

**HPLC and Mass Spectrometry-based Proteomics and Glycoproteomics
for Biomarker Discovery**

by

Yinghua Qiu

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2008**

Doctoral Committee:

**Professor David M. Lubman, Chair
Professor Michael D. Morris
Professor Richard D. Sacks. Deceased
Professor Robert Zand
Assistant Professor Kristina I. Hakansson**

To my family

Acknowledgements

First, I would like to give my most sincere thanks to my advisor Dr. David M. Lubman for his consistent support, guidance, and supervision through these years. I am very thankful for his understanding, aid and encouragement. I am fortunate to have an advisor like Dr. Lubman who is never tired of helping his students. Sincerest appreciation also goes to the members of my dissertation committee, Professor Michael D. Morris, Professor Robert Zand, and Professor Kristina Hakansson. This thesis would not be possible without their help and support.

I would also like to thank many collaborators for the research, Professor Subramaniam Pennathur M.D. from Department of Internal Medicine, Professor Dean E. Brenner M.D. from the UM Cancer Center, Professor David G. Beer Ph.D. from the Department of Surgery at the University of Michigan, as well as professor Sophia Kathariou Ph.D. from the Department of Food Science at North Carolina State University for providing valuable samples and constructive suggestions. I would also like to thank professor Kerby A. Shedden Ph.D. from the Department of Statistics at the University of Michigan for providing statistic tools in data analysis.

I am grateful to all the members of my research group for all of their help and friendship. Here, I would like to extend my special thanks to Suping Zheng, Kan Zhu, and Manoj Pal for their selfless help especially when I first joined the group. I would also like to thank research assistant professor Dr. David Misek and other members in the group.

Most importantly, I would like to express my appreciation to my husband, my parents, my parents-in-law, and my unborn son for their support through all the difficult times. It is a new beginning after graduation, their love will continue to accompany me, encourage me, and bless me at my new work.

Table of Contents

Dedication.....	ii
Acknowledgements.....	iii
List of Figures	ix
List of Tables.....	xii
Abbreviations.....	xiii

Chapter 1. Introduction

1.1 Proteomics	1
1.1.1 Introduction of Proteomics	1
1.1.2 Protein Separation and Protein Expression Mapping	2
1.1.3 Protein Identification by Mass Spectrometry and Database Searching	4
1.2 Glycoproteomics	7
1.2.1 Introduction of Protein Glycosylation.....	7
1.2.2 Isolation of Glycoproteins by Lectin Affinity Chromatography	8
1.2.3 Identification of N-glycosylation Changes by Glycoprotein Microarrays	9
1.2.4 Statistical Analysis of Glycoprotein Microarray Data	10
1.3 References	12

Chapter 2. Proteomic Analysis of Cold Adaptation in a Siberian Permafrost Bacterium – *Exiguobacterium Sibiricum* 255-15 by Two Dimensional Liquid Separation Coupled with Mass Spectrometry

2.1 Introduction.....	15
2.2 Materials and Methods.....	17
2.2.1 Chemicals.....	17
2.2.2 Cell Culture and Sample Preparation	18
2.2.3 Chromatofocusing	19
2.2.4 NPS-RP-HPLC Online Couple with ESI-TOF MS	20
2.2.5 Tryptic Digestion	21
2.2.6 MALDI-TOF MS for Peptide Map Fingerprinting	21
2.2.7 MALDI-QIT-TOF MS MS/MS Peptide Sequencing	22
2.3 Results and Discussion	23
2.3.1 Protein Purification Using 2-D Liquid Phase Separation	23
2.3.2 Protein Identification by PMF Using MALDI-TOF MS	24

2.3.3	Protein Identification by Peptide Sequencing Using MALDI-QIT-TOF MS.....	25
2.3.4	Homologous Cold Shock Proteins (Csps).....	27
2.3.5	Cold Acclimation Proteins (Caps).....	27
2.3.5.1	Caps Characterized as Csps in <i>E. coli</i>	27
2.3.5.2	Caps Characterized as Other Stress-induced Proteins	28
2.3.5.3	Enzymes	29
2.3.5.4	Possible Posttranslational Modifications	30
2.4	Concluding Remarks	30
2.5	References	42

