

**Applications of Mass Spectrometric Assays for
Analysis of Serum Protein N-glycosylation
Associated with Pancreatic Cancer**

by

Zhenxin Lin

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2013**

Doctoral Committee:

**Professor David M. Lubman, Chair
Associate Professor Kristina Hakansson
Professor Jairam Menon
Professor Michael D. Morris**

© Zhenxin Lin
All Rights Reserved
2013

DEDICATION

**To my parents, my husband, and my
daughter Annabelle**

ACKNOWLEDGEMENT

Foremost, I would like to express my deepest gratitude to my advisor Dr. David M. Lubman for his insightful guidance and continuous support in my research projects during the past five years at the University of Michigan. I could not have imagined a better advisor for my Ph.D. study.

Besides my advisor, I would like to sincerely thank the rest of my committee: Professor Michael Morris, Professor Kristina Hakansson, and Professor Jairam Menon, for their wonderful comments and encouragement.

My gratitude goes to the current and former colleagues in the Lubman group for their generous help and support in both my research and life. Special thanks go to Dr. Andy Lo and Dr. Chen Li for the helpful training and advice.

I also want to thank my research collaborators, Dr. Mack Ruffin, Dr. Diane Simeone, Dr. Michelle Anderson, Dr. Randall Brand, and Dr. Kerby Shedden, for providing the clinical samples and helping revision of the manuscripts.

Last but not the least, I owe my gratitude to my parents, my parents-in-law, my husband for their love and support. I would love to specially thank my daughter Annabelle, for bringing me tremendous joy and giving my strength!

TABLE OF CONTENTS

DEDICATION.....	ii
ACKNOWLEDGEMENT	iii
LIST OF FIGURES.....	viii
LIST OF TABLES.....	xiv
ABSTRACT	xvi
Chapter 1 Introduction	1
1.1 Pancreatic cancer.....	1
1.2 N-Glycosylation	2
1.3 Global N-glycosylation analysis	4
1.4 Site-specific N-glycosylation analysis	6
1.5 Outline of the dissertation	11
Chapter 2 A Mass Spectrometric Assay for Analysis of Haptoglobin Fucosylation in Pancreatic Cancer27	
2.1 Introduction.....	27
2.2. Experimental Section	29
2.2.1 Serum samples.....	29
2.2.2 Separation of haptoglobin from serum.	30
2.2.3 Deglycosylation and desialylation of haptoglobin.	32
2.2.4 Extraction of desialylated glycans.....	32
2.2.5 Permethylation of glycans.....	32
2.2.6 MALDI-QIT-TOF instrument	33
2.2.7 Data evaluation.....	33
2.3 Results and Discussion.....	34
2.3.1 Purification of haptoglobin from human serum	34

2.3.2 N-glycan profiles of haptoglobin with and without desialylation reveal elevated fucosylation in pancreatic cancer	35
2.3.3 Fucosylation degree indices	37
2.3.4 Reproducibility study and influence of freeze/thaw cycles.....	39
2.3.5 MS/MS study confirms glycan composition and location of fucosylation.....	40
2.4 Conclusion.....	42
Chapter 3 An N-glycosylation Analysis of Human Alpha-2-Macroglobulin Using an Integrated Approach.....	59
3.1 Introduction.....	59
3.2 Experimental Section	61
3.2.1 Serum samples.....	61
3.2.2 Purification of alpha-2-macroglobulin from serum.....	61
3.2.3 Deglycosylation, purification, permethylation and identification of N-glycans	63
3.2.4 LC-ESI-CID/ETD-MS analysis of chymotryptic glycopeptides.....	63
3.2.5 LC-ESI-CID-MS analysis of endo- β -N-acetylglucosaminidase F3 (Endo F3) treated chymotryptic glycopeptides	64
3.2.6 Data analysis	65
3.3 Results and Discussion.....	66
3.3.1 Purification of Alpha-2-macroglobulin from Human Serum.....	67
3.3.2 N-glycan Analysis of Alpha-2-macroglobulin.....	67
3.3.3 MS/MS analysis of N-glycopeptides.....	68
3.3.4 MS/MS analysis of partially deglycosylated N-glycopeptides.....	71
3.4 Conclusion.....	74
Chapter 4 A Method for Label-free Relative Quantitation of site-specific Core-fucosylation by LC-MS/MS.....	85
4.1 Introduction	85
4.2. Experimental Section	87

4.2.1 Serum samples.....	87
4.2.2 Sample preparation.....	87
4.2.3 NanoLC-LTQ-MS analysis of partially deglycosylated peptides.....	88
4.2.4 Quantitative data analysis.....	88
4.2.5 Statistical analysis of fucosylation ratios	89
4.3. Results and Discussion.....	90
4.3.1 Selection of enzyme for proteolysis	90
4.3.2 Label-free quantitative analysis.....	91
4.3.3 Fucosylation ratio indices.....	93
4.3.4 Statistical analysis	94
4.4. Conclusion.....	96
Chapter 5 A Strategy for Profiling of Core-fucosylation in Human Serum Using Lectin Peptide Enrichment and HCD-MS/MS	112
5.1 Introduction.....	112
5.2 Materials and methods	115
5.2.1 Serum immunoaffinity depletion.....	115
5.2.2 Peptide level LCA enrichment	116
5.2.3 Protein level LCA enrichment.....	117
5.2.4 Isobaric iTRAQ labeling	117
5.2.5 LC-ESI-MS/MS analysis.....	118
5.2.6 Data analysis	119
5.3 Results and Discussion.....	120
5.3.1 Analytical strategy.....	120
5.3.2 MS/MS methods.....	122
5.3.3 Assignment of core-fucosylation sites.....	124
5.3.4 Bioinformatic concerns due to neutral loss	125
5.3.5 Quantitative study of the assay by iTRAQ labeling.....	126

5.4 Conclusion.....	127
Chapter 6 Conclusion.....	143

LIST OF FIGURES

Figure 1. 1. Legends of monosaccharides of human N-glycans.....	14
Figure 1. 2. Three types of N-glycans.....	15
Figure 1. 3. MS ⁿ spectra of an ovalbumin glycan at m/z 1923.2 with composition of Man ₃ GlcNAc ₅ . The diagnostic ion of m/z 893.5 in MS ⁵ confirms the isomer structure B. (Reprinted with permission from (49). Copyright (2007) American Chemical Society) ..	16
Figure 1. 4. HCD-triggered-ETD analysis of a ribonuclease B glycopeptide. (a) MS survey scan, (b) HCD MS/MS spectrum of precursor ion at m/z 645.6194, and (c) ETD-MS/MS spectrum of precursor ion at m/z 645.6194. (Reprinted with permission from (66). Copyright (2012) American Chemical Society) ..	17
Figure 1.5. Quantitative mass spectrometric analysis with isobaric TMT labeling. Differentially labeled peptides are not distinguishable in MS ¹ , MS ² generates reporter ions with different intensities on which quantification is based. (Reprinted with permission from (78). Copyright (2011) Nature Publishing Group).....	18
Figure 1. 6. In an MRM-MS analysis, precursor ions are preselected in Q1 and fragmented in Q2. Predefined fragment ions are filtered in Q3. Synthetic peptides are labeled with a stable isotopic tag as shown with asterisks and spiked in the sample to improve the precision of relative quantitation. (Reprinted with permission from (79). Copyright (2013) Nature Publishing Group).....	19
Figure 1. 7. Research strategies of the work discussed in this dissertation.	20
Figure 2. 1. Workflow to characterize N-glycan structures of haptoglobin and study	

fucosylation differences between pancreatic cancer and other benign pancreatic diseases/normal controls.....	45
Figure 2. 2. Purification of haptoglobin from 10uL of human serum. (a) Tryptic peptide profile of haptoglobin resulted from 10-min on-plate digestion followed by MALDI-QIT-TOF MS analysis. (b) Patterns of haptoglobin α -1, α -2 and β chain on SDS-PAGE followed by silver staining.....	46
Figure 2. 3. (a) MALDI-QIT-TOF MS spectra of haptoglobin glycans from pooled serum of normal controls (upper panel) and pancreatic cancer patients (lower panel). (b) MALDI-QIT-TOF spectrum of desialylated haptoglobin glycans from serum of a normal control (upper panel) and a stage IV pancreatic cancer patient (lower panel). (red triangle-Fuc, blue square-GlcNAc, green circle-Man, yellow circle-Gal, purple diamond-NeuAc).....	47
Figure 2. 4. Zoom-in MALDI-QIT-TOF mass spectra showing fucosylation difference in tri-antennary and tetra-antennary N-glycans of haptoglobin from serum of pancreatic cancer, normal control, chronic pancreatitis and type II diabetes.....	49
Figure 2. 5. Scatter plot of fucosylation degrees of haptoglobin N-glycans of pancreatic cancer patients and non-cancer samples. Fucosylation degrees are elevated in pancreatic cancer. P-value of pair-wise t-test is 1.9×10^{-7}	50
Figure 2. 6. MALDI-QIT-TOF MS/MS spectra of glycans at m/z 2867.46 (a), m/z 2244.20 (b), and m/z 3142.59 (c). Potential structures of fragment ions are labeled.....	51
Figure 3. 1. Strategy to characterize site-specific N-glycosylation of alpha-2-macroglobulin from human serum.....	75
Figure 3. 2. (a) Peptide finger print of immunoprecipitated alpha-2-macroglobulin after on	

plate tryptic digestion followed by MALDI-QIT-TOF MS analysis. (b) SDS-PAGE followed by silver staining of 1/10 of immunoprecipitated alpha-2-macroglobulin.76

Figure 3. 3. (a) Representative mass spectrum of N-glycan profile of alpha-2-macroglobulin of a normal serum sample. Four biantennary complex type glycans were identified. (b) MS/MS spectrum of a fucosylated biantennary glycan at m/z 2966.44.77

Figure 3. 4. (a) Base peak chromatogram of LC-MS analysis of chymotryptic N-glycopeptides purified by ZIC-HILIC. The peaks at 27.30 min, 29.85 min, 37.47 min, 38.22 min and 41.47 min are the primary glycopeptides. (b) Extracted ion chromatogram of oxonium fragment ion at m/z 366 (GlcNAc-Gal) from CID MS/MS spectra. (c) Summed mass spectrum at 37.60-38.60 min revealed some major N-glycopeptides.....78

Figure 3. 5. (a) CID MS/MS spectrum of the glycopeptide at m/z 1026.7 (sequence ESVRGNRS LF). Glycosidic bond cleavages were observed, resulting in b,y type ions. (b) ETD MS/MS spectrum of the same glycopeptide. c,z type ions were observed with the intact glycan structure.79

Figure 3. 6. (a) CID MS/MS of the fucosylated glycopeptide (SN(+GlcNAc-Fuc)ATTDEHGLVQF) at m/z 884.6. The major fragment at m/z 811.6 is the product ion after neutral loss of fucose from the precursor ion. Minimal peptide backbone fragmentation was observed. (b) CID MS3 of m/z 811.6 from (a) showed extensive fragmentation along the peptide backbone, providing both peptide sequence information and the glycosylation site. (c) CID MS/MS of the non-fucosylated glycopeptide (SN(+GlcNAc)ATTDEHGLVQF) at m/z 811.6. (c) has great similarity with (b), revealing presence of both fucosylation and non-fucosylation at the same site.80

