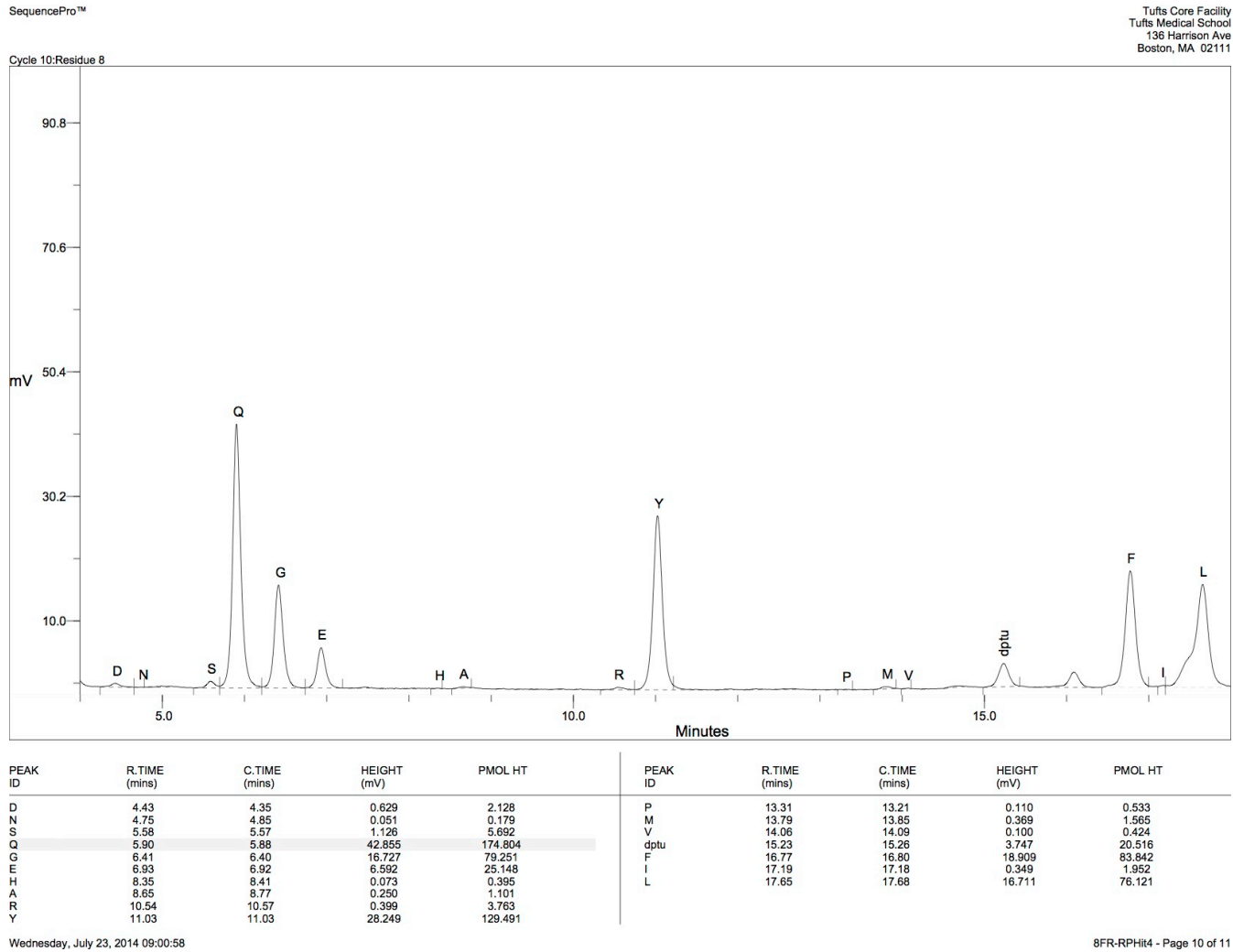


Figure 5.24: SL12 AA 1. The sequence of SL12 is R-L-D\*-G-F-Y-D\*-Q-B-B-R-M. Q is a random position in library.

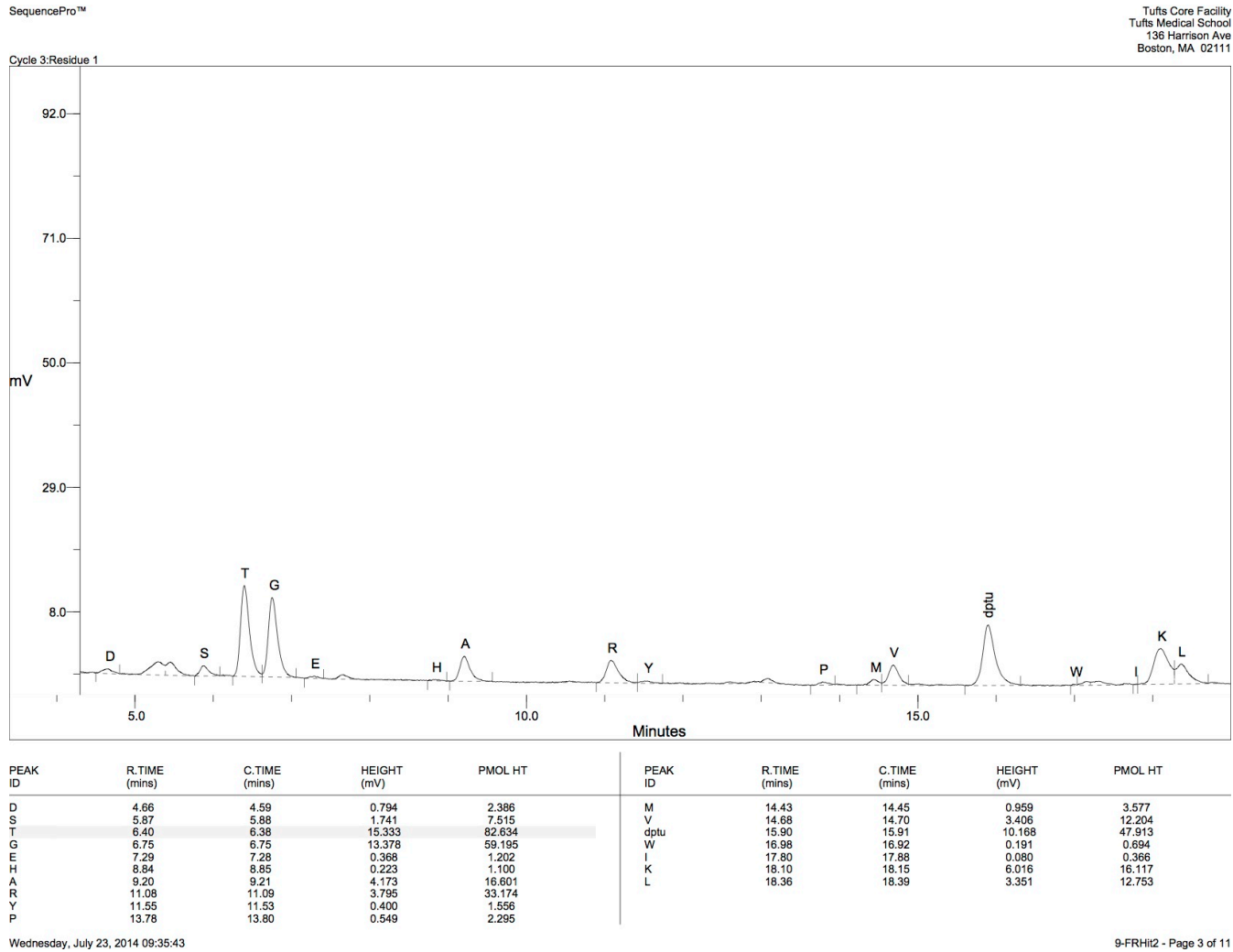


Figure 5.25: SL13 AA 8. The sequence of SL13 is R-T-D\*-G-L-A-D\*-V-B-B-R-M. R is a fixed position in library.