Chapter 3. Effect of Growth Temperature and Culture Medium on the Cryotolerance of Permafrost *Exiguobacterium Sibiricum* 255-15 by Proteome-wide Mass Mapping

3.1	Introduction.....	46
3.2	Materials and Methods.....	48
3.2.1	Chemicals.....	48
3.2.2	<i>E. sibiricum</i> 255-15 Cell Culture	49
3.2.3	Freeze-thawing	49
3.2.4	Cell Lysis	50
3.2.5	2-D Liquid Phase Separation	50
3.2.6	Protein <i>Mr</i> Measurement and Interlysate Quantification by ESI-TOF MS.....	51
3.2.7	Tryptic Digestion.....	52
3.2.8	Protein Identification by PMF and Peptide Sequencing.....	52
3.3	Results.....	53
3.3.1	Freeze-thaw Tolerance	53
3.3.2	Protein Separation and Comparison of 2-D Mass Maps of Protein Expression in Cells Grown Under Different Conditions	54
3.3.3	Proteins Identified by PMF and Peptide Sequencing.....	55
3.3.4	Cold Shock Proteins in Inducing Cryotolerance.....	56
3.3.5	Variation of Cellular Proteins According to Different Growth Conditions.....	56
3.3.5.1	Down-regulation of Proteins Associated with the Improved Cryotolerance.....	57

3.3.5.2 Up-regulation of Proteins Associated with the Improved Cryotolerance.....	58
3.4 Discussion.....	58
3.5 Concluding Remarks.....	64
3.6 References	75

Chapter 4. Plasma Glycoprotein Profiling for Colorectal Cancer Biomarker Identification by Lectin Glycoarray and Lectin Blot

4.1 Introduction.....	78
4.2 Materials and Methods.....	80
4.2.1 Plasma Samples.....	80
4.2.2 Preparation of Glycoprotein Samples.....	81
4.2.2.1 Delipidation and Immunodepletion of the Plasma Samples.....	81
4.2.2.2 <i>N</i> -Glycoprotein Enrichment with ConA Affinity Capture.....	82
4.2.2.3 HPLC Separation of Glycoproteins.....	82
4.2.3 Lectin Glycoarrays.....	83
4.2.4 Statistical Analysis of Lectin Glycoarray Data.....	83
4.2.4.1 Principal Components Analysis (PCA).....	83
4.2.4.2 Hierarchical Clustering.....	84
4.2.4.3 Z-statistics.....	84
4.2.5 SDS-PAGE and Lectin Blot.....	84
4.2.6 Protein Digestion.....	85
4.2.6.1 Tryptic Digestion and <i>N</i> -deglycosylation of NPS-RP-HPLC Fractions.....	85
4.2.6.2 Tryptic Digestion and <i>N</i> -deglycosylation of SDS Gel Bands.....	85
4.2.7 LC-MS/MS for Protein Identification and Glycosylation Site Determination.....	86
4.3 Results and Discussion.....	87
4.3.1 Reduction in the Complexity of Plasma Glycoprotein Mixtures by Immunoaffinity Depletion and Lectin Affinity Enrichment.....	87
4.3.2 Glycoarrays for Identification of <i>N</i> -glycosylation Pattern Changes.....	89
4.3.3 Statistical Analysis of <i>N</i> -glycosylation Pattern Changes	90
4.3.4 Identification of Plasma Biomarkers with Altered <i>N</i> -glycosylation	91
4.3.5 Lectin Blot of a Control Set for Detection of Potential Biomarkers for Differentiating the Different Clinical States.....	93