Figure 4. 1. Workflow of the study	98
Figure 4. 2. Inconsistent labeling efficiency of chymotrypsin-catalyzed 218O labeling. Top trace: labeling pattern of peptide SNATTDEHGLVQF. Bottom trace: labeling pattern of peptide IYLDKVSQNQL.....	99
Figure 4. 3. XIC of the peptides before and after TMT labeling. Top: before TMT labeling; bottom: after TMT labeling. (RT: retention time, in mins, AA: peak area, SN: signal over noise ratio, BP: base peak m/z).....	100
Figure 4. 4. Extracted ion chromatograms (XIC) of 5 pairs of partially deglycosylated peptides. The top chromatogram of each panel is the non-core-fucosylated peptide, and the bottom chromatogram in the same panel is the fucosylated counterpart with the same peptide backbone. (RT: retention time, in mins, AA: peak area, SN: signal over noise ratio, BP: base peak m/z).....	101
Figure 4. 5. Scatterplots of log fucosylation ratios at 5 peptide sequences corresponding to 3 glycosylation sites. Twenty samples per disease states (normal, chronic pancreatitis and pancreatic cancer) were included in comparison, except in Figure 4.3(b) where only 19 data points are shown in the pancreatitis group due to removal of an outlier.....	102
Figure 4. 6. Comparison of A2MG N-glycan profiles between normal control (top), chronic pancreatitis (middle) and pancreatic cancer (bottom).....	103
Figure 4. 7. Correlation analysis of fucosylation ratios between peptide sequences of the same site ((a) and (b)), and between peptide sequences of different sites ((c)-(j)). Pearson's r higher than 0.5 indicates strong correlation. Fucosylation ratios of peptide sequences corresponding to the same site (site 396: 811.5/884.5 and 893.0/966.0 in (a) and site 1424: 611.0/684.0 and 749.2/822.2 in (b)). Fucosylation ratios of site 396 and site 410	

((c),(d)), and site 1424 and site 410 ((e) and (f)) are not correlated. Fucosylation ratios of site 396 and site 1424 are correlated ((g)-(j)). 104

Figure 5. 1. Enrichment workflow comparisons for protocol optimization. (a) label-free strategies, where Strategy A and B are peptide level LCA enrichment, and Strategy C is protein level LCA enrichment. Two hundred and fifty micrograms of serum proteins were used.. (b) iTRAQ labeling strategy, where four aliquots of 100 µg serum proteins tryptic digests were differentially labeled and combined for LCA enrichment 128

Figure 5. 2. Comparison of numbers of identified unique core-fucosylated peptides (a) and core-fucosylated proteins (b) using different workflows as shown in Figure 5.1(a). 129

Figure 5. 3. Comparison of neutral loss-triggered CID-MS3 and HCD MS/MS of the same core-fucosylated peptide. (a) CID MS/MS of the core-fucosylated peptide which is dominated by the neutral loss of core fucose (m/z 934.81). (b) Neutral loss triggered-CID MS3 of fragment m/z 934.81. (c) HCD MS/MS of the same core-fucosylated peptide.. 130

Figure 5. 4. Comparison of numbers of identified unique core-fucosylated peptides (a) and core-fucosylated proteins (b) using HCD, neutral loss-triggered HCD and neutral loss-triggered CID..... 132

Figure 5. 5. Histogram of mass accuracy distribution of identified core-fucosylated peptides by HCD-MS/MS..... 133

Figure 5. 6. (a) Distribution of singly and multiply core-fucosylated proteins identified and (b) proportion of novel identifications, reported identifications and non-motif identifications..... 134

Figure 5. 7. HCD-MS/MS spectrum of a newly identified glycosylation site of Sushi, nidogen and EGF-like domain-containing protein 1. The blue annotations show identified

fragments, and the red annotations show unidentified but manually assigned fragments.135

Figure 5. 8. Summary of ratios of reporter ion abundances in an isobaric iTRAQ experiment where four aliquots of 100 µg of serum protein digests were labeled with 4-plex reagents (reporter ion masses 114, 115, 116 and 117) and mixed for subsequent analysis as shown in Figure 5.1 (b). Ratios were constructed against the abundances of reporter ions 114.....136

LIST OF TABLES

Table 2. 1. Eight desialylated N-glycans identified in human serum.	53
Table 2. 2. Summary of sample disease states and number of correct classification with fucosylation degree index	54
Table 2. 3. (a) Reproducibility test of the assay for four aliquots of one normal serum sample processed following the workflow in Figure 1, (b) Influence of freeze/thaw cycles evaluated for three aliquots of one pancreatic cancer serum sample frozen/thawed once, twice and four times, and processed independently.....	55
Table 3. 1. N-glycopeptides identified in LC-CID/ETD-MS/MS analysis.....	81
Table 3. 2. Glycosylation sites identified by CID MS/MS of Endo F3 treated chymotryptic N-glycopeptides of alpha-2-macroglobulin.	82
Table 4. 1. Demographic information and cancer stage information of human serum samples enrolled in this study.	106
Table 4. 2. Glycosylation sites identified with different proteolysis enzymes or enzyme combinations. Glycosylation site is labeled as N. NF and F indicate fucosylated glycopeptide and non-core-fucosylated glycopeptide respectively. NA indicates that no fucosylated peptide was detected.....	107
Table 4. 3. Reproducibility test of the assay for four aliquots of the same normal serum sample processed on four different days. The values in row 2-5 are fucosylation ratios.	108
Table 4. 4. Statistical summary of core-fucosylation ratios in different disease states. SD stands for standard deviation	109

Table 5. 1. List of core-fucosylated peptide containing the N-X-S/T motif in this study.137

ABSTRACT

Pancreatic cancer has the worst prognosis among all cancers, mainly due to lack of effective diagnostics capable of early-stage detection. There have been many studies on pancreatic cancer biomarkers, among which serum protein markers are of particular interest as a minimally-invasive technique.

In this dissertation, multiple mass spectrometric assays have been utilized to characterize protein N-glycosylation at the glycan, glycopeptide, and peptide levels both qualitatively and quantitatively, for identification of serum-based pancreatic cancer biomarkers. The assays incorporated optimization of sample pre-processing, liquid chromatography separation, and analysis by mass spectrometry. State-of-the-art mass spectrometric fragmentation methods including collision-induced dissociation (CID), electron transfer dissociation (ETD) and higher-energy collisional dissociation (HCD) have been utilized to study N-glycosylation of both individual proteins isolated from human serum and depleted human serum.

Chapter 2 describes a MALDI-MS/MS based method for analysis of N-glycans which integrates N-glycan extraction, desialylation, permethylation, structure elucidation, and fucosylation degree measurement. The fucosylation degree of human serum haptoglobin is elevated in pancreatic cancer relative to non-cancer conditions, including normal controls, chronic pancreatitis and type II diabetes. Chapter 3 presents a CID/ETD-MS/MS method for detailing the site-specific glycosylation patterns of human serum alpha-2-macroglobulin (A2MG). For targeted analysis of A2MG site-specific

core-fucosylation, an endoglycosidase-assisted strategy was utilized. All eight potential N-glycosylation sites were identified with six of them found to be core-fucosylated. Chapter 4 extends the work in Chapter 3 to a quantitative level and investigates the alterations of A2MG core-fucosylation at specific sites in pancreatic cancer using a label-free method. Core-fucosylation ratios at sites N396, N410 and N1424 were found to decrease in pancreatic cancer relative to normal controls. Chapter 5 describes total core-fucosylation profiling of human serum, where the core-fucosylated glycopeptide enrichment and mass spectrometric methods were optimized, leading to identification of 135 core-fucosylation sites in serum. The quantitative aspect of this assay used an isobaric labeling strategy, which may prove potentially useful for future high-throughput serum pancreatic cancer biomarker screening.

The results of this thesis provide insight into the potential of protein fucosylation alterations as pancreatic cancer biomarkers, and relevant strategies will be useful in other clinical applications.

Chapter 1

Introduction

1.1 Pancreatic cancer

Pancreatic ductal adenocarcinoma, the most common form of pancreatic cancer, is the fourth leading cause of cancer death in United States. In the year 2008, 37,700 people were diagnosed with pancreatic cancer, and 34,300 people died from it.^{1, 2} Pancreatic cancer has the worst prognosis among all types of cancers with a five year survival rate at 4%.¹⁻³ Pancreatic cancer is known to be a silent cancer, whose symptoms do not become noticeable until an advanced stage, where the tumor migrates to the surrounding or distant organs and surgery is no longer an effective option.² While the cause of pancreatic cancer is not well understood, risk factors include family history, age, gender and smoking.

Current diagnostics of pancreatic cancer include imaging techniques (Computer Tomography, Magnetic Resonance Imaging and Positron Emission Tomography) and biopsy, which are either expensive or invasive. It is desirable to have a routine blood test which is capable of detecting pancreatic cancer at an early stage, especially for high-risk populations⁴. The only FDA-approved blood test for pancreatic cancer is the level of protein CA 19-9. However, this diagnostic has its own limitations including: (1) low sensitivity and specificity; (2) lack of power in detection of early stage pancreatic cancer; and (3) false positives as 10% of the patients are not able to synthesize CA 19-9 even at an advanced stage of pancreatic cancer.^{1, 5} With the advent of LC-MS/MS techniques, the

last decade has seen a plethora of research exploring human serum proteomics for the identification of pancreatic cancer biomarkers.⁵⁻¹⁰

1.2 N-Glycosylation

Glycosylation where carbohydrates attach to proteins, is one of the most prevalent post-translational modifications. There are two major types of glycosylation. N-glycosylation, where N-glycans attach to the amide nitrogen of Asn in the sequence of Asn-Xxx-Ser/Thr (Xxx cannot be Pro), and O-glycosylation, where glycans attach to hydroxyl oxygen of Ser and Thr. N-glycosylation is usually more complex compared to O-glycosylation and is the focus of this dissertation. Common monosaccharides involved in N-glycosylation in humans are N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), mannose (Man), galactose (Gal), fucose (Fuc), and N-acetylneuraminic acid (NeuAc) as shown in Figure 1.1. N-glycosylation is initiated in the endoplasmic reticulum where a 14-monomer structure is assembled and transferred to the asparagine residues on proteins, and is further extended and modified in the Golgi apparatus.¹¹ Various glycosidases which truncate the monosaccharide units (eg. glucosidase and mannosidase) and glycosyltransferases which add monosaccharide units (eg. fucosyltransferase and N-acetylglucosaminyltransferase) are involved in the process of N-glycosylation.^{12, 13}

N-glycans are usually highly branched and diverse in structure. It is speculated that over 3000 glycoforms potentially exist based on the available monosaccharides and linkages.¹⁴ There are three types of N-glycans, all sharing the same pentasaccharide core structure (Man₃GlcNAc₂). As shown in Figure 1.2, high-mannose type glycans extend the

core structure with mannoses only; complex type glycans add any monosaccharide other than mannose to the core structure; hybrid type glycans are a combination of high-mannose and complex type glycans.

It is important to note that unlike protein expression, N-glycosylation occurs in a non-template manner with glycosidic bonds formed in multiple ways¹⁵; hence, changes in the physiological and pathological cellular environment may alter the N-glycosylation. N-glycosylation exerts profound impacts on the activity and functions of proteins and cells in various ways, including protein folding, protein binding, protein recognition, cell-cell interaction, cell-metastasis and immune responses.¹⁶⁻¹⁸ Aberrations of N-glycosylation have been found to be associated with different types of diseases, including cancers, and may serve as promising blood-based biomarkers. The common types of N-glycosylation aberrations in serum/plasma proteins include increased branching¹⁹, and hyper or under expression of mannosylation^{20, 21}, sialylation^{22, 23} and fucosylation^{24, 25}.

Early studies of N-glycosylation utilized lectin blotting or lectin microarray^{26, 27}. Lectins are a family of proteins which recognize and specifically bind to different glycan epitopes. By conjugating lectins with chemiluminescent or fluorescent tags and measuring the chemiluminescence or fluorescence signals, different N-glycosylation patterns can be identified and quantified. However, limited specificity of lectins is a concern, and this method does not provide enough structural information of the glycan. Methods providing more structural information, such as exoglycosidase-assisted HPLC and tandem MS analysis, are gaining popularity in the field of glycoproteomics. A comprehensive study of N-glycosylation includes identifying the composition of

monosaccharides, revealing the glycan structures (including branching, linkage positions and substitution of fucose or sialic acids), quantifying the glycoforms, locating the glycosylation site, and determining the site occupancy.