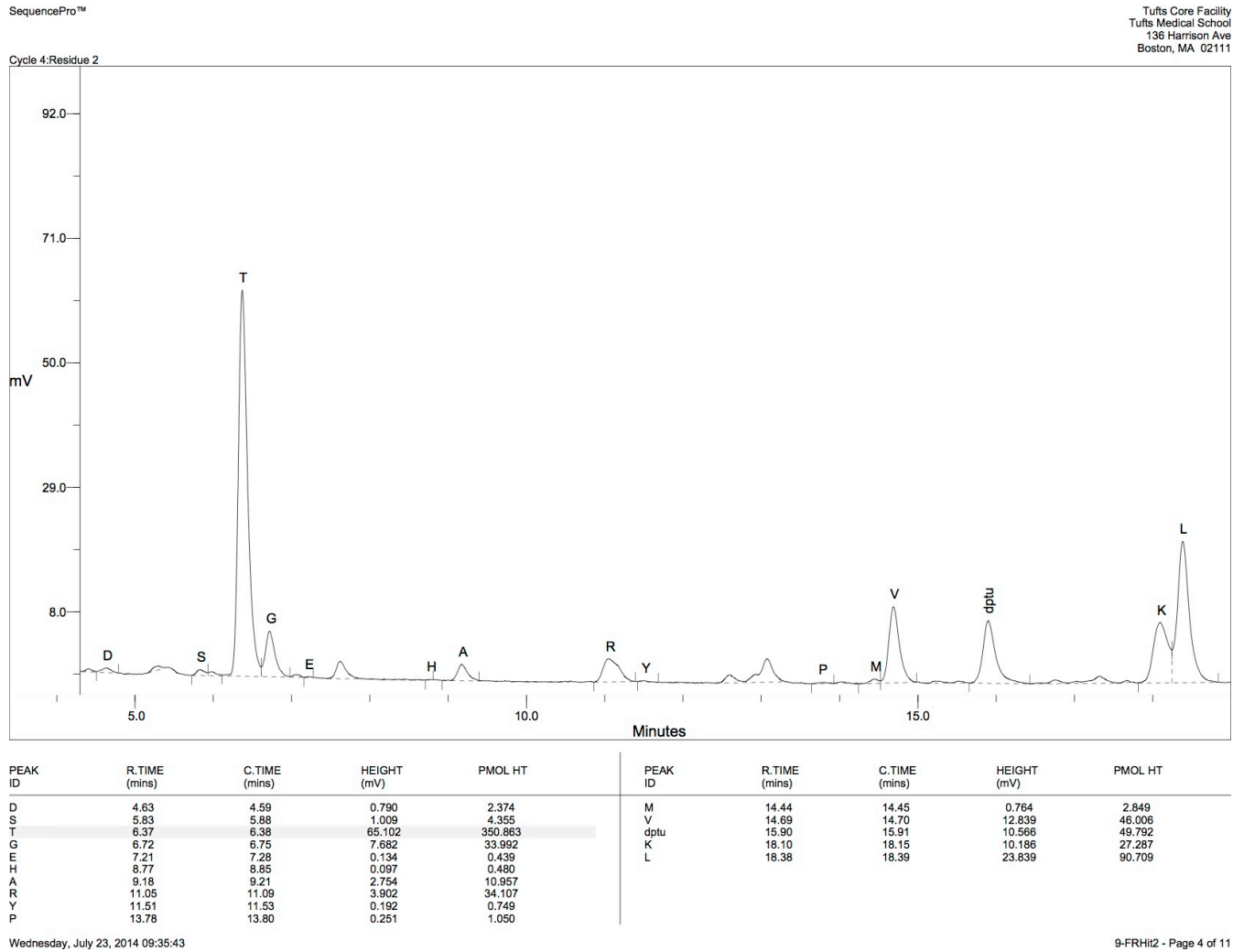


Figure 5.26: SL13 AA 7. The sequence of SL13 is R-T-D\*-G-L-A-D\*-V-B-B-R-M. T is a random position in library.

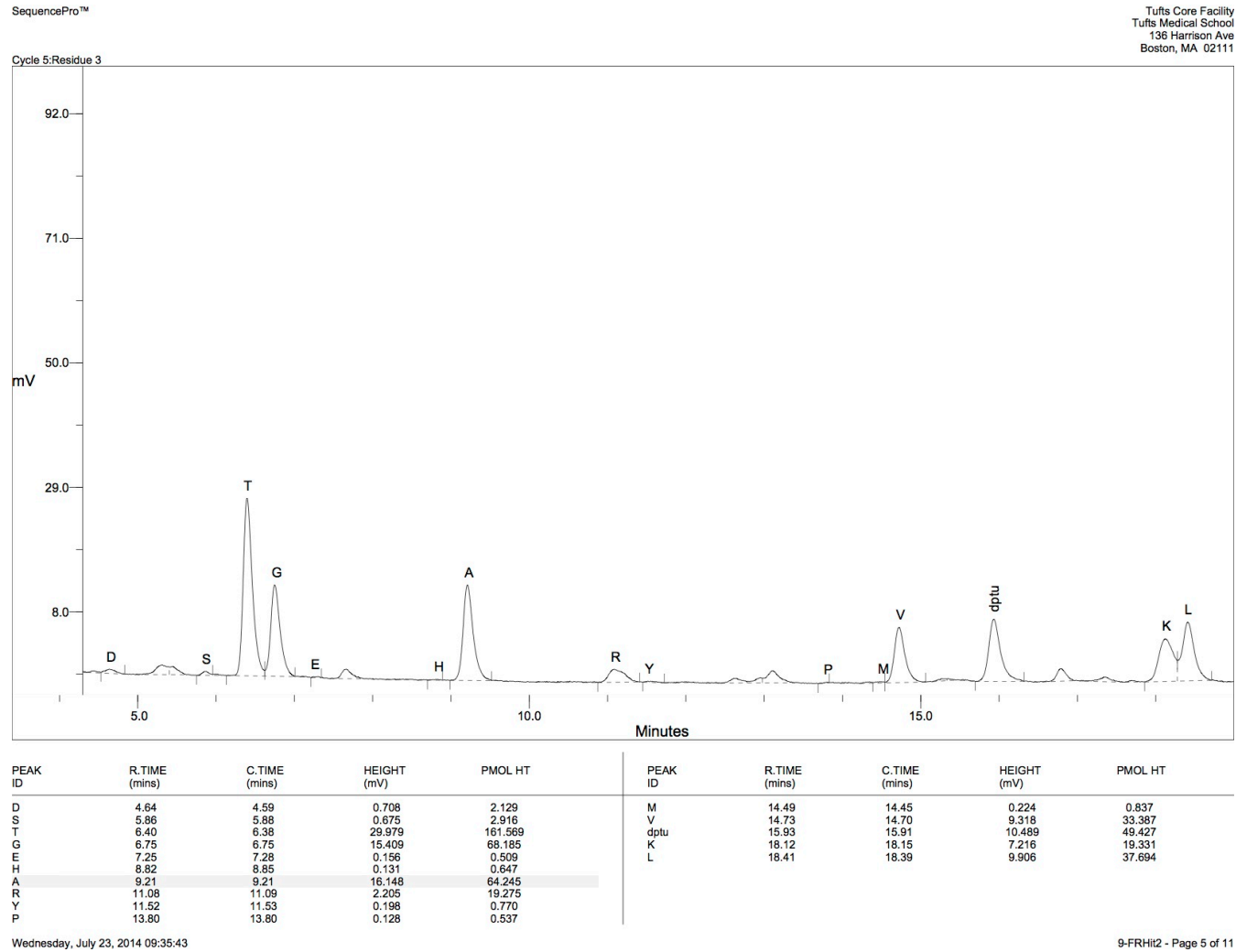


Figure 5.27: SL13 AA 6. The sequence of SL13 is R-T-D\*-G-L-A-D\*-V-B-B-R-M. D\* is a fixed position in library.