4.4 Conclusion.....	94
4.5 References	107
Chapter 5. Serum Glycoproteomics of Esophageal Adenocarcinoma by Multi-Lectin Detection Based Glycoprotein Microarrays and Mass Spectrometry	
5.1 Introduction.....	109
5.2 Materials and Methods.....	110
5.2.1 Samples.....	110
5.2.2 Sample Preparation for Protein Microarrays and Lectin Blot	111
5.2.2.1 Delipidation and Immunodepletion	111
5.2.2.2 <i>N</i> -Glycoprotein Enrichment with ConA Affinity Capture.....	112
5.2.2.3 NPS-RPHPLC Separation	112
5.2.3 Lectin Glycoarrays.....	113
5.2.4 Data Analysis and Clustering.....	114
5.2.5 SDS-PAGE and Lectin Blot.....	114
5.2.6 Tryptic Digestion.....	115
5.2.6.1 Tryptic Digestion of NPS-RP-HPLC Fractions.....	115
5.2.6.2 In-gel Trypsin Digestion.....	115
5.2.7 Protein Identification.....	116
5.3 Results and Discussion.....	117
5.3.1 Reduction in the Complexity of Serum Glycoprotein Mixtures by Immunodepletion and Lectin Affinity Enrichment	117
5.3.2 Lectin Glycoarrays for Identification of <i>N</i> -glycosylation Pattern Changes	119
5.3.3 Statistical Analysis of <i>N</i> -glycosylation Pattern Changes	119
5.3.4 Identification of Plasma Biomarkers with Altered <i>N</i> -glycosylation	120
5.4 References	128
Chapter 6. Conclusions.....	129

List of Figures

Figure		
2.1	Comparison of 2-D mass maps between <i>E. sibiricum</i> 255-15 grown at 4°C and 25°C.....	35
2.2	MALDI-QIT-TOF peptide sequencing of 1098.7 (SLDEGQEVEF EITEGAR) from Csp (7.150 kDa, pH 4.54).	36
2.3	ESI-TOF spectra of NPS-RP-HPLC fraction containing 7.150 kDa and 7.414 kDa Csps at 25°C (A) and its corresponding HPLC fraction at 4°C (B). Their deconvoluted spectra are shown in (C) and (D). A protein of 7.442 kDa was exclusively expressed at 4°C as shown in (D)	37
2.4	MALDI-TOF peptide sequencing of NPS-RP-HPLC fraction containing proteins of 7.150 kDa, 7.414 kDa, and 7.442.....	38
2.5	MALDI-QIT-TOF peptide sequencing of 1937.9 (SLDEGQEV SFEVEEGQR) from homologous Csps of 7.409 kDa (pH 4.55) and 7.409 kDa (pH 4.41).....	39
2.6	MALDI-QIT-TOF peptide sequencing of 2098.0 (ESGDDVVFVH FSAIQTDGFK) from Csp (7.409 kDa, pH 4.55)	40
2.7	MALDI-QIT-TOF peptide sequencing of 2125.1 (ENGDDVVFVH FSAIQTDGFK) from Csps (7.409 kDa, pH 4.41).....	41
3.1	Viability of <i>E. sibiricum</i> 255-15 cells grown in liquid broth or on solid agar medium at both 25°C and 4°C after repetitive cycles of freeze-thaw treatment.....	69
3.2	Experimental overview of 2-D liquid phase separation combined with MS for proteomic analysis of cryotolerance in <i>E. sibiricum</i> 255-15.....	70
3.3	2-D mass map of <i>E. sibiricum</i> 255-15 after 25°C TSB-YE liquid broth growth and 25°C TSA-YE agar surface growth of all CF fractions in <i>Mr</i> range of 5-95 kDa with the mass map of lane 5 from each sample highlighted on the left and right plot respectively	71