1.3 Global N-glycosylation analysis

Although it is desirable to characterize the N-glycosylation in aforementioned levels simultaneously, it is technically difficult. Usually N-glycans are released from the protein or peptide backbones with N-glycans and deglycosylated peptides studied separately. The analysis of N-glycans offers information in glycan composition and structure details, while the study of deglycosylated peptides reveals glycosylation site and site occupancy.

A commonly used enzyme which liberates the N-glycans is peptide-N-glycosidase F (PNGase F), which cleaves the amide bond between the GlcNAc and Asn. This hydrolysis is accompanied by converting Asn to Asp and introducing a 0.98 Da mass shift on the peptide which is frequently used as an indicator of the N-glycosylation site^{28,29}. However, deamidation of the Asn may occur spontaneously, introducing ambiguity in the site-identification.³⁰ To alleviate, but not entirely overcome this problem, PNGase F digestion may be performed in H₂¹⁸O, producing a 3.98 Da mass increment²⁹. An alternative is partial deglycosylation with endoglycosidases, which hydrolyze the glycosidic bond between the two core GlcNAc³¹. This generates a 203.08 Da (if the site is not core-fucosylated) or 349.13 Da (if the site is core-fucosylated) mass increase which enables site-identification even with a low-resolution mass spectrometer.

Similar to phosphorylation studies, enrichment of N-glycopeptides and N-glycans is essential. Various enrichment strategies including lectin affinity^{22, 32}, hydrazide

chemistry^{33, 34}, hydrophilic affinity^{35, 36}, and hydrophilic interaction (HILIC)^{37, 38} have been applied to separate the N-glycopeptides from more abundant non-glycopeptides. Porous graphitized carbon (PGC)³⁹, cation exchange⁴⁰ and HILIC⁴¹ are commonly used for desalting and pre-concentration of N-glycans.

N-glycans are usually derivatized for both HPLC and MS analysis. Unlike peptides, which have UV absorbance at 210 nm, the natural N-glycans do not absorb UV light sufficiently to give a reasonable detector signal. For HPLC analysis, N-glycans are usually labeled with 2-aminobenzamide (2-AB) or 2-aminobenzoic acid (2-AA) at the reducing end through reductive amination, separated by C₁₈, HILIC or PGC, and detected by fluorescence with an excitation wavelength set at 320 nm and emission wavelength set at 420 nm⁴². The composition information is obtained by matching the retention time with a glucose ladder index¹⁵. Although quantification information may be obtained by measuring the fluorescence, the chromatographic resolution and sensitivity of this method are not satisfying⁴³.

An alternative to analyzing N-glycans is to eliminate the LC separation and use MALDI-MS. For this purpose, glycans are usually permethylated, where the hydroxyl groups are substituted by methoxy groups. Permethylation increases the ionization efficiency and stability of glycans while enabling the detection of acidic and neutral glycans simultaneously in positive ion mode⁴⁴. CID of N-glycans usually produces glycosidic bond cleavages, generating B, Y ions which could be used to infer the monosaccharide compositions. When using high-energy CID in a time-of-flight mass spectrometer, cross-ring cleavages (A,X ions) are also observed and may be used to reveal glycosidic linkage patterns⁴⁵⁻⁴⁷.

Notably, a unique advantage of using an ion trap instrument for permethylated glycan analysis is that the MSⁿ experiment is made possible. Isobaric fragments or fragments with close masses can appear in MS² spectra although they result from different fragmentation pathways, and such ambiguity of assignment can be reduced by further analysis of MS³ which probes the formation of the ions in MS²^{48, 49}. Furthermore, glycan branching isomerism information can be obtained by additional fragmentation of MS² ions^{48, 50}. For example, Reinhold etc. reported using MSⁿ to determine the structural isomer of a bovine ovalbumin glycan at *m/z* 1923.1 (composition Man₃GlcNAc₅) as shown in Figure 1.3 where Structure A was confirmed based on the presence of the fragment ion at *m/z* 893.2 in MS⁵.

Quantitative analysis can be performed both label-free^{21, 51} or with stable isotopic labeling^{52, 53} (CH₃I, CH₂DI, CHD₂I and CD₃I or ¹³CH₃I) for comparative glycomic studies across various conditions.

1.4 Site-specific N-glycosylation analysis

The aforementioned methods analyze glycan and deglycosylated peptides separately, and the information provided is limited due to the inability to characterize site-specific glycosylation, which is crucial in disease biomarker studies. Previous study of α -1-acid glycoproteins identified increased triantennary α 1,3 fucosylation at sites 3,4 and increased tetraantennary α 1,3 fucosylation at site 3, 4, and 5 for chronic inflammation.⁵⁴ Another study reported an increase of terminal fucosylation at site 3 in serum haptoglobin of pancreatic cancer³⁶.

The analysis of native N-glycopeptides for site-specific glycosylation is much more

challenging compared to the separate analysis of glycan and deglycosylated peptides. A single protein may have multiple N-glycosylation sites and there is usually more than one glycoform associated with each site. The abundance of a single glycopeptide is usually very low, requiring mass spectrometric analysis methods with high sensitivity. Furthermore, the low ionization efficiency of glycopeptides and ionization suppression from the non-glycopeptides pose another technical obstacle for analysis. Advances in the enrichment and separation methods, and novel mass spectrometric methods can be used to circumvent the problems. For example, the use of new generation mass spectrometers which enable the selection of top 10-20 precursor ions for MS/MS scans with significantly increased scan speed greatly enhances the chance of glycopeptides to be analyzed⁵⁵.

With respect to fragmentation methods, conventional CID mainly produces fragments across glycosidic bonds, leaving the peptide backbone intact. The production of oxonium ions ($[\text{HexNAc}+\text{H}]^+$ (204.1 Da), $[\text{NeuAc}+\text{H}]^+$ (292.1 Da), $[\text{HexNAc-Hex}+\text{H}]^+$ (366.1 Da) and $[\text{NeuAc-Hex-HexNAc}+\text{H}]^+$ (657.2 Da)) are normally used as markers for presence of the glycopeptides⁵⁶. Although some studies use the masses of Y_1 (peptide+HexNAc) ions to deduce the peptide sequences^{36, 57, 58}, the false positive rate is higher and requires prior knowledge of the protein, which limits this method to profiling of single proteins or simple protein mixtures. Aside from CID, electron transfer dissociation (ETD) is also widely utilized, often in tandem with CID. ETD preferentially produces extensive fragmentation on the peptide backbone, generating sequence-related c,z- ions, and normally leaving the glycan portion intact^{59, 60}. By using CID and ETD in the same LC-MS run, both glycan composition and peptide sequence information can be obtained

simultaneously^{56, 61, 62}. However, ETD has comparatively low fragmentation efficiency, especially for ions with low charge states⁶³. Recently, the emerging higher-energy collisional dissociation (HCD) has been used for intact glycopeptide analysis^{38, 64-66}. HCD mainly generates glycosidic bond cleavages, producing oxonium ions, and to a lesser extent peptide fragments (b, y ions similar to CID). Compared to CID in an ion trap, the 1/3 cut-off issue is not present for HCD fragmentation method and low-mass oxonium ions, such as [HexNAc+H]⁺ (204.1 Da) and [NeuAc+H]⁺ (292.1 Da), can be measured. Compared to CID in high-resolution instruments such as Q-TOF, HCD may be used in conjunction with ETD in an “oxonium ion-triggered ETD” manner which enables higher-efficiency identification of glycopeptides^{65, 67}. An example is demonstrated in Singh and coworkers' study which identifies the glycoforms at various N-glycosylation sites in ribonuclease B⁶⁷ as shown in Figure 1.4.

Analysis of MS/MS spectra of intact glycopeptides is often done manually or semi-automatically by experienced personnel for targeted studies of single proteins. For large scale glycoproteomic profiling, automatic database search is required. However, it is difficult to implement the most widely used database search algorithms such as Sequest, Mascot and X!Tandem in analysis of ETD data for glycosylation site identification. The main reason is the high heterogeneity of N-glycosylation, where over a thousand glycoforms have to be specified as possible modifications in a search. This greatly increases the search time and storage space, and favors random assignment which increases false-positive rates⁶⁸.

In order to reduce the microheterogeneity of N-glycosylation and enable automatic database search, a series of exoglycosidases (eg. neuraminidases, fucosidases,

galactosidases) or endoglycosidases (eg. Endo H, D, and F) may be used to target subtypes of N-glycosylation, such as galactosylation or fucosylation. Endoglycosidases are a family of enzymes which cleave between the two innermost GlcNAc to leave only the core GlcNAc or GlcNAc-Fuc attached to the peptides. Endoglycosidase digestion produces mass increments of 203.1 Da (GlcNAc) or 349.1 Da (GlcNAc-Fuc) which may be utilized to identify protein N-glycosylation sites and core-fucosylation sites^{31, 69-72}. Compared to PNGase F deglycosylation, this strategy eliminates the false-positive identification due to spontaneous deamidation, and enables detection using low-resolution mass spectrometers. Compared to intact glycopeptide analysis, this strategy greatly increases the ionization efficiency of glycopeptides, and reduces glycosylation microheterogeneity while retaining the core-fucosylation patterns, which has been found to alter in various proteins in different cancers^{24, 72-74}.

Site-specific glycosylation quantification may be achieved by both label-free and label-based strategies. The label-free method is usually based on the intensities of extracted precursor ion chromatograms. Label-based methods introduce stable isotopic or isobaric mass tags to different samples, and compare respective mass spectrometric signal abundances for relative quantification. The differentially labeled species are identical or very similar in chemical properties. Various labeling methods may be used, including chemical labeling (isotopic reductive dimethylation, isotopic/isobaric tag for relative and absolute quantification (iTRAQ)⁷⁵ and tandem mass tag (TMT))⁷⁶, metabolic labeling (stable isotope labeling by amino acids in cell culture (SILAC))⁷⁷, and enzymatic labeling (H₂¹⁸O peptide C-terminal labeling)⁷⁰.

Both precursor ion intensities and fragment ion intensities can be utilized for

quantification based on the assumption that the mass spectrometric signal is proportional to the analyte concentration. In precursor ion intensity-based quantification, normalization is usually required to account for the variation in ionization efficiency, either against an internal standard which may or may not be isotopically labeled, or against the sum of peak intensities of all analytes. Although precursor ion intensity based quantification is known to be less accurate and reproducible compared to other strategies such as multiple reaction monitoring (MRM) or iTRAQ labeling, it is known to provide the deepest proteome coverage⁷⁸. In the field of glycoproteomics, it is commonly used as compared to the other strategies which will be detailed in the following paragraph.

Fragment ion intensity-based quantification methods usually incorporate isobaric tags (iTRAQ or TMT). Typical iTRAQ or TMT reagents are composed of three groups- a reporter ion group, a mass balance group, and an amine-reactive group. Owing to the isobaric nature, the differentially labeled peptides appear as one peak in the full MS scan. Upon fragmentation, the reporter ions are readily cleaved from the reagent-peptide complex and relative quantification information can be obtained by comparing their intensities. An illustration of how isobaric 6-plex TMT labeling works is shown in Figure 1.5⁷⁹. A recent publication reported the application of TMT in intact glycopeptide analysis for quantification of site-specific glycosylation⁷⁶. However, this work observed very low-intensity reporter ions with HCD caused by the preference in lower-energy glycosidic bond cleavage rather than amide bond cleavage which produces reporter ions. Furthermore, iTRAQ labeling is generally believed to be inferior in characterizing low abundance peptides⁷⁸.