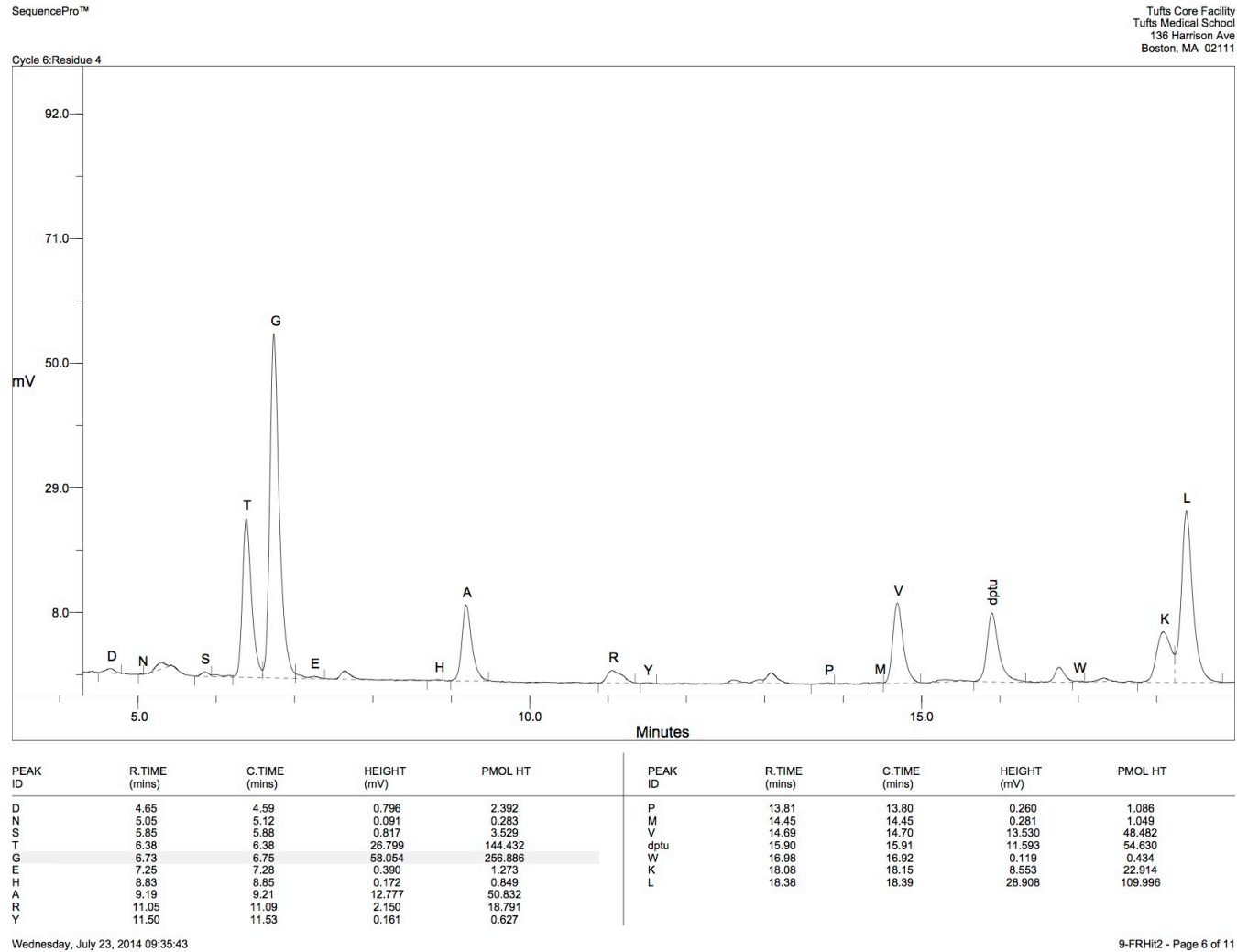


Figure 5.28: SL13 AA 5. The sequence of SL13 is R-T-D\*-G-L-A-D\*-V-B-B-R-M. G is a random position in library.

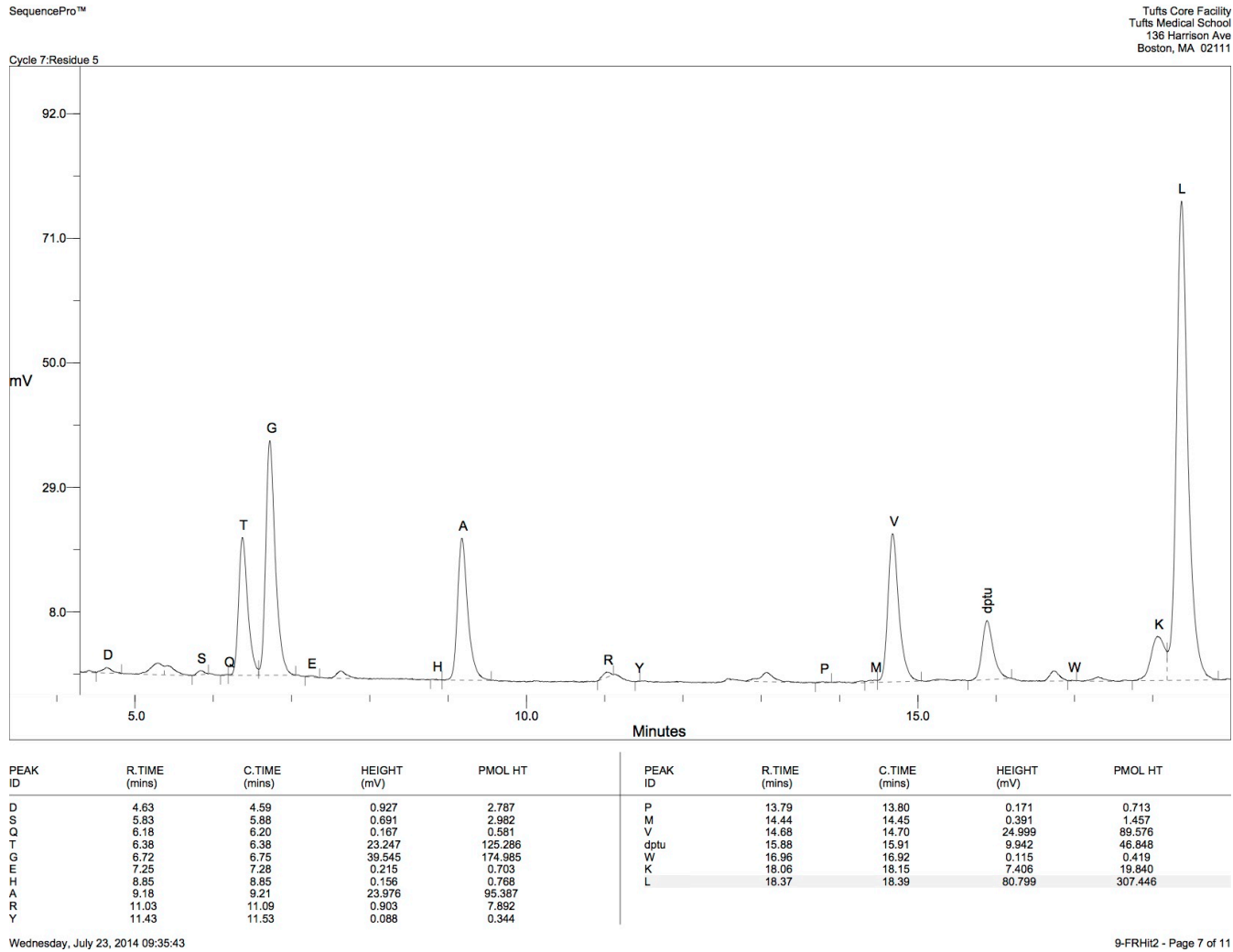


Figure 5.29: SL13 AA 4. The sequence of SL13 is R-T-D\*-G-L-A-D\*-V-B-B-R-M. L is a random position in library.