3.4	Reproducibility demonstration of protein profiling by 2-D mass mapping is generated by two independent CF/NPS-RP-HPLC/ESI-TOF MS experiments with two groups of samples cultured separately.....	72
3.5	(A) MALDI-TOF MS spectrum of protein (18.900kDa, pI 5.44). MALDI-QIT-TOF MS/MS fragmentation of peptide 1837.8 (DDATDETSYGASWIDQVK) was shown in (B) and peptide 1657.9 (FIGIFHDESSLHQK) in (C). Both MS/MS spectra confirm the protein identity.	73
3.6	Alignment of hypothetical protein of 24460 Da in <i>E. sibiricum</i> 255-15 with putative homologous phage shock protein A from <i>Bacillus cereus</i> , <i>Bacillus anthracis</i> , <i>Trichodesmium erythraeum</i> , <i>Clostridium perfringens</i> and <i>Yersinia berovieri</i>	74
4.1	Schematic presentation for high throughput analysis of plasma N-glycosylation pattern changes in colorectal cancer.	99
4.2	(A) Representative chromatographic profiles of immunoaffinity depletion of plasma from normal, adenoma, and colorectal cancer patients using ProteomeLab IgY-12 kit. (B) UV chromatograms of all the plasma samples from colorectal cancer, adenoma, and normal controls.....	100
4.3	Microarray images of N-glycosylated proteins separated from NPS-RP-HPLC with different lectins..	101
4.4A	PCA plot for normalized glycoprotein microarray data derived from the replicates of healthy individuals, adenoma, and colorectal cancer patients. Ovals indicate the areas where the data points of the three groups are distributed.	102
4.4B	Reproducibility demonstration of PCA for normalized glycoprotein microarray data derived from the replicates of healthy individuals, adenoma, and colorectal cancer patients.	103
4.5	Unsupervised hierarchical clustering of glycoprotein microarray data distinguishes colorectal cancer (c1-c6) from adenoma (a1-a5) and normal controls (n1-n9).....	104
4.6	Nano LC-MS/MS mass spectra of (A) doubly charged N-glycosylated peptide GLN*VTLSSGH (m/z = 553.28) from complement 4 and (B) doubly charged N-glycosylated peptide LANENN*ATFYFK from kininogen-1.....	105
4.7	Elevated fucosylation and sialylation of complement C3 (A) and histidine-rich glycoprotein (C) by AAL and SNA blot analysis. The	

	corresponding protein expression levels are shown in (B) for complement C3 and (D) for histidine-rich glycoprotein respectively.....	106
5.1	Representive chromatographic profiles of immunodepletion of serum form normal, high grade dysplasia and esophageal adenocarcinoma patients using ProteomLab IgY-12 kit.....	123
5.2	UV chromatograms of all the serum samples from normal, high grade dysplasia and esophageal adenocarcinoma patients. The similarity among these UV chromatograms among different samples indicated that proteins undergo heterogeneity of glycosylation modifications without necessarily changing protein..	124
5.3	PCA plot for normalized glycoprotein microarray data derived from the replicates of healthy individuals (red), high grade dysplasia (green), and esophageal cancer patients (blue). Ovals indicate the areas where the data points of the three groups are distributed.....	125
5.4	Reproducibility demonstration of PCA for normalized glycoprotein microarray data derived from the replicates of healthy individuals (red), high grade dysplasia (green), and esophageal cancer patients (blue).....	126
5.5	Unsupervised hierarchical clustering of glycoprotein microarray data distinguishes healthy individuals (red), high grade dysplasia (green), and esophageal cancer patients (blue). The clustering method was the average linkage, and the dissimilarity was obtained from the Pearson correlation coefficient....	127

List of Tables

Table	
2.1	Caps identified in <i>E. sibiricum</i> 255-15.....32
2.2	Sequence alignment of the homologous Csps in <i>E. sibiricum</i> 255-15.....34
3.1	Cold shock proteins expressed at different growth conditions66
3.2	Proteins down-regulated at 4°C TSB-YE, 4°C TSA-YE AND 25°C TSA-YE growth that are associated with the improved cryotolerance67
3.3	Proteins up--regulated at 4°C TSB-YE, 4°C TSA-YE AND 25°C TSA-YE growth that are associated with the improved cryotolerance68
4.1	The amount of protein processed through the IgY antibody column and recovered in the depleted fraction from 250 µL plasma samples.....96
4.2A	Z-statistics of differentially glycosylated proteins detected by lectins.97
4.2B	Differentially glycosylated proteins identified with the glycosylation sites.98
5.1	Z-statistics of differentially glycosylated proteins detected by lectins.122

Abbreviations

2-D	two-dimensional
CF	chromatofocusing
<i>pI</i>	isoelectric point
RP	reversed-phase
NPS	nonporous
SB	start buffer
EB	elute buffer
ESI	electrospray ionization
TOF	time-of-flight mass spectrometer
<i>M_r</i>	molecular weight
PMF	peptide mass fingerprinting
MALDI	matrix-assisted laser desorption/ionization
QIT	quadrupole ion trap
ConA	Concanavalin A
AAL	<i>Aleuria aurentia</i> lectin
MAL	<i>Maackia amurensis</i> lectin
PNA	peanut agglutinin
SNA	<i>Sambucus nigra</i> bark lectin