MRM is a highly-specific quantification strategy based on the intensities of

transitions, where a transition is defined as a pair of a precursor ion m/z and its fragment ion m/z . MRM is usually performed in a triple quadrupole instrument where the first quadrupole only allows a designated precursor ion m/z to pass, the second quadrupole fragments the precursor ion, and the third quadrupole only passes selected fragment ion m/z as shown in Figure 1.6⁸⁰. Compared to precursor ion based quantification, the sensitivity and specificity of MRM is greatly enhanced. Normally, commercial isotopic peptide standards are utilized for selection of fragment ions for the transitions and optimization of collision conditions. In the glycopeptide (both intact and partially deglycosylated) study, such glycopeptides standards are not available. Due to the prevalence of glycosidic cleavage in CID of glycopeptides, oxonium ions are commonly used as transitions rather than peptide backbone fragments, which greatly reduces the specificity of the MRM assay⁸¹⁻⁸³.

Due to the aforementioned limitation of iTRAQ/TMT and MRM strategies, precursor ion intensity based quantification is more often used for measuring site-specific glycosylation^{72, 84, 85}.

1.5 Outline of the dissertation

This dissertation consists of four research chapters which describe mass spectrometric analytical strategies for qualitative and quantitative analysis of protein N-glycosylation at the glycan, intact glycopeptide and partially deglycosylated glycopeptide levels as shown in Figure 1.7. The studies highlight the applications of mass spectrometric assays in discovery of glycosylation aberrations in human serum as potential pancreatic cancer biomarkers, with minimal sample consumption.

Chapter 2 describes a MALDI-MS assay which incorporates exoglycosidase digestion and permethylation to characterize haptoglobin N-glycans with the quantification information obtained based on peak area integration. A fucosylation-degree index was utilized to measure the extent of fucosylation in haptoglobin and the indices were compared between disease states including normal control, chronic pancreatitis, type II diabetes and pancreatic cancer. Fucosylation of haptoglobin was found to increase in pancreatic cancer. This study provides sensitive profiling of fucosylated glycans both qualitatively and quantitatively, and may prove useful in cancer biomarker study. The work was published as detailed in Reference 44 and 51.

Chapter 3 describes characterization of individual protein site-specific glycosylation by combining ETD/CID MS/MS to analyze the intact glycopeptides and CID MS/MS or MS³ to analyze endoglycosidase treated glycopeptides. The assay was used to comprehensively analyze alpha-2-macroglobin (A2MG) N-glycosylation at glycan, glycopeptide and peptide levels, and should have a wide application in study of site-specific glycosylation, particularly core-fucosylation. The work was published as detailed in Reference 62.

Chapter 4 extends the peptide-level work of Chapter 3 to the quantitative level, where the site-specific core-fucosylation of A2MG was quantified in a label-free manner using precursor ion intensity. The core-fucosylation ratios of A2MG were obtained at various glycosylation sites and compared between normal controls, chronic pancreatitis and pancreatic cancer. Core-fucosylation at three sites was found to decrease in both pancreatic cancer and chronic pancreatitis. This analytical assay is robust and straightforward, allowing probing of site-specific glycosylation change which may be a

novel cancer biomarker. Furthermore, the alteration of A2MG core-fucosylation in pancreatic cancer may be used in conjunction with other pancreatic cancer biomarkers to improve the accuracy of the diagnostics.

Chapter 5 applies the endoglycosidase-assisted strategy to large-scale serum core-fucosylation profiling. Effective optimization and combination of multiple experiment steps enables the most comprehensive profiling of the serum core-fucosylation. Furthermore, the potential of combining iTRAQ labeling to the workflow for discovery of pancreatic cancer-related core-fucosylation aberrations was explored. This high-throughput assay allows simultaneous screening of core-fucosylation pattern changes of hundreds of proteins, which should be valuable for cancer biomarker discovery purposes.

Figures

Figure 1. 1. Legends of monosaccharides of human N-glycans.

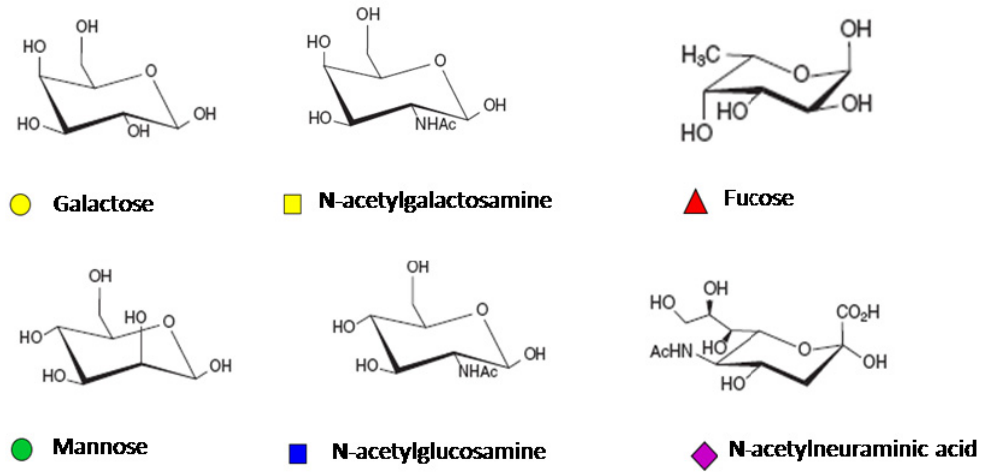


Figure 1. 2. Three types of N-glycans.

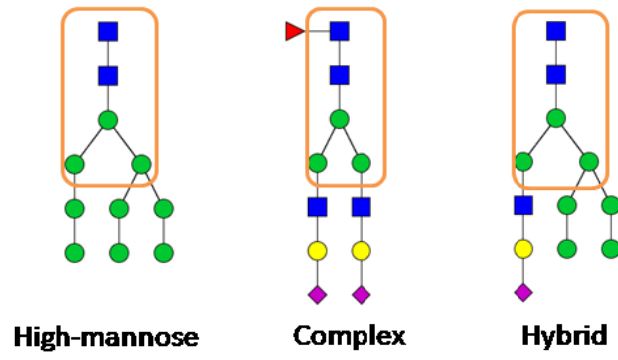
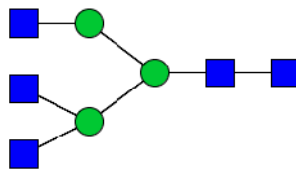
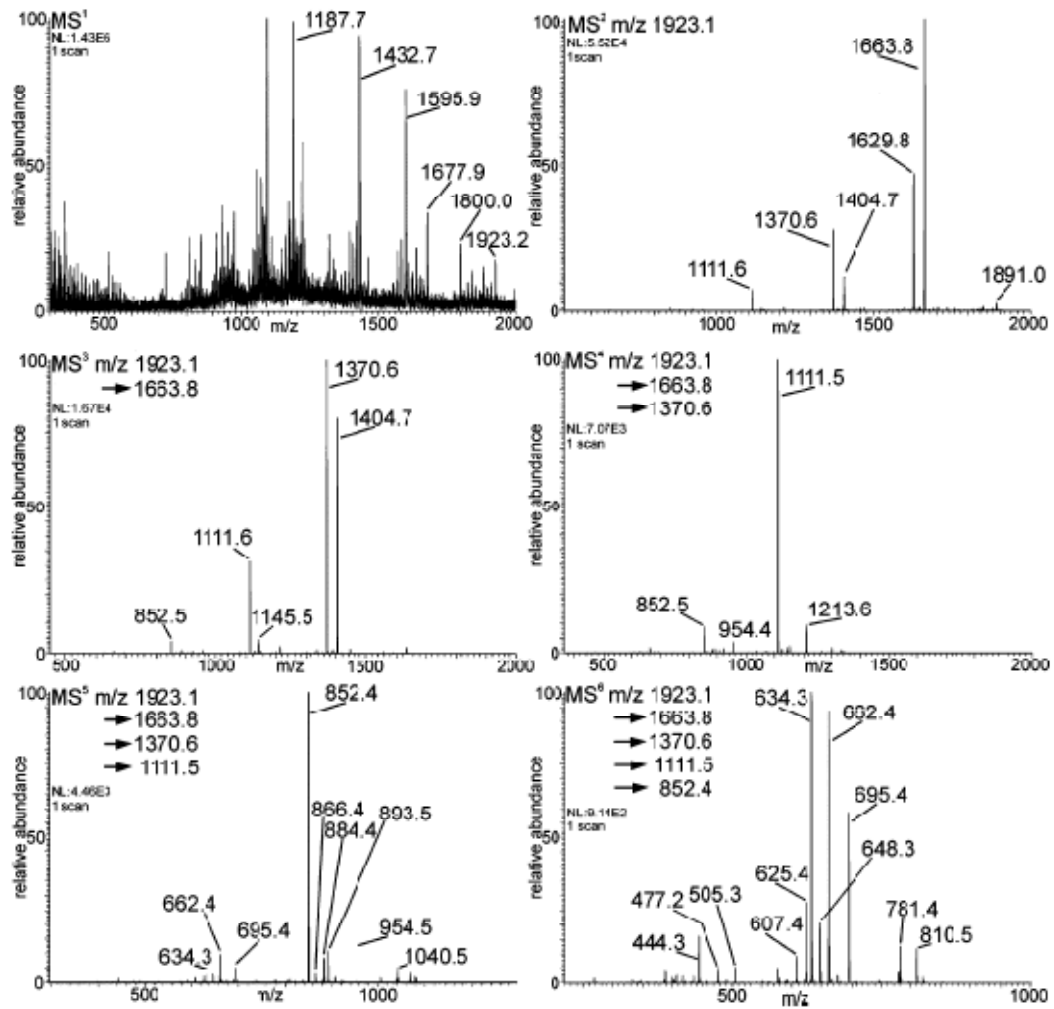
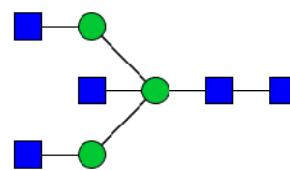


Figure 1. 3. MSⁿ spectra of an ovalbumin glycan at m/z 1923.2 with composition of Man₃GlcNAc₅. The diagnostic ion of m/z 893.5 in MS⁵ confirms the isomer structure B. (Reprinted with permission from (49). Copyright (2007) American Chemical Society)



Structure A



Structure B

Figure 1. 4. HCD-triggered-ETD analysis of a ribonuclease B glycopeptide. (a) MS survey scan, (b) HCD MS/MS spectrum of precursor ion at m/z 645.6194, and (c) ETD-MS/MS spectrum of precursor ion at m/z 645.6194. (Reprinted with permission from (66). Copyright (2012) American Chemical Society)