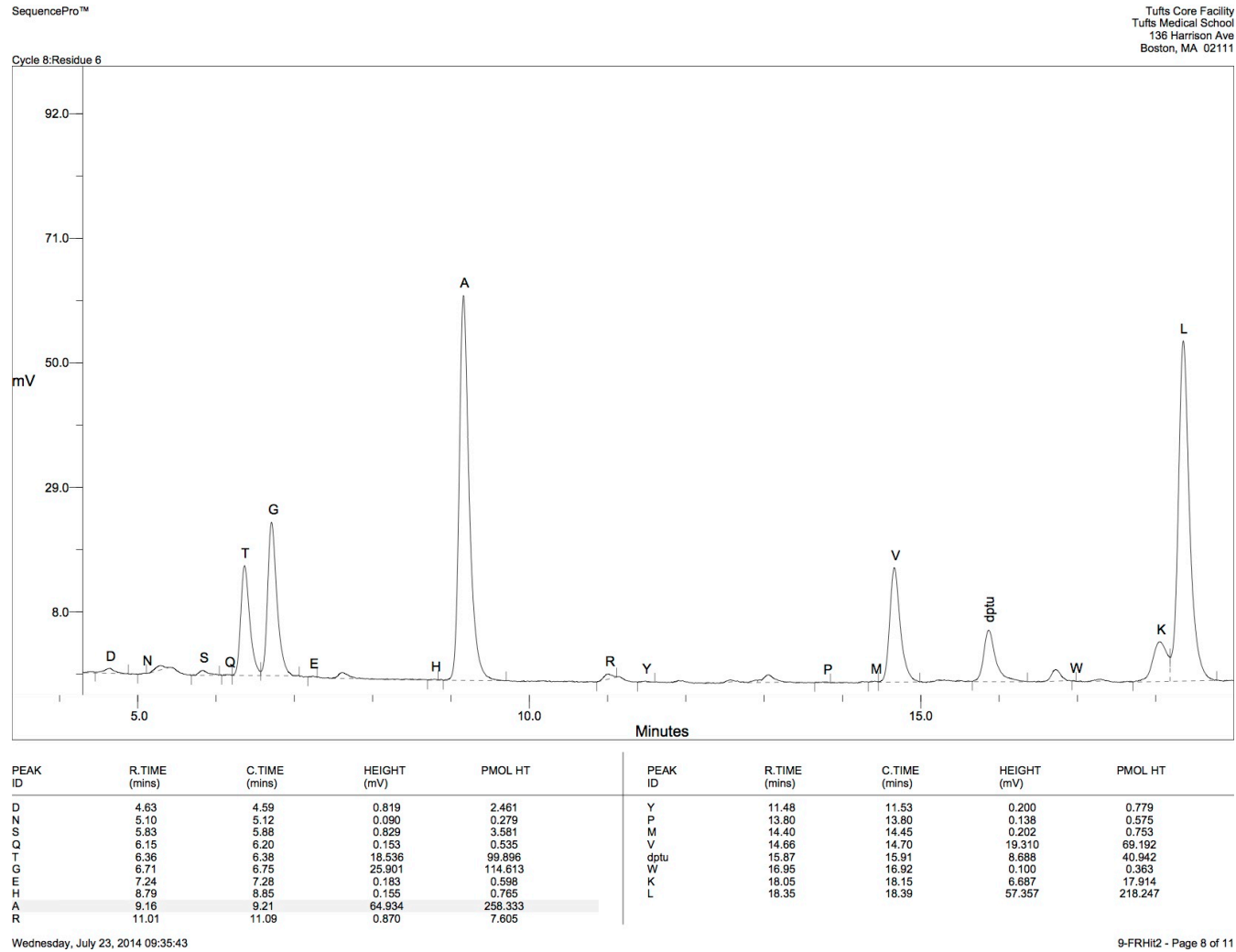


Figure 5.30: SL13 AA 3. The sequence of SL13 is R-T-D\*-G-L-A-D\*-V-B-B-R-M. A is a random position in library.

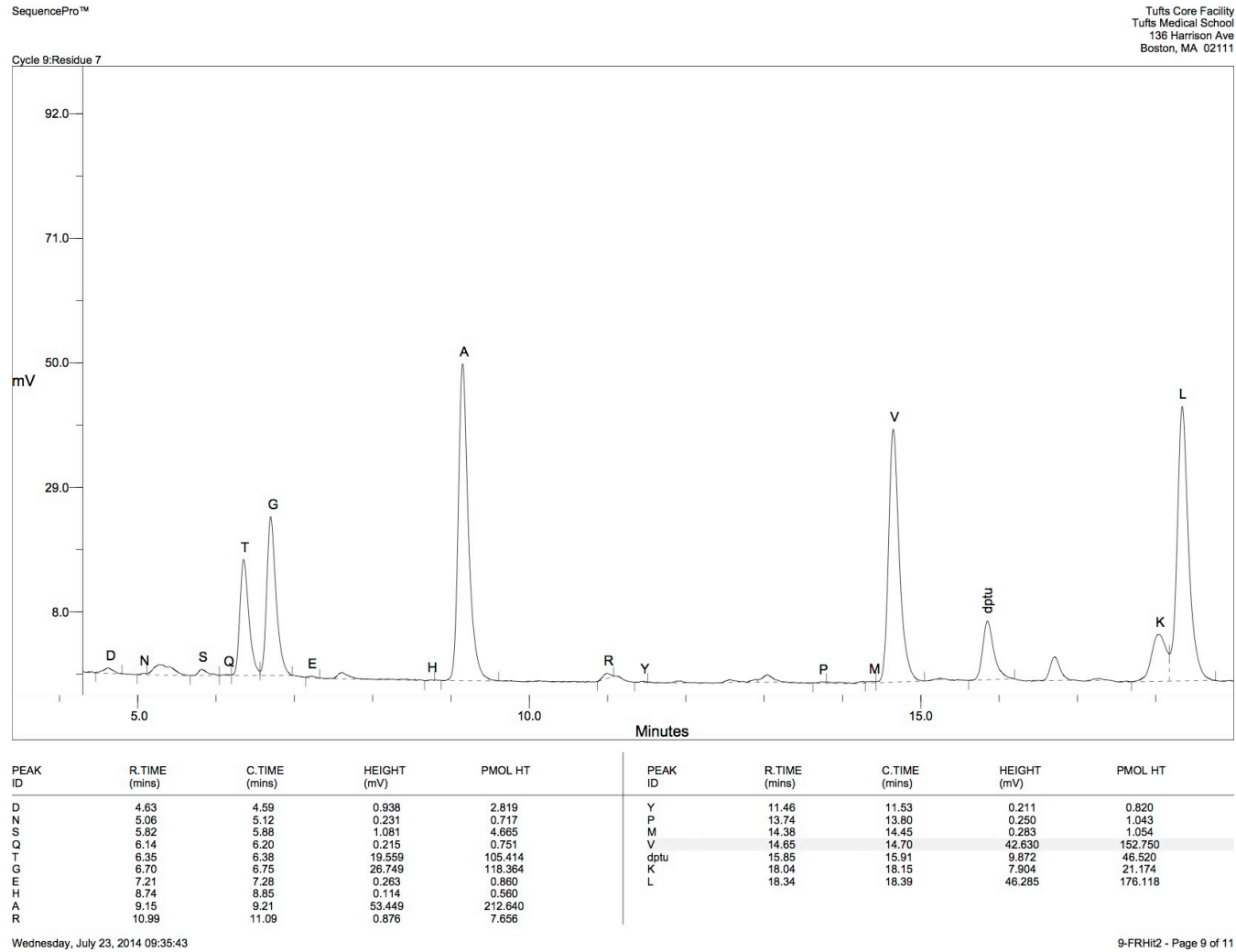


Figure 5.31: SL13 AA 2. The sequence of SL13 is R-T-D\*-G-L-A-D\*-V-B-B-R-M. D\* is a fixed position in library.