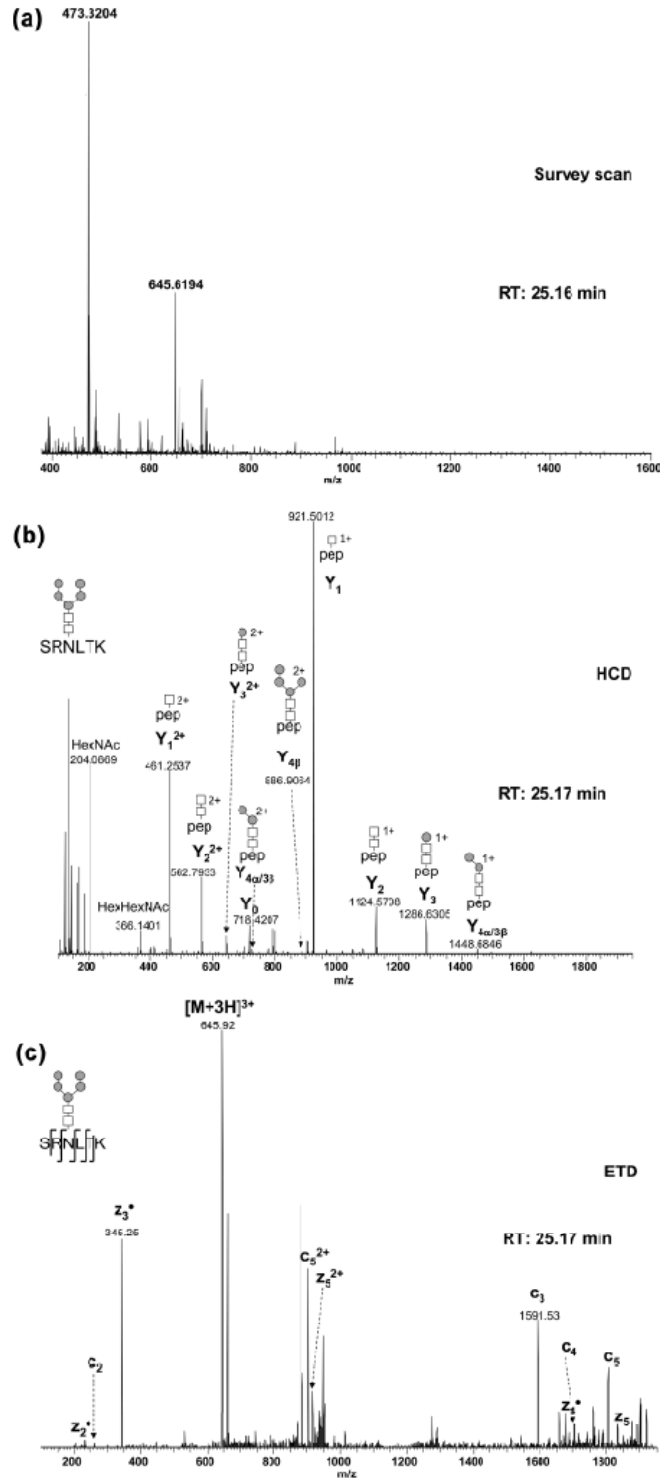


Figure 1.5. Quantitative mass spectrometric analysis with isobaric TMT labeling. Differentially labeled peptides are not distinguishable in MS¹, MS² generates reporter ions with different intensities on which quantification is based. (Reprinted with permission from (78). Copyright (2011) Nature Publishing Group)

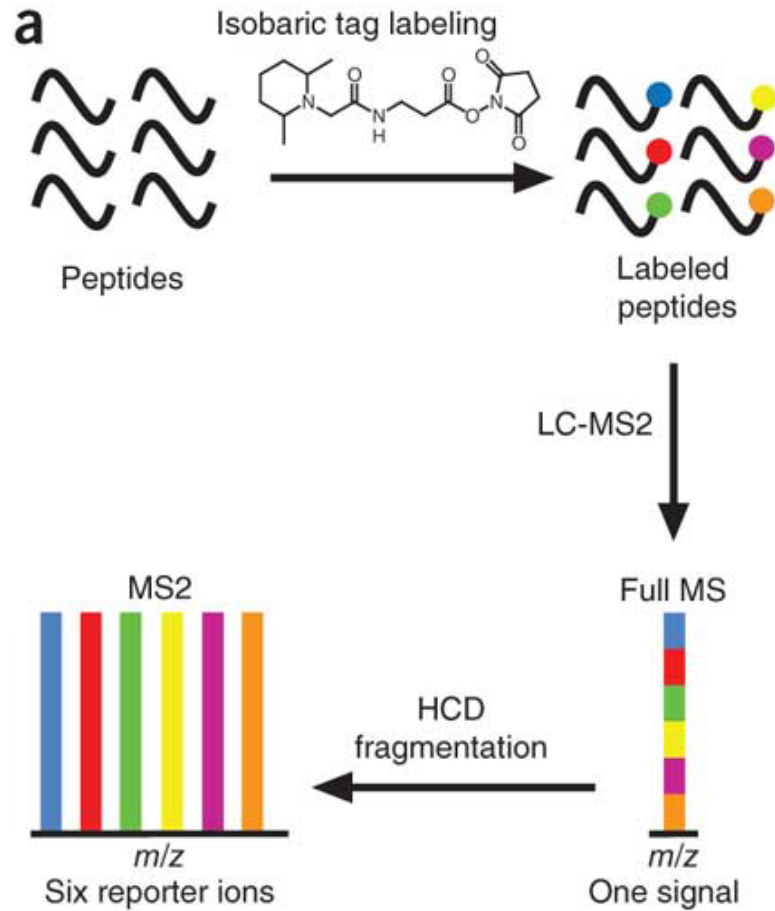


Figure 1. 6. In an MRM-MS analysis, precursor ions are preselected in Q1 and fragmented in Q2. Predefined fragment ions are filtered in Q3. Synthetic peptides are labeled with a stable isotopic tag as shown with asterisks and spiked in the sample to improve the precision of relative quantitation. (Reprinted with permission from (79). Copyright (2013) Nature Publishing Group)

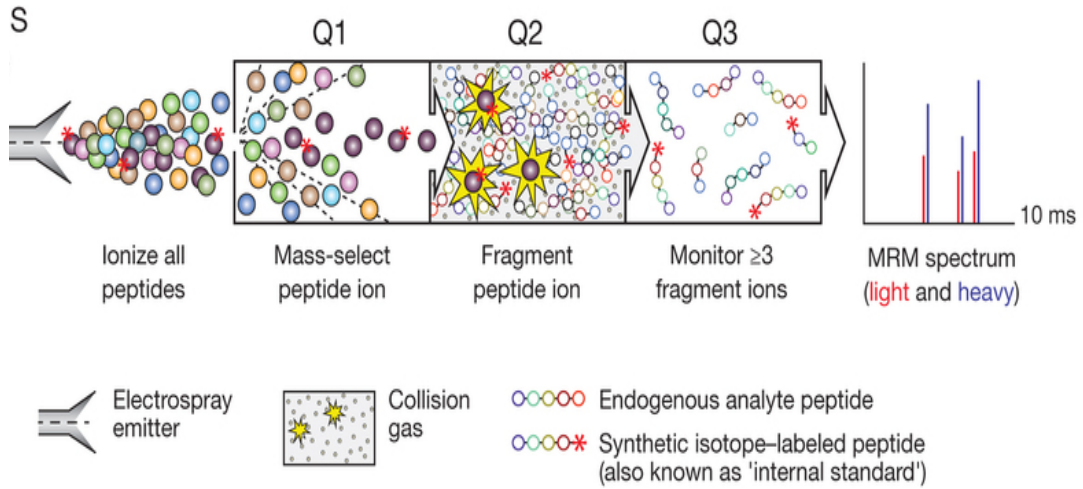
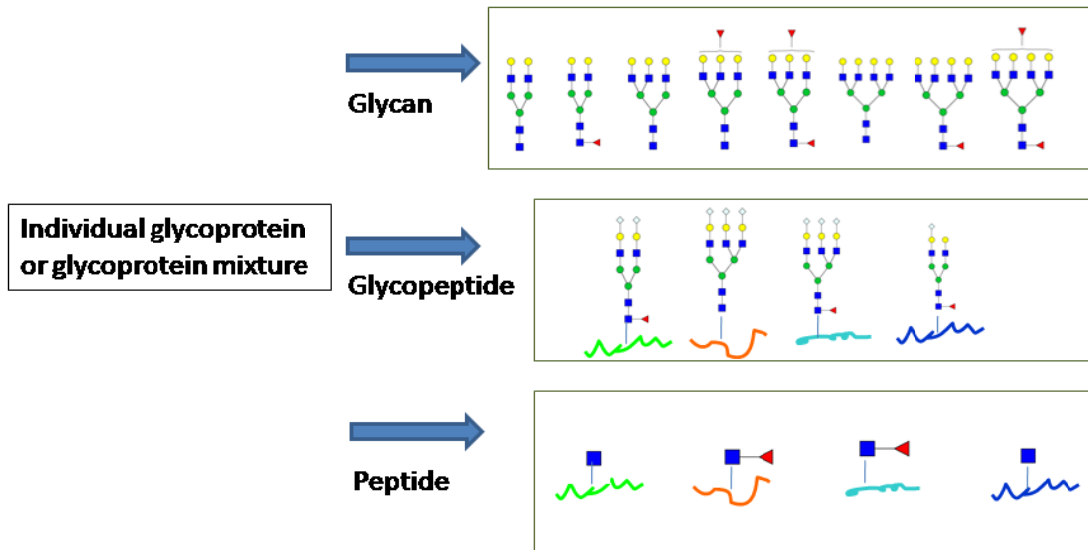


Figure 1. 7. Research strategies of the work discussed in this dissertation.



References

- (1) Hidalgo, M., Pancreatic Cancer. *New Engl. J. Med.* **2010**, 362, (17), 1605-1617.
- (2) Vincent, A.; Herman, J.; Schulick, R.; Hruban, R. H.; Goggins, M., Pancreatic cancer. *Lancet* **2011**, 378, (9791), 607-620.
- (3) Li, D. H.; Xie, K. P.; Wolff, R.; Abbruzzese, J. L., Pancreatic cancer. *Lancet* **2004**, 363, (9414), 1049-1057.
- (4) Kaur, S.; Baine, M. J.; Jain, M.; Sasson, A. R.; Batra, S. K., Early diagnosis of pancreatic cancer: challenges and new developments. *Biomarkers in Medicine* **2012**, 6, (5), 597-612.
- (5) Winter, J. M.; Yeo, C. J.; Brody, J. R., Diagnostic, prognostic, and predictive biomarkers in pancreatic cancer. *J. Surg. Oncol.* **2013**, 107, (1), 15-22.
- (6) Harsha, H. C.; Kandasamy, K.; Ranganathan, P.; Rani, S.; Ramabadrnan, S.; Gollapudi, S.; Balakrishnan, L.; Dwivedi, S. B.; Telikicherla, D.; Selvan, L. D. N.; Goel, R.; Mathivanan, S.; Marimuthu, A.; Kashyap, M.; Vizza, R. F.; Mayer, R. J.; DeCaprio, J. A.; Srivastava, S.; Hanash, S. M.; Hruban, R. H.; Pandey, A., A Compendium of Potential Biomarkers of Pancreatic Cancer. *Plos Medicine* **2009**, 6, (4).
- (7) Faca, V. M.; Song, K. S.; Wang, H.; Zhang, Q.; Krasnoselsky, A. L.; Newcomb, L. F.; Plentz, R. R.; Gurusurthy, S.; Redston, M. S.; Pitteri, S. J.; Pereira-Faca, S. R.; Ireton, R. C.; Katayama, H.; Glukhova, V.; Phanstiel, D.; Brenner, D. E.; Anderson, M. A.; Misek, D.; Scholler, N.; Urban, N. D.; Barnett, M. J.; Edelstein, C.; Goodman, G. E.; Thornquist, M. D.; McIntosh, M. W.; DePinho, R. A.; Bardeesy, N.; Hanash, S. M., A mouse to human search for plasma Proteome changes associated with pancreatic tumor development. *Plos Medicine* **2008**, 5, (6), 953-967.
- (8) Cho, W. C. S., Contribution of oncoproteomics to cancer biomarker discovery. *Molecular Cancer* **2007**, 6.
- (9) Chen, R.; Yi, E. C.; Donohoe, S.; Pan, S.; Eng, J.; Cooke, K.; Crispin, D. A.; Lane, Z. L.; Goodlett, D. R.; Bronner, M. P.; Aebersold, R.; Brentnall, T. A., Pancreatic cancer proteome: The proteins that underlie invasion, metastasis, and immunologic escape. *Gastroenterology* **2005**, 129, (4), 1187-1197.
- (10) Chen, R.; Pan, S.; Brentnall, T. A.; Aebersold, R., Proteomic profiling of pancreatic cancer for biomarker discovery. *Molecular & Cellular Proteomics* **2005**, 4, (4), 523-533.
- (11) Bailey, U.-M.; Jamaluddin, M. F.; Schulz, B. L., Analysis of Congenital Disorder of Glycosylation-Id in a Yeast Model System Shows Diverse Site-Specific Under-glycosylation of Glycoproteins. *J Proteome Res* **2012**, 11, (11), 5376-5383.
- (12) Walsh, G.; Jefferis, R., Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* **2006**, 24, (10), 1241-1252.
- (13) Wang, H.; Wong, C.-H.; Chin, A.; Taguchi, A.; Taylor, A.; Hanash, S.; Sekiya, S.; Takahashi, H.; Murase, M.; Kajihara, S.; Iwamoto, S.; Tanaka, K., Integrated mass spectrometry-based analysis of plasma glycoproteins and their glycan modifications. *Nature Protocols* **2011**, 6, (3).
- (14) Goldberg, D.; Sutton-Smith, M.; Paulson, J.; Dell, A., Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. *Proteomics* **2005**, 5, (4), 865-875.
- (15) Ruhaak, L. R.; Miyamoto, S.; Lebrilla, C. B., Developments in the Identification of Glycan Biomarkers for the Detection of Cancer. *Molecular & Cellular Proteomics* **2013**, 12, (4), 846-855.

- (16) Helenius, A.; Aebi, M., Intracellular functions of N-linked glycans. *Science* **2001**, 291, (5512), 2364-2369.
- (17) Nilsson, J.; Halim, A.; Grahn, A.; Larson, G., Targeting the glycoproteome. *Glycoconjugate J.* **2013**, 30, (2), 119-136.
- (18) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A., Glycosylation and the immune system. *Science* **2001**, 291, (5512), 2370-2376.
- (19) Alley, W. R.; Vasseur, J. A.; Goetz, J. A.; Syoboda, M.; Mann, B. F.; Matei, D. E.; Menning, N.; Hussein, A.; Mechref, Y.; Novotny, M. V., N-linked Glycan Structures and Their Expressions Change in the Blood Sera of Ovarian Cancer Patients. *J Proteome Res* **2012**, 11, (4), 2282-2300.
- (20) de Leoz, M. L. A.; Young, L. J. T.; An, H. J.; Kronewitter, S. R.; Kim, J.; Miyamoto, S.; Borowsky, A. D.; Chew, H. K.; Lebrilla, C. B., High-Mannose Glycans are Elevated during Breast Cancer Progression. *Molecular & Cellular Proteomics* **2011**, 10, (1).
- (21) Kyselova, Z.; Mechref, Y.; Al Bataineh, M. M.; Dobrolecki, L. E.; Hickey, R. J.; Vinson, J.; Sweeney, C. J.; Novotny, M. V., Alterations in the serum glycome due to metastatic prostate cancer. *J Proteome Res* **2007**, 6, (5), 1822-1832.
- (22) Zhao, J.; Simeone, D. M.; Heidt, D.; Anderson, M. A.; Lubman, D. M., Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: Application to pancreatic cancer serum. *J Proteome Res* **2006**, 5, (7), 1792-1802.
- (23) Hedlund, M.; Ng, E.; Varki, A.; Varki, N. M., alpha 2-6-linked sialic acids on N-glycans modulate carcinoma differentiation in vivo. *Cancer Res.* **2008**, 68, (2), 388-394.
- (24) Saldova, R.; Fan, Y.; Fitzpatrick, J. M.; Watson, R. W. G.; Rudd, P. M., Core fucosylation and alpha 2-3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia. *Glycobiology* **2011**, 21, (2), 195-205.
- (25) Wang, M.; Long, R. E.; Comunale, M. A.; Junaidi, O.; Marrero, J.; Di Bisceglie, A. M.; Block, T. M.; Mehta, A. S., Novel Fucosylated Biomarkers for the Early Detection of Hepatocellular Carcinoma. *Cancer Epidemiology Biomarkers & Prevention* **2009**, 18, (6), 1914-1921.
- (26) Qiu, Y.; Patwa, T. H.; Xu, L.; Shedden, K.; Misek, D. E.; Tuck, M.; Jin, G.; Ruffin, M. T.; Turgeon, D. K.; Synal, S.; Bresalier, R.; Marcon, N.; Brenner, D. E.; Lubman, D. M., Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. *J Proteome Res* **2008**, 7, (4), 1693-1703.
- (27) Li, C.; Simeone, D. M.; Brenner, D. E.; Anderson, M. A.; Shedden, K. A.; Ruffin, M. T.; Lubman, D. M., Pancreatic Cancer Serum Detection Using a Lectin/Glyco-Antibody Array Method. *J Proteome Res* **2009**, 8, (2), 483-492.
- (28) Wang, L.; Aryal, U. K.; Dai, Z.; Mason, A. C.; Monroe, M. E.; Tian, Z.-X.; Zhou, J.-Y.; Su, D.; Weitz, K. K.; Liu, T.; Camp, D. G., II; Smith, R. D.; Baker, S. E.; Qian, W.-J., Mapping N-Linked Glycosylation Sites in the Secretome and Whole Cells of *Aspergillus niger* Using Hydrazide Chemistry and Mass Spectrometry. *J Proteome Res* **2012**, 11, (1), 143-156.
- (29) Malerod, H.; Graham, R. L. J.; Sweredoski, M. J.; Hess, S., Comprehensive Profiling of N-Linked Glycosylation Sites in HeLa Cells Using Hydrazide Enrichment. *J Proteome Res* **2013**, 12, (1), 337-348.

- (30) Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R., Chemical Deamidation: A Common Pitfall in Large-Scale N-Linked Glycoproteomic Mass Spectrometry-Based Analyses. *J Proteome Res* **2012**, 11, (3), 1949-1957.
- (31) Hagglund, P.; Bunkenborg, J.; Elortza, F.; Jensen, O. N.; Roepstorff, P., A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* **2004**, 3, (3), 556-566.
- (32) Yang, Z.; Harris, L. E.; Palmer-Toy, D. E.; Hancock, W. S., Multilectin affinity chromatography for characterization of multiple glycoprotein biomarker candidates in serum from breast cancer patients. *Clin. Chem.* **2006**, 52, (10), 1897-1905.
- (33) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **2003**, 21, (6), 660-666.
- (34) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D., Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J Proteome Res* **2005**, 4, (6), 2070-2080.
- (35) Wada, Y.; Tajiri, M.; Yoshida, S., Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. *Anal Chem* **2004**, 76, (22), 6560-6565.
- (36) Nakano, M.; Nakagawa, T.; Ito, T.; Kitada, T.; Hijioka, T.; Kasahara, A.; Tajiri, M.; Wada, Y.; Taniguchi, N.; Miyoshi, E., Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: A novel approach for the development of tumor markers. *Int J Cancer* **2008**, 122, (10), 2301-2309.
- (37) Yeh, C. H.; Chen, S. H.; Li, D. T.; Lin, H. P.; Huang, H. J.; Chang, C. I.; Shih, W. L.; Chern, C. L.; Shi, F. K.; Hsu, J. L., Magnetic bead-based hydrophilic interaction liquid chromatography for glycopeptide enrichments. *J. Chromatogr.* **2012**, 1224, 70-78.
- (38) Scott, N. E.; Parker, B. L.; Connolly, A. M.; Paulech, J.; Edwards, A. V. G.; Crossett, B.; Falconer, L.; Kolarich, D.; Djordjevic, S. P.; Hojrup, P.; Packer, N. H.; Larsen, M. R.; Cordwell, S. J., Simultaneous Glycan-Peptide Characterization Using Hydrophilic Interaction Chromatography and Parallel Fragmentation by CID, Higher Energy Collisional Dissociation, and Electron Transfer Dissociation MS Applied to the N-Linked Glycoproteome of *Campylobacter jejuni*. *Molecular & Cellular Proteomics* **2011**, 10, (2).
- (39) An, H. J.; Gip, P.; Kim, J.; Wu, S.; Park, K. W.; McVaugh, C. T.; Schaffer, D. V.; Bertozzi, C. R.; Lebrilla, C. B., Extensive determination of glycan heterogeneity reveals an unusual abundance of high mannose glycans in enriched plasma membranes of human embryonic stem cells. *Mol. Cell. Proteomics* **2012**, 11, (4), M111 010660.
- (40) Jensen, P. H.; Karlsson, N. G.; Kolarich, D.; Packer, N. H., Structural analysis of N- and O-glycans released from glycoproteins. *Nature Protocols* **2012**, 7, (7), 1299-1310.
- (41) Selman, M. H. J.; Hemayatkar, M.; Deelder, A. M.; Wuhrer, M., Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Anal Chem* **2011**, 83, (7), 2492-2499.
- (42) Ruhaak, L. R.; Zauner, G.; Huhn, C.; Bruggink, C.; Deelder, A. M.; Wuhrer, M., Glycan labeling strategies and their use in identification and quantification. *Anal. Bioanal. Chem.* **2010**, 397, (8), 3457-3481.
- (43) Pabst, M.; Altmann, F., Glycan analysis by modern instrumental methods. *Proteomics* **2011**, 11, (4), 631-643.

- (44) Lin, Z.; Lubman, D. M., Permethylated N-glycan analysis with mass spectrometry. *Methods in molecular biology (Clifton, N.J.)* **2013**, 1007, 289-300.
- (45) Mechref, Y.; Novotny, M. V.; Krishnan, C., Structural characterization of oligosaccharides using MALDI-TOF/TOF tandem mass spectrometry. *Anal Chem* **2003**, 75, (18), 4895-4903.
- (46) Tang, H. X.; Mechref, Y.; Novotny, M. V., Automated interpretation of MS/MS spectra of oligosaccharides. *Bioinformatics* **2005**, 21, I431-I439.
- (47) Morelle, W.; Slomianny, M. C.; Diemer, H.; Schaeffer, C.; van Dorsselaer, A.; Michalski, J. C., Fragmentation characteristics of permethylated oligosaccharides using a matrix-assisted laser desorption/ionization two-stage time-of-flight (TOF/TOF) tandem mass spectrometer. *Rapid Commun Mass Spectrom* **2004**, 18, (22), 2637-2649.
- (48) Weiskopf, A. S.; Vouros, P.; Harvey, D. J., Characterization of oligosaccharide composition and structure by quadrupole ion trap mass spectrometry. *Rapid Commun Mass Spectrom* **1997**, 11, (14), 1493-1504.
- (49) Ashline, D. J.; Lapadula, A. J.; Liu, Y. H.; Lin, M.; Grace, M.; Pramanik, B.; Reinhold, V. N., Carbohydrate structural isomers analyzed by sequential mass spectrometry. *Anal Chem* **2007**, 79, (10), 3830-3842.
- (50) Harvey, D. J.; Martin, R. L.; Jackson, K. A.; Sutton, C. W., Fragmentation of N-linked glycans with a matrix-assisted laser desorption/ionization ion trap time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* **2004**, 18, (24), 2997-3007.
- (51) Lin, Z.; Simeone, D. M.; Anderson, M. A.; Brand, R. E.; Xie, X.; Shedden, K. A.; Ruffin, M. T.; Lubman, D. M., Mass Spectrometric Assay for Analysis of Haptoglobin Fucosylation in Pancreatic Cancer. *J Proteome Res* **2011**, 10, (5), 2602-2611.
- (52) Alvarez-Manilla, G.; Warren, N. L.; Abney, T.; Atwood, J., III; Azadi, P.; York, W. S.; Pierce, M.; Orlando, R., Tools for glycomics: relative quantitation of glycans by isotopic permethylation using (CH₃I)-C-13. *Glycobiology* **2007**, 17, (7), 677-687.
- (53) Mechref, Y.; Hu, Y.; Desantos-Garcia, J. L.; Hussein, A.; Tang, H., Quantitative Glycomics Strategies. *Molecular & Cellular Proteomics* **2013**, 12, (4), 874-884.
- (54) Higai, K.; Aoki, Y.; Azuma, Y.; Matsumoto, K., Glycosylation of site-specific glycans of alpha(1)-acid glycoprotein and alterations in acute and chronic inflammation. *Biochimica Et Biophysica Acta-General Subjects* **2005**, 1725, (1), 128-135.
- (55) Kolarich, D.; Jensen, P. H.; Altmann, F.; Packer, N. H., Determination of site-specific glycan heterogeneity on glycoproteins. *Nature Protocols* **2012**, 7, (7), 1285-1298.
- (56) Wang, D. D.; Hincapie, M.; Rejtar, T.; Karger, B. L., Ultrasensitive Characterization of Site-Specific Glycosylation of Affinity-Purified Haptoglobin from Lung Cancer Patient Plasma Using 10 μ m i.d. Porous Layer Open Tubular Liquid Chromatography-Linear Ion Trap Collision-Induced Dissociation/Electron Transfer Dissociation Mass Spectrometry. *Anal Chem* **2011**, 83, (6), 2029-2037.
- (57) Hua, S.; Nwosu, C. C.; Strum, J. S.; Seipert, R. R.; An, H. J.; Zivkovic, A. M.; German, J. B.; Lebrilla, C. B., Site-specific protein glycosylation analysis with glycan isomer differentiation. *Anal. Bioanal. Chem.* **2012**, 403, (5), 1291-302.
- (58) Pompach, P.; Brnakova, Z.; Sanda, M.; Wu, J.; Edwards, N.; Goldman, R., Site-specific Glycoforms of Haptoglobin in Liver Cirrhosis and Hepatocellular Carcinoma. *Molecular & cellular proteomics : MCP* **2013**, 12, (5), 1281-93.
- (59) Mikesch, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F., The utility of ETD mass spectrometry in proteomic analysis. *Biochim*