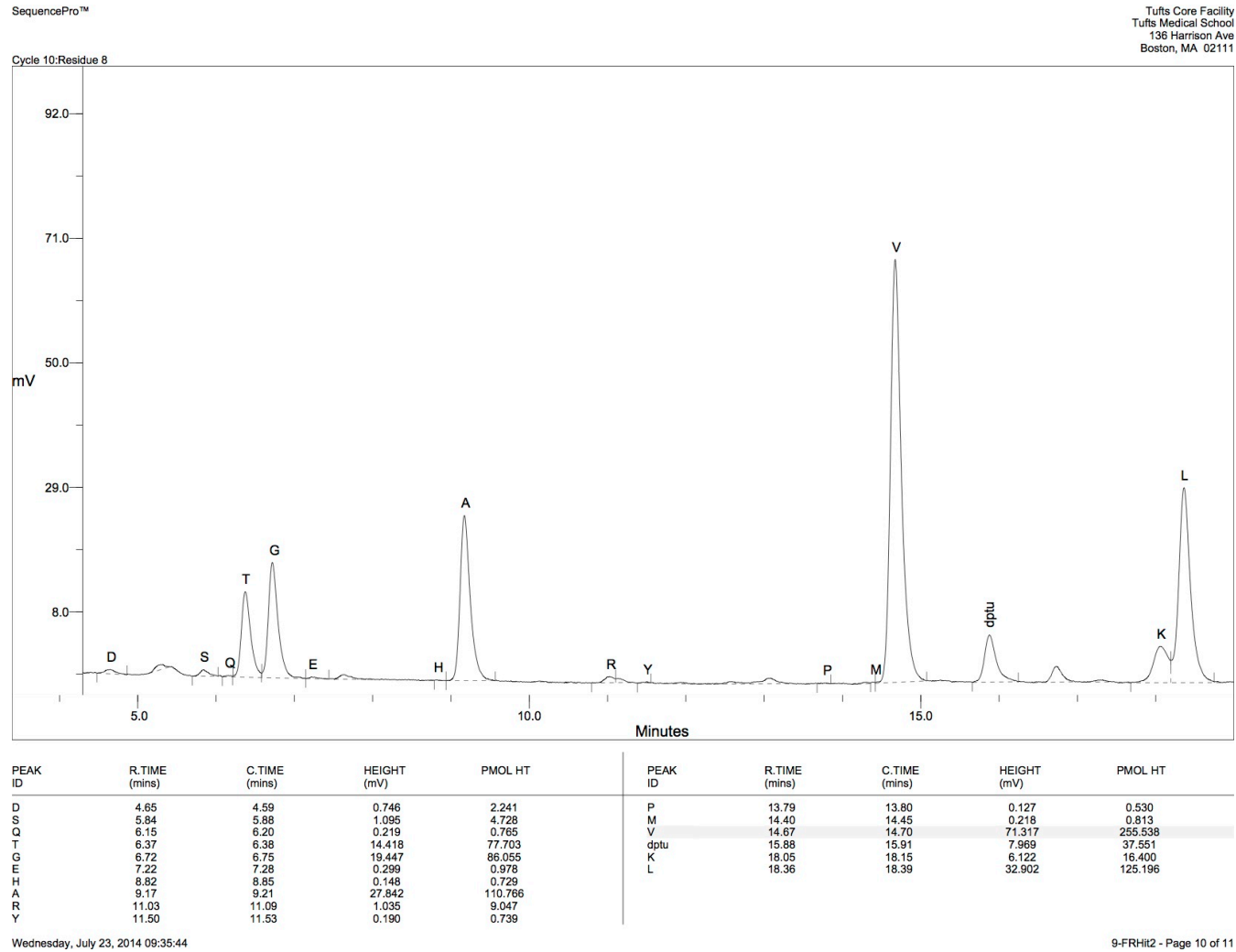


Figure 5.32: SL13 AA 1. The sequence of SL13 is R-T-D\*-G-L-A-D\*-V-B-B-R-M. V is a random position in library.

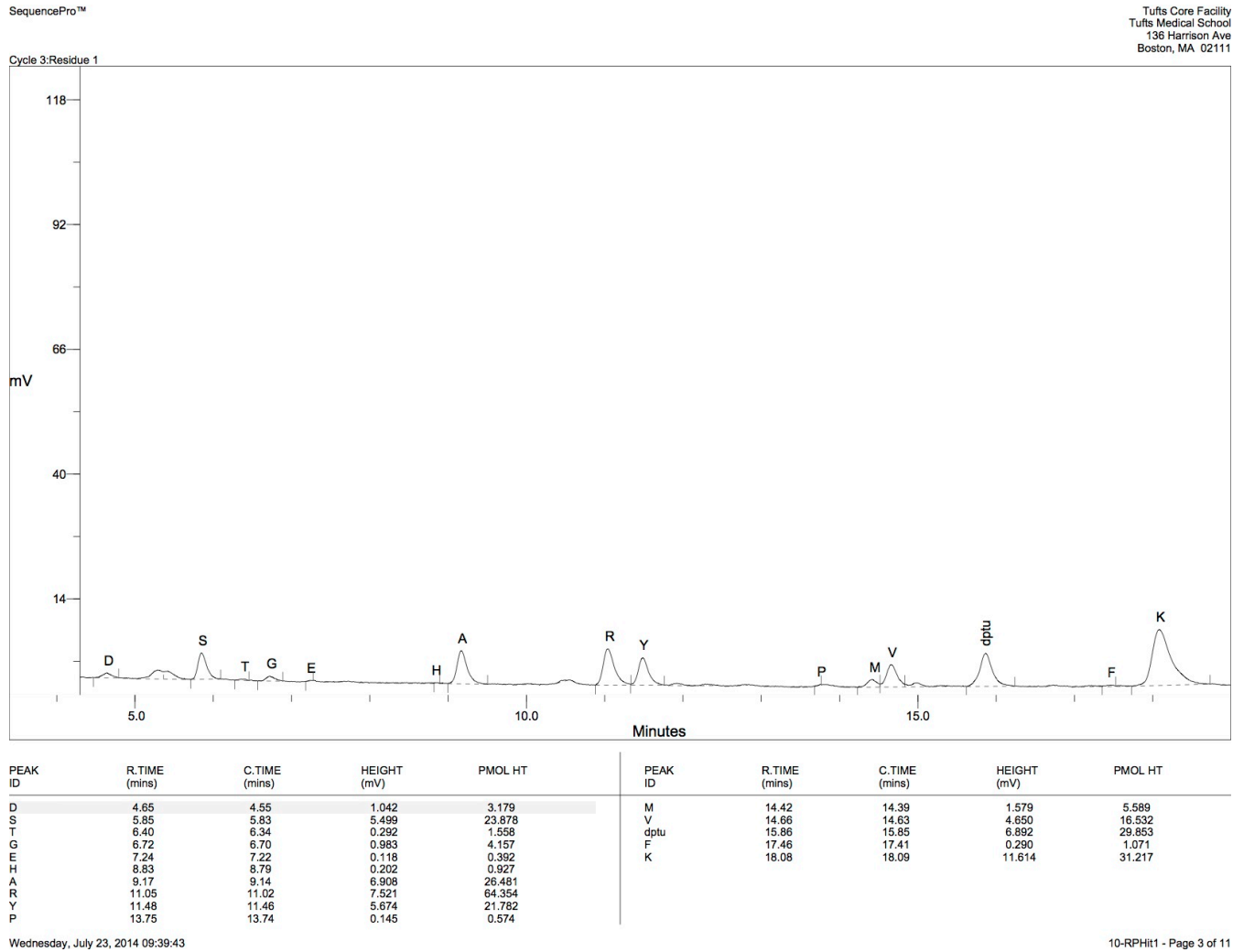


Figure 5.33: SL14 AA 8. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. R is a fixed position in library.



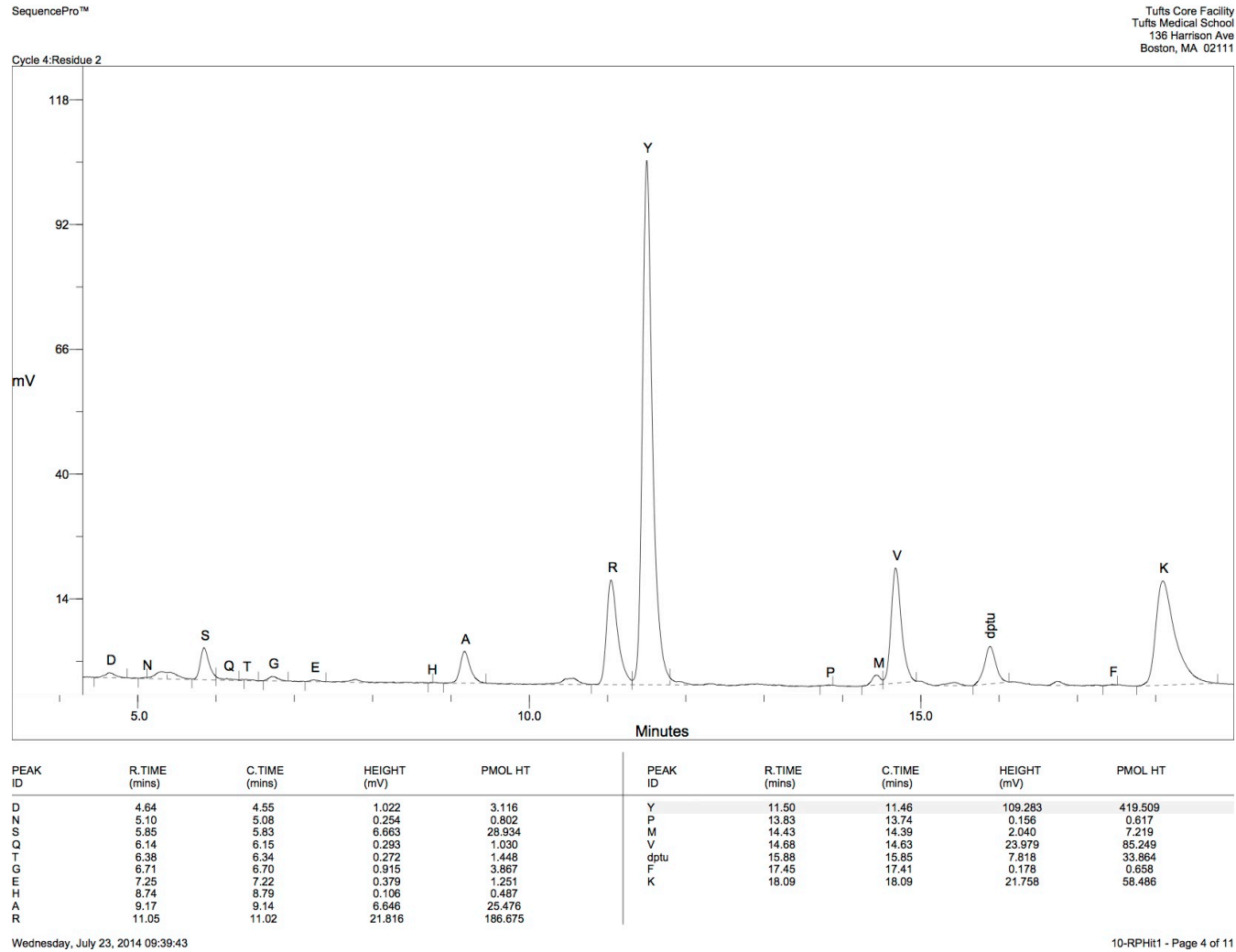


Figure 5.34: SL14 AA 7. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. Y is a random position in library.

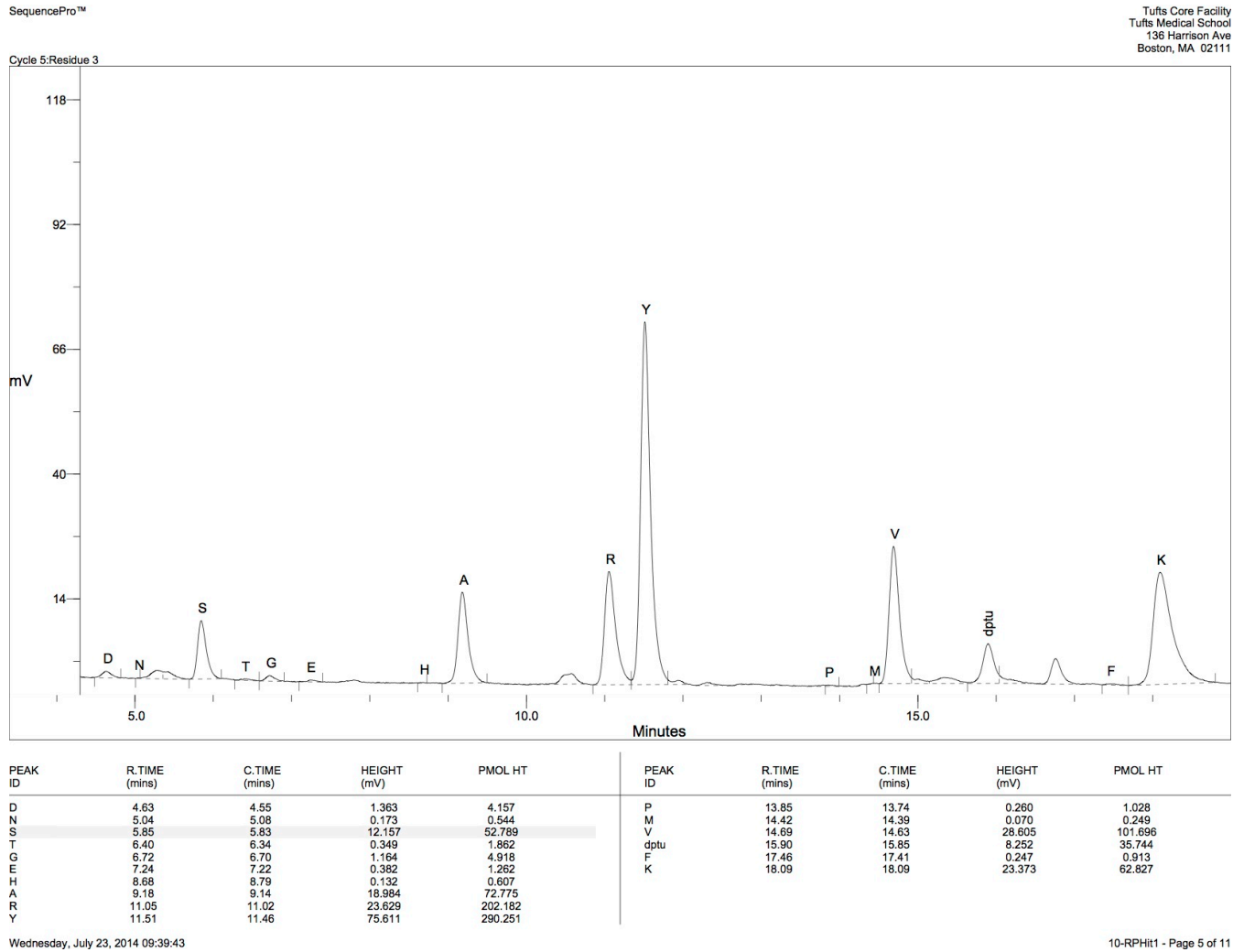


Figure 5.35: SL14 AA 6. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. D\* is a fixed position in library.

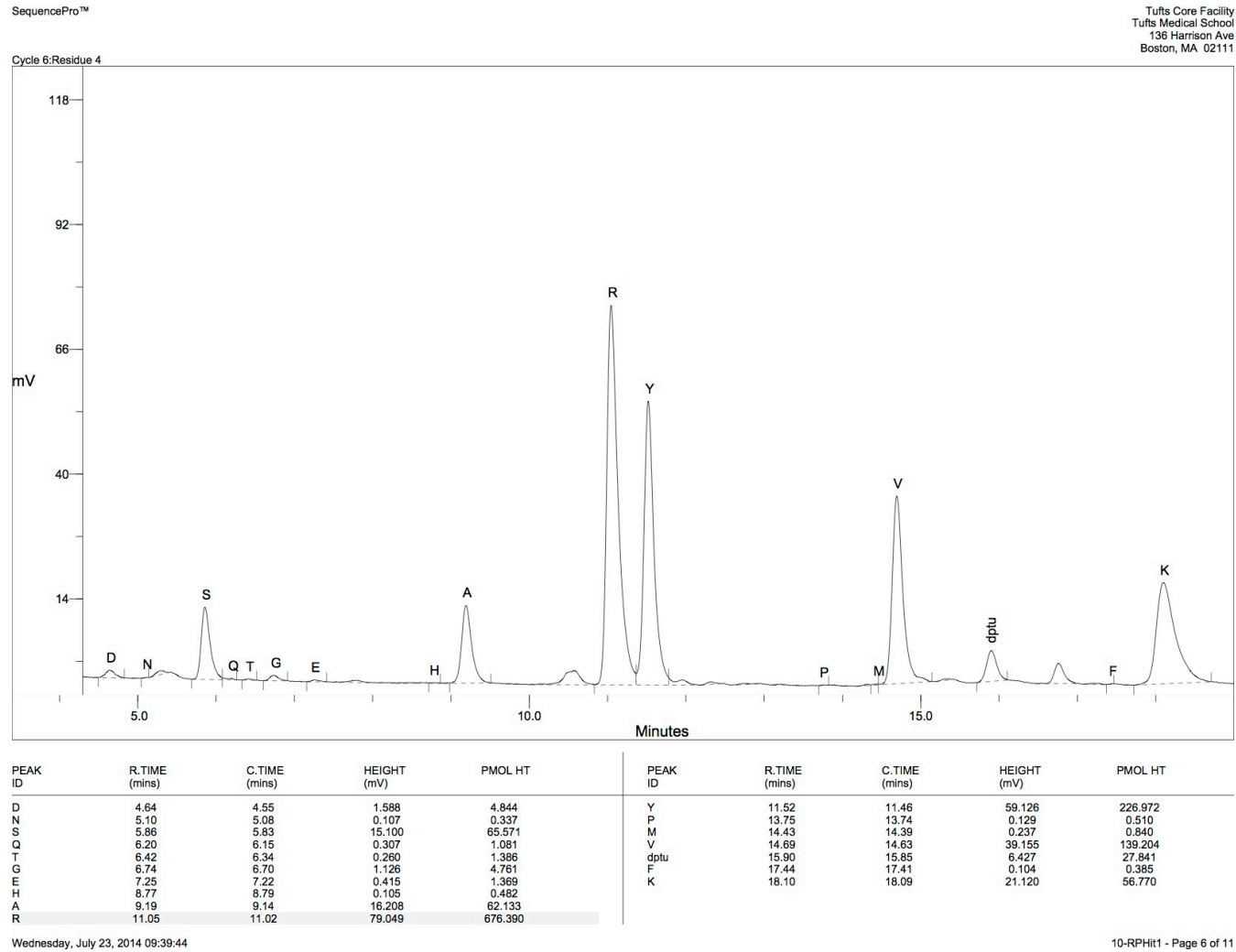


Figure 5.36: SL14 AA 5. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. R is a random position in library.