- Biophys Acta, Proteins Proteomics* **2006**, 1764, (12), 1811-1822.
- (60) Mechref, Y., Use of CID/ETD mass spectrometry to analyze glycopeptides. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **2012**, Chapter 12, Unit 12.11.1-11.
- (61) Alley, W. R., Jr.; Mechref, Y.; Novotny, M. V., Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom* **2009**, 23, (1), 161-170.
- (62) Lin, Z.; Lo, A.; Simeone, D. M.; Ruffin, M. T.; Lubman, D. M., An N-glycosylation Analysis of Human Alpha-2-Macroglobulin Using an Integrated Approach. *J Proteomics Bioinform* **2012**, 5, 127-134.
- (63) Darula, Z.; Medzihradsky, K. F., Affinity Enrichment and Characterization of Mucin Core-1 Type Glycopeptides from Bovine Serum. *Molecular & Cellular Proteomics* **2009**, 8, (11), 2515-2526.
- (64) Hart-Smith, G.; Raftery, M. J., Detection and Characterization of Low Abundance Glycopeptides Via Higher-Energy C-Trap Dissociation and Orbitrap Mass Analysis. *J. Am. Soc. Mass Spectrom.* **2012**, 23, (1), 124-140.
- (65) Saba, J.; Dutta, S.; Hemenway, E.; Viner, R., Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *International journal of proteomics* **2012**, 2012, 560391-560391.
- (66) Segu, Z. M.; Mechref, Y., Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun Mass Spectrom* **2010**, 24, (9), 1217-1225.
- (67) Singh, C.; Zampronio, C. G.; Creese, A. J.; Cooper, H. J., Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of N-Linked Glycoproteins. *J Proteome Res* **2012**, 11, (9), 4517-4525.
- (68) Pasing, Y.; Sickmann, A.; Lewandrowski, U., N-glycoproteomics: mass spectrometry-based glycosylation site annotation. *Biol. Chem.* **2012**, 393, (4), 249-258.
- (69) Segu, Z. M.; Hussein, A.; Novotny, M. V.; Mechref, Y., Assigning N-Glycosylation Sites of Glycoproteins Using LC/MSMS in Conjunction with Endo-M/Exoglycosidase Mixture. *J Proteome Res* **2010**, 9, (7), 3598-3607.
- (70) Zhao, Y.; Jia, W.; Wang, J. F.; Ying, W. T.; Zhang, Y. J.; Qian, X. H., Fragmentation and Site-Specific Quantification of Core Fucosylated Glycoprotein by Multiple Reaction Monitoring-Mass Spectrometry. *Anal Chem* **2011**, 83, (22), 8802-8809.
- (71) Hagglund, P.; Matthiesen, R.; Elortza, F.; Hojrup, P.; Roepstorff, P.; Jensen, O. N.; Bunkenborg, J., An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins. *J Proteome Res* **2007**, 6, (8), 3021-3031.
- (72) Chen, R.; Wang, F.; Tan, Y.; Sun, Z.; Song, C.; Ye, M.; Wang, H.; Zou, H., Development of a combined chemical and enzymatic approach for the mass spectrometric identification and quantification of aberrant N-glycosylation. *J Proteomics* **2012**, 75, (5), 1666-74.
- (73) Okuyama, N.; Ide, Y.; Nakano, M.; Nakagawa, T.; Yamanaka, K.; Moriwaki, K.; Murata, K.; Ohigashi, H.; Yokoyama, S.; Eguchi, H.; Ishikawa, O.; Ito, T.; Kato, M.; Kasahara, A.; Kawano, S.; Gu, J. G.; Miyoshi, E., Fucosylated haptoglobin is a novel

marker for pancreatic cancer: A detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. *Int J Cancer* **2006**, 118, (11), 2803-2808.

(74)Geng, F.; Shi, B. Z.; Yuan, Y. F.; Wu, X. Z., The expression of core fucosylated E-cadherin in cancer cells and lung cancer patients: prognostic implications. *Cell Res.* **2004**, 14, (5), 423-433.

(75)Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B., Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics* **2007**, 7, (3), 340-350.

(76)Ye, H.; Boyne, M. T., II; Buhse, L. F.; Hill, J., Direct Approach for Qualitative and Quantitative Characterization of Glycoproteins Using Tandem Mass Tags and an LTQ Orbitrap XL Electron Transfer Dissociation Hybrid Mass Spectrometer. *Anal Chem* **2013**, 85, (3), 1531-1539.

(77)Taga, Y.; Kusubata, M.; Ogawa-Goto, K.; Hattori, S., Site-specific Quantitative Analysis of Overglycosylation of Collagen in Osteogenesis Imperfecta Using Hydrazide Chemistry and SILAC. *J Proteome Res* **2013**, 12, (5), 2225-32.

(78)Li, Z.; Adams, R. M.; Chourey, K.; Hurst, G. B.; Hettich, R. L.; Pan, C. L., Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos. *J Proteome Res* **2012**, 11, (3), 1582-1590.

(79)Ting, L.; Rad, R.; Gygi, S. P.; Haas, W., MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **2011**, 8, (11), 937-940.

(80)Gillette, M. A.; Carr, S. A., METHOD OF THE YEAR Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* **2013**, 10, (1), 28-34.

(81)Kuroguchi, M.; Matsushita, T.; Amano, M.; Furukawa, J.; Shinohara, Y.; Aoshima, M.; Nishimura, S. I., Sialic Acid-focused Quantitative Mouse Serum Glycoproteomics by Multiple Reaction Monitoring Assay. *Molecular & Cellular Proteomics* **2010**, 9, (11), 2354-2368.

(82)Song, E.; Pyreddy, S.; Mechref, Y., Quantification of glycopeptides by multiple reaction monitoring liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* **2012**, 26, (17), 1941-1954.

(83)Toyama, A.; Nakagawa, H.; Matsuda, K.; Sato, T.-A.; Nakamura, Y.; Ueda, K., Quantitative Structural Characterization of Local N-Glycan Microheterogeneity in Therapeutic Antibodies by Energy-Resolved Oxonium Ion Monitoring. *Anal Chem* **2012**, 84, (22), 9655-9662.

(84)Ivancic, M. M.; Gadgil, H. S.; Halsall, H. B.; Treuheit, M. J., LC/MS analysis of complex multiglycosylated human alpha(1)-acid glycoprotein as a model for developing identification and quantitation methods for intact glycopeptide analysis. *Anal. Biochem.* **2010**, 400, (1), 25-32.

(85)Rebecchi, K. R.; Wenke, J. L.; Go, E. P.; Desaire, H., Label-Free Quantitation: A New Glycoproteomics Approach. *J. Am. Soc. Mass Spectrom.* **2009**, 20, (6), 1048-1059.

Chapter 2

A Mass Spectrometric Assay for Analysis of Haptoglobin Fucosylation in Pancreatic Cancer

2.1 Introduction

Pancreatic cancer is the fourth leading cause of cancer deaths in the United States with the worst prognosis among all cancers. One of the causes of poor prognosis is the lack of a reliable early-detection method of the disease. Currently, the most widely used serum-based marker is CA 19-9 whose diagnostic value is limited because of a high false positive rate, further it does not allow early detection and can not readily discriminate between chronic pancreatitis and pancreatic cancer.^{1,2} Thus, there is an urgent need for reliable noninvasive or low-invasive methods for early detection of pancreatic cancer.

Unique protein glycosylation patterns have recently been explored as a potential target for cancer biomarker detection. Distinctive serum glycomic patterns have been reported to be associated with various types of cancers and other malignancies.³⁻⁷ Unlike proteins which have genetic templates that determine the structures, glycans have greater variability which is determined by both genetic polymorphisms and the physiological environment of the cells.⁸ The microheterogeneity of glycans depends on activities of glycosidases and glycosyltransferases which are influenced by the physiological and pathological states of cells.⁹ In general, changes of branching, and alternation of levels of sialylation and fucosylation are the most common tumor-associated glycan aberrations.^{9,}
¹⁰ The most common types of fucosylation in human serum glycoproteins are core

fucosylation where fucoses attach to the core N-acetylglucosamine (GlcNAc) via α 1-6 linkage and antennary fucosylation where fucoses attach to terminal GlcNAc via α 1-3 or α 1-4 linkage. Fucosylation is controlled by fucosyltransferases capable of different types of linkages where it has been reported that the activity of fucosyltransferases are related to cancer progression.¹¹⁻¹³ The potential clinical utility of fucosylation changes as a cancer marker has been explored.^{14, 15}

Acute phase proteins (APP) secreted by the liver are attractive potential markers because they display changes in both protein levels and glycosylation modification in response to inflammation and diseases including cancer.¹⁶ Haptoglobin is one of the APPs which binds to liver hemoglobin and plays an important role in defense response to inflammation and infection.¹⁷ There has been increasing evidence that the glycosylation status of haptoglobin is associated with various cancers such as prostate cancer, ovarian cancer, liver cancer, and colon cancer.¹⁸⁻²² In particular, the elevation of haptoglobin fucosylation in pancreatic cancer using various methods such as lectin blot, lectin-antibody microarray and lectin-antibody ELISA has been shown.²³⁻²⁵ The structures of 2-aminopyridine-labeled N-glycans were elucidated with NP-HPLC and MALDI-TOF MS analysis.^{24, 26} Site-specific glycan analysis was also performed with LC-ESI-MS, revealing that fucosylated glycans are markedly increased at N211.²⁶ However, in their study, mass spectrometric analysis of haptoglobin glycans was performed in a qualitative instead of quantitative manner and was not performed on every individual sample.

In the current study, we have examined the unique fucosylation patterns of haptoglobin in serum samples obtained from patients with pancreatic cancer, chronic

pancreatitis, type II diabetes, and normal controls using a mass spectrometry based approach. A method has been developed to assay fucosylation using an antibody to extract the protein from serum followed by deglycosylation, desialylation and extraction of the N-glycan units. The N-glycans were then permethylated to increase the sensitivity of the assay. The MALDI-QIT MS was then used to study the fucosylation pattern of the glycans in MS and MS/MS modes for all stages of cancer versus chronic pancreatitis, type II diabetes and normal controls. We found that there were distinct changes in the level of both core and antennary fucosylation associated with all stages of cancer compared to the noncancerous samples. An index comparing the changes in fucosylated glycans has been developed and the results over a limited analytical test set show the potential of using haptoglobin fucosylation changes as a marker of disease state for pancreatic cancer.