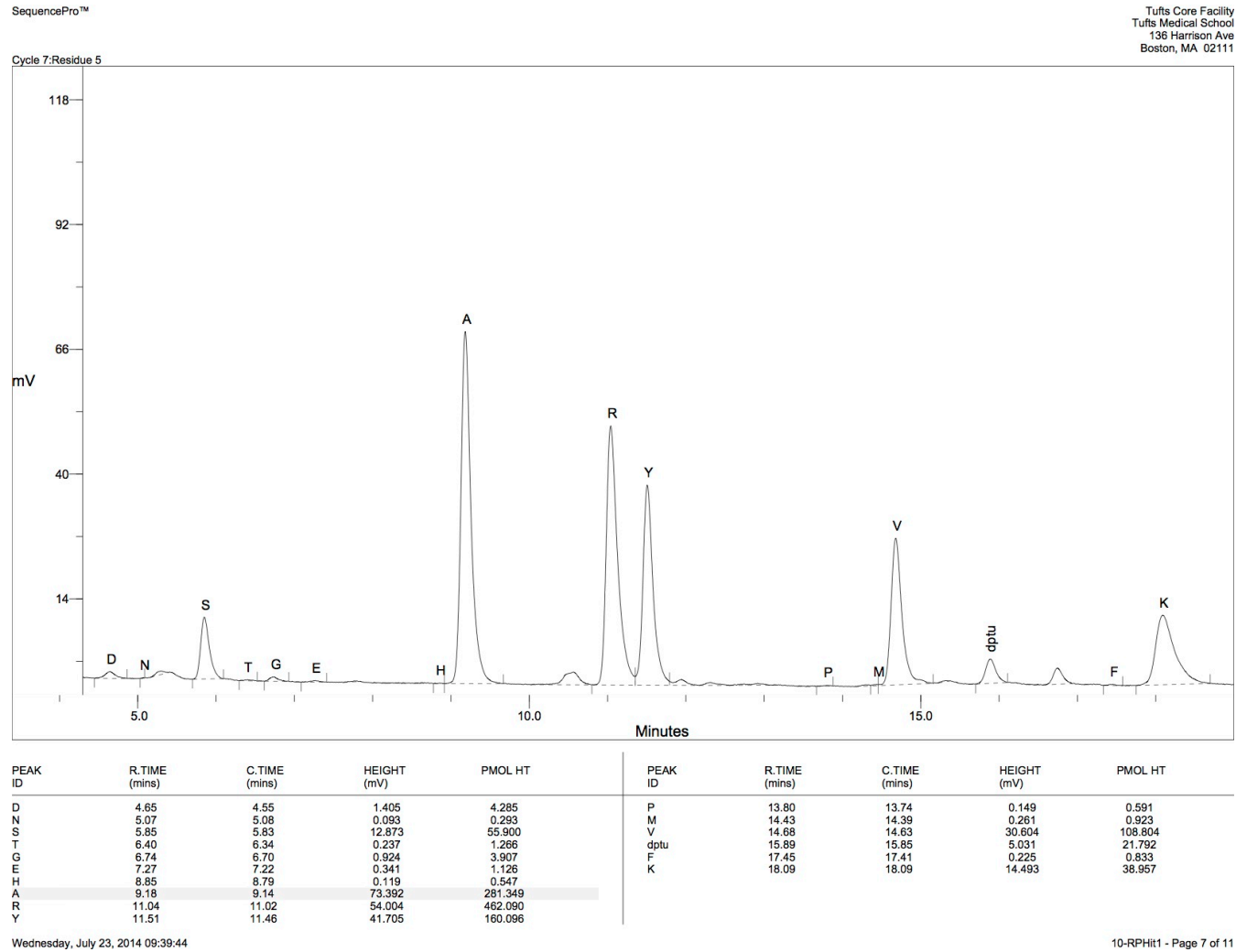


Figure 5.37: SL14 AA 4. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. A is a random position in library.

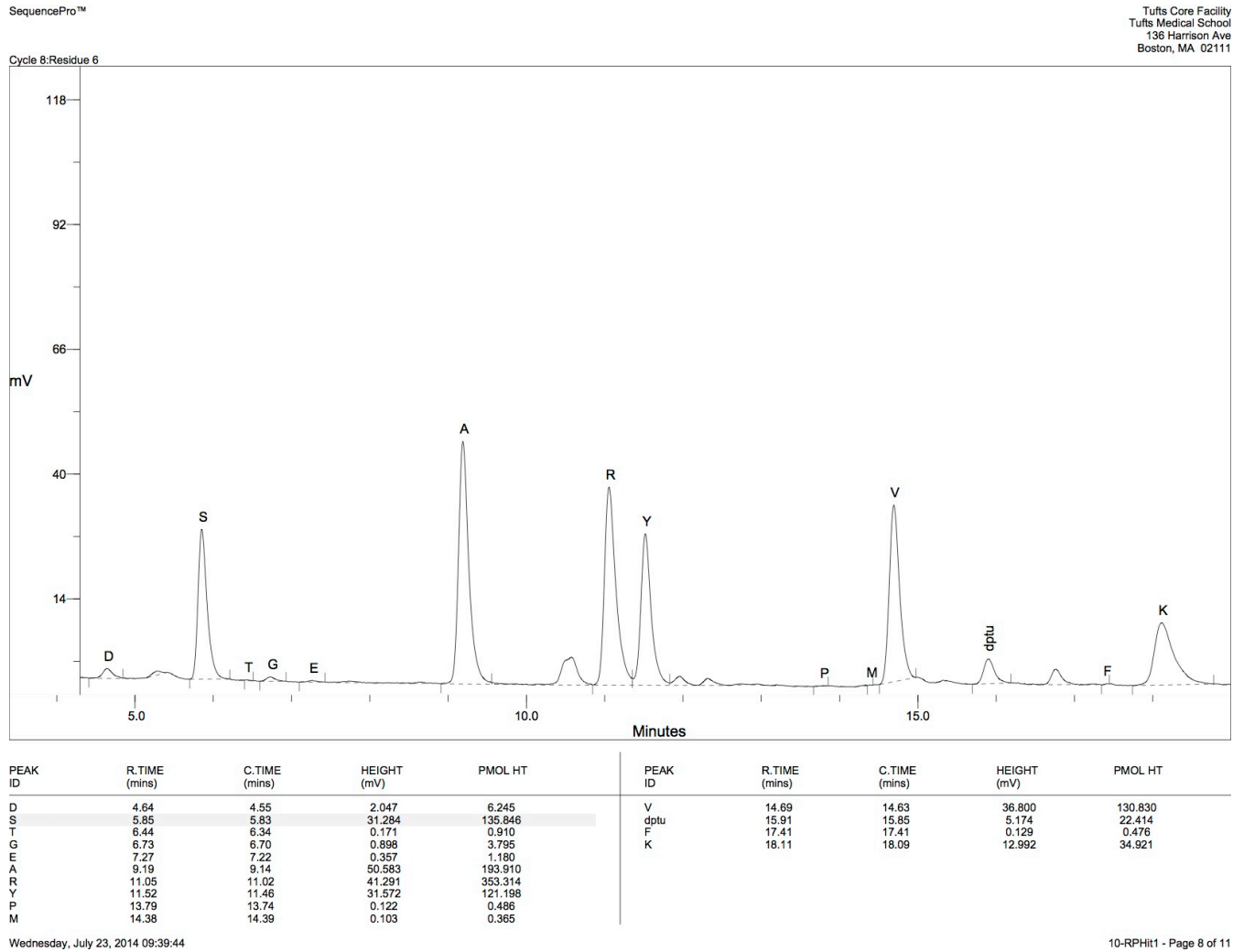


Figure 5.38: SL14 AA 3. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. S is a random position in library.

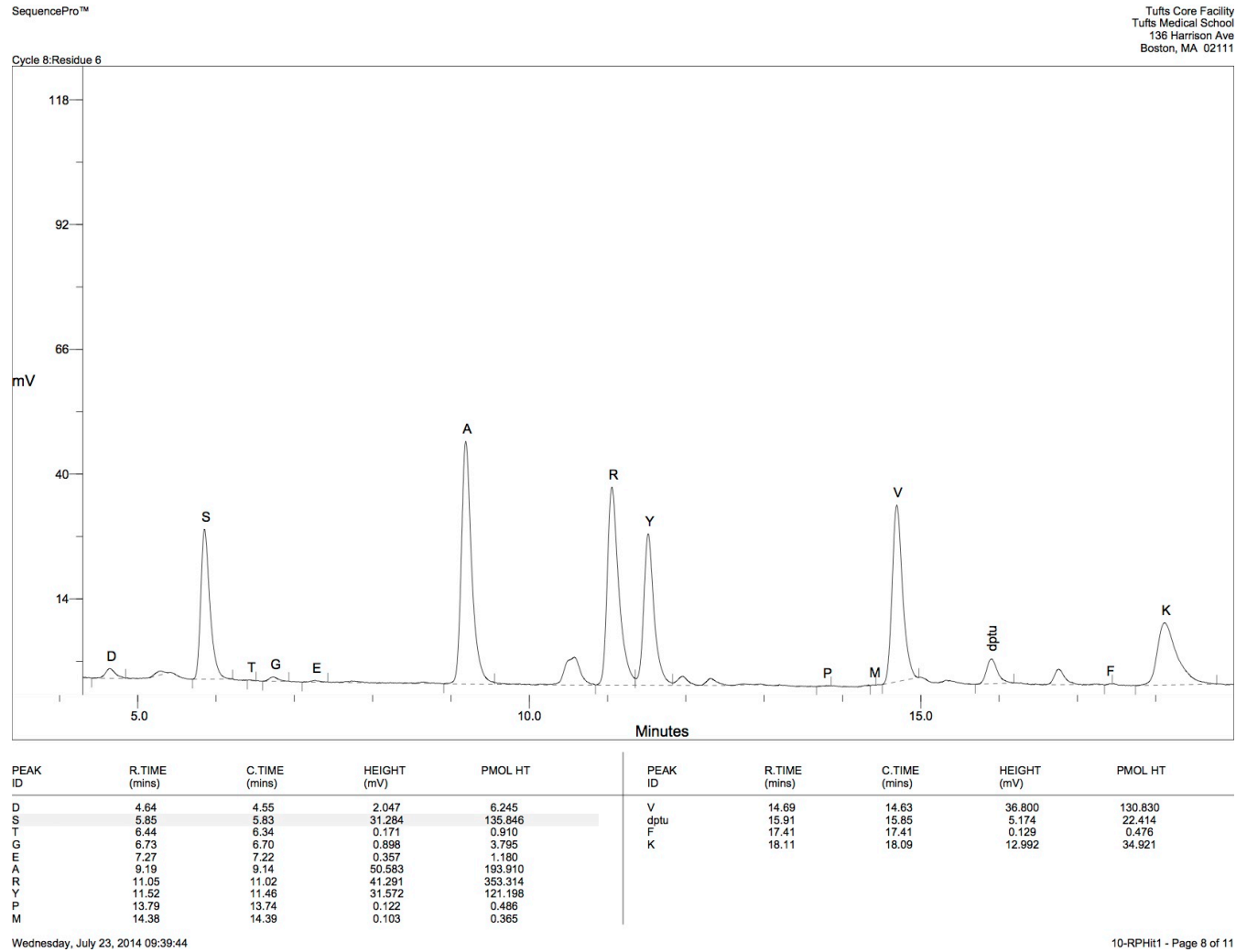


Figure 5.39: SL14 AA 2. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. D\* is a fixed position in library.

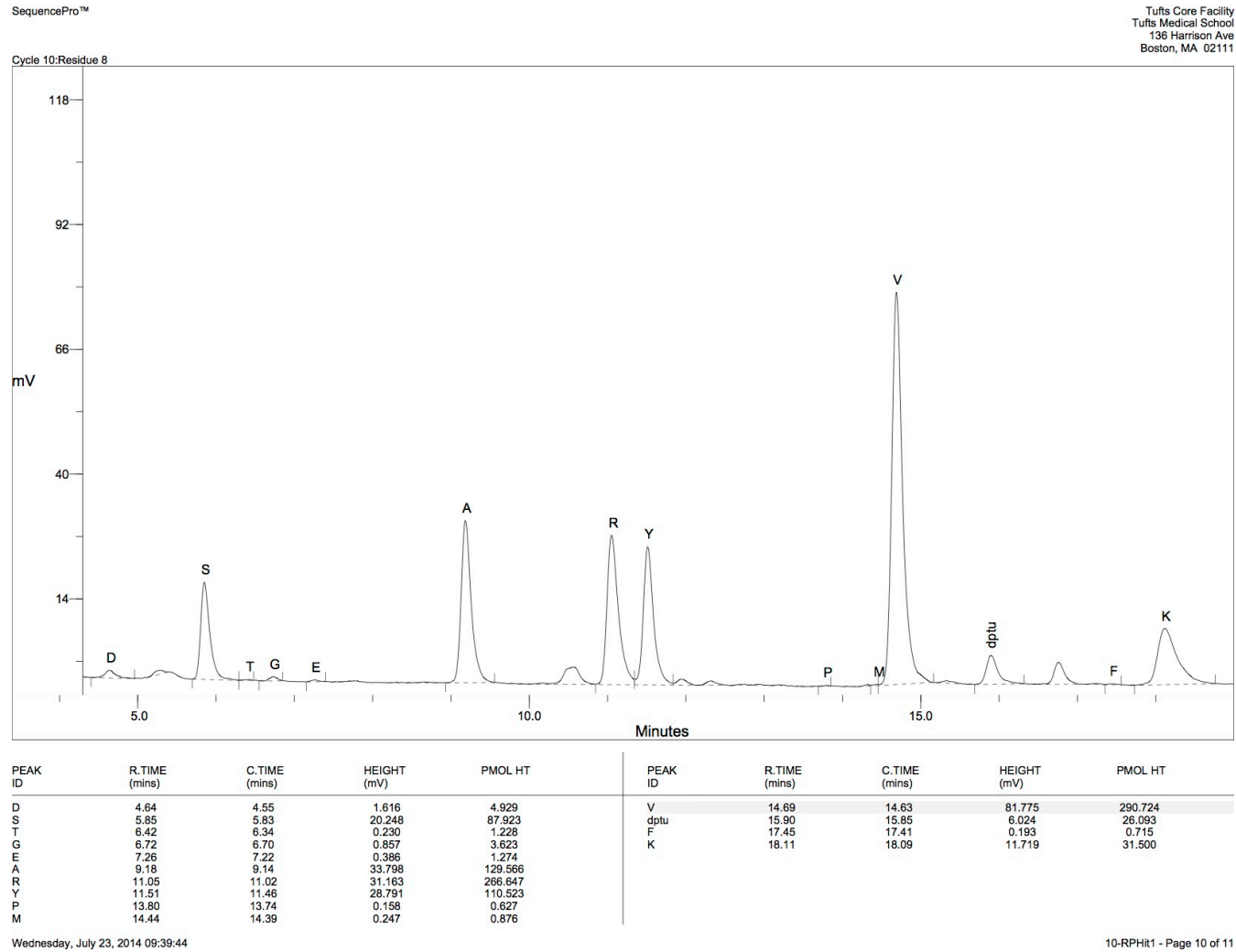


Figure 5.40: SL14 AA 1. The sequence of SL14 is R-Y-D\*-R-A-S-D\*-V-B-B-R-M. V is a random position in library.