2.2. Experimental Section

2.2.1 Serum samples.

Human normal serum (n=5, 3 females and 2 males, median age 59), chronic pancreatitis serum (n=5, 3 females and 2 males, median age 59), type II diabetes serum(n=5, 4 females and 1male, median age 58) and pancreatic cancer serum(n=16, 1 stage IA, 3 stage IIA, 4 stage IIB, 4 stage III, and 4 stage IV, 7 females and 9 males, median age 69) were provided by the University Hospital, Ann Arbor, Michigan and the University of Pittsburgh according to IRB approval.

The samples were aliquoted and stored in a -80°C freezer until further use. All samples were frozen and thawed only once except for the freeze-thaw study. In the

freeze-thaw study, serum from a pancreatic cancer IIB patient was aliquoted into three fractions, frozen in a -80°C freezer, and then thawed at room temperature. The freeze-thaw cycles were repeated for once, twice and four times, respectively.

2.2.2 Separation of haptoglobin from serum.

10uL of human serum was thawed and diluted to 250uL using coupling buffer (10mM sodium phosphate, 150mM sodium chloride, pH7.2) in Cross-link IP kit (Pierce Scientific, Rockford, IL). Ig G is the most abundant glycoprotein in serum with a concentration of 8-16 mg/mL and may interfere with protein A/G based immunoprecipitation. Hence, IgG was depleted prior to haptoglobin capture using Protein A/G agarose beads (Pierce Scientific, Rockford, IL). In this procedure, 100uL bead slurry was incubated with diluted serum samples at 4 °C for 3 hours in a 900uL spin column on an end-to-end rotator. The depleted serum dilution was spun down in a centrifuge at 1000×g for 1 minute. The beads were washed once with 100uL coupling buffer.

Immunoprecipitation was performed using the Cross-link IP kit according to the supplier protocol. Briefly, 10ug monoclonal haptoglobin antibody (Abcam, Cambridge, MA) was bound to 20uL protein A/G plus agarose slurry at room temperature for 30 minutes, and cross-linked with the beads by 1×disuccinimidyl suberate (DSS) crosslinker at room temperature for 30 minutes. Unbound antibody that was not cross-linked was removed by extensive washing with coupling buffer and elution buffer (100mM Glycine-HCl, pH 2.8) respectively. The antibody-conjugated beads were then incubated with IgG depleted serum at 4 °C overnight and elution was carried out with 60uL elution buffer. The eluted haptoglobin was dried down in a SpeedVac concentrator (Labconco, Kansas city, MO) at room temperature, redissolved in 10uL water, and desalted by 75uL

Zeba desalting spin columns (Pierce Scientific, Rockford, IL) according to the protocol supplied.

Fast on-plate digestion and mass spectrometric analysis were performed for identification of haptoglobin. 0.5uL desalted haptoglobin was spotted on a MALDI plate, and dried in air. 0.4ug (1uL) trypsin (Promega, Madison, WI) was added to 10uL 100mM ammonium bicarbonate solution with 20% acetonitrile. 0.5uL trypsin solution was deposited on top of the haptoglobin spot and the plate was placed in a covered humid chamber at 37 °C for 10 minutes. 10mg/mL 2,5-dihydroxybenzoic acid (DHB) (Laser Biolabs, France) was prepared in 50% acetonitrile and added on top of the dried spot. Mass spectrometric analysis was carried out using an Axima MALDI quadrupole ion trap-TOF instrument (Shimadzu Biotech, Manchester, UK). Ionization was performed with a pulsed N₂ laser (337nm) at 5HZ where two laser shots generated one profile. Helium was used to cool the trapped ions and Argon was used for CID fragmentation. MALDI spectra were recorded only in positive ion mode. The TOF detector was calibrated with 1nmol/uL peptide mixtures of Angiotensin II (m/z 1046.54), Angiotensin I (m/z 1296.68), Substance P (m/z 1347.74), Bombesin (m/z 1619.82), ACTH 1-17 (m/z 2093.09), and ACTH 18-39 (m/z 2465.20). The mass accuracy with calibration was 50ppm. The peptide peaks were searched in the Mascot database with methionine oxidation as the variable modification.

SDS-PAGE followed by silver staining was used to evaluate the yield of haptoglobin. 1/4 of the haptoglobin eluant was boiled for 3 minutes and separated by 4-20% precast gel (Bio-Rad, Hercules, CA) in a MINI-PROTEAN cell (Bio-Rad, Hercules, CA) at 120V supplied by Power Pac3000 (Bio-Rad, Hercules, CA). 5uL of the Kaleidoscope

protein marker (Bio-Rad, Hercules, CA) was used. Silver staining was performed according to the protocol provided by the manufacturer.

2.2.3 Deglycosylation and desialylation of haptoglobin.

The denaturing solution (0.2% SDS, 100mM 2-mercaptoethanol) was added to 10uL haptoglobin solution to make the final denature solution concentration 10%. The mixture was incubated in a 60 °C oven for 30 minutes. Ammonium bicarbonate solution was then added to make a final concentration of 15mM. 1U of N-glycosidase F (PNGase F) (New England Biolabs, Ipswich, MA) was added. Deglycosylation was performed at 37°C for 18 hours. PNGase F was deactivated by boiling for 5 minutes and the protein-glycan mixture was dried down in a SpeedVac and reconstituted in 30uL 20mM ammonium acetate solution. 40mU neuraminidase from *Clostridium perfringens* (Sigma Aldrich, St. Louis, MO) was added. Desialylation was performed at 37 °C for 20 hours.

2.2.4 Extraction of desialylated glycans.

Desialylated glycans and the protein mixture was dried in a SpeedVac and redissolved in 10uL water (with 0.1% TFA). 10uL porous graphitized carbon tips (PGC tips) (Sigma Aldrich, St. Louis, MO) were used to separate glycans from proteins and other impurities. The tip was activated by 50% acetonitrile (with 0.1% TFA) and equilibrated by water (with 0.1% TFA). The samples were then loaded and the tips were washed with water (with 0.1% TFA) to remove non-specific binding. 10uL 10% acetonitrile (with 0.1% TFA) and 10uL 25% acetonitrile (with 0.1% TFA) were used for glycan elution where the two elutions were combined and dried with the SpeedVac.

2.2.5 Permethylation of glycans.

Permethylation was performed according to the procedure of Kang in the literature.²⁷

3mg grounded sodium hydroxide powder was added to the glycans, and mixed with 20uL DMSO, 3.8uL methyl iodide and 0.2uL water at room temperature for 10 minutes. 24uL water and 24uL chloroform were added, and the chloroform phase was washed with 24uL water for 5 times. The water phase was discarded and the permethylated glycans were dried and redissolved in 2uL 20% acetonitrile for mass spectrometric analysis.

2.2.6 MALDI-QIT-TOF instrument

0.5uL of permethylated glycans were spotted on a MALDI plate and allowed to dry in air. 0.5uL of sodiated DHB (10mg/mL DHB in 50% acetonitrile with 100mM sodium chloride) was added on top. The parameters of the MALDI-QIT-TOF were the same as previously described. Glycomod tool (<http://www.expasy.org/tools/glycomod>) was utilized to predict the glycan composition. Only glycan structures included in GlycoSuite database (<http://glycosuitedb.expasy.org/glycosuite/glycodb>) were selected. The glycan compositions were further confirmed by MS/MS analysis. All glycans were sodiated and analyzed in positive ion mode in this study.

2.2.7 Data evaluation

The MALDI MS data were acquired and processed in Launch-pad software (Karatos, Manchester, UK). The m/z values and intensities were exported as ASCII files and plotted in SigmaPlot (San Jose, CA) and peak intensities were scaled with the highest peak as 100%. Glycan peak area integration was performed with Matlab (Natick, MA). The peak area of each glycan was the addition of both permethylated glycan peak and the most abundant underpermethylated glycan peak detected 14 Da lower than the fully permethylated peak. For data visualization, a column scatter plot of the calculated fucosylation index was generated with Prism (La Jolla, CA).

2.3 Results and Discussion

In our work, we sought to develop a mass spectrometric assay to identify and evaluate haptoglobin fucosylation patterns to discriminate pancreatic cancer samples from benign pancreatic diseases and normal controls. The work flow of this study is outlined in Figure 2.1. Briefly, haptoglobin was immunoprecipitated from IgG depleted human serum. On-plate digestion followed by MALDI-QIT MS peptide analysis was used to verify the success of immunoprecipitation. The haptoglobin yield was evaluated by SDS-PAGE analysis and silver staining. The haptoglobin was then deglycosylated and desialylated and the glycans were purified using the PGC tips and permethylated. Permethylated glycans were subject to mass spectrometric analysis for structural elucidation. Analysis of fucosylation degrees was performed by Matlab and visualized with Prism.

2.3.1 Purification of haptoglobin from human serum

Prior to immunoprecipitation with antibody-conjugated protein A/G agarose, depletion of IgG is required where IgG will bind to unoccupied protein A/G and coelute with the target protein, interfering with subsequent glycomic analysis. Haptoglobin is one of the plasma acute-phase proteins produced in liver with a molecular weight around 45kDa. It is a tetramer composed of α -1, α -2 and two β chains. Haptoglobin has four potential N-glycosylation sites at N184, N207, N211 and N241 which are all located in the β chain. In this study, haptoglobin was purified from the sera of 5 normal volunteers, 5 chronic pancreatitis patients, 5 type II diabetes patients, and 16 pancreatic cancer patients (1 Stage IA, 3 Stage IIA, 4 Stage IIB, 4 Stage III, and 4 Stage IV). 10-minute on-plate digestion and MALDI-QIT analysis were performed to confirm the identity and purity of haptoglobin. The mass spectrum as shown in Figure 2.2a was searched in the

Mascot database and returned human haptoglobin as the only significant protein with 13 matched peptides. SDS-PAGE followed by silver staining was used to visualize the yield of haptoglobin (Figure 2.2b). 1/4 of the eluent from immunoprecipitation was used for silver-staining analysis. Haptoglobin β chain (~ 42 kDa), α -2 chain (~ 18 kDa) and α -1 chain (~13 kDa) were all observed. As shown in Figure 2.2b, there was no contamination from other proteins in the eluted haptoglobin, and the total yield of haptoglobin β chain was estimated to be around 1-2ug per 10uL serum. The yield of immunoprecipitation was limited by the starting volume of serum and amount and efficiency of haptoglobin antibody. In this method, 1ug of protein is sufficient for subsequent glycan analysis.

2.3.2 N-glycan profiles of haptoglobin with and without desialylation reveal elevated fucosylation in pancreatic cancer

After purification of haptoglobin glycans, in-solution permethylation was performed. Permethylation stabilizes the relatively labile sialic acids and fucoses, and significantly improves the sensitivity and signal-over-noise ratio of glycans (comparison not shown), so that 10uL of serum aliquot and 1~2 ug of haptoglobin suffice for the identification of glycan structures.

Glycan structures of haptoglobin were first analyzed without desialylation. N-glycans were extracted from two 10uL pooled serum samples of 5 normal controls and 5 pancreatic cancers respectively. The representative spectra are shown in Figure 2.3a. All glycans observed from haptoglobin are complex-type. The major glycans are nonfucosylated mono- and disialylated bi-antennary structures for both normal controls and pancreatic cancer patients. Minor peaks corresponding to tri-antennary nonfucosylated glycans with one, two and three sialic acids are also observed in both

samples. The abundances of two bi-antennary fucosylated glycans and one tri-antennary fucosylated glycan with three sialic acids (m/z 2605.21, m/z 2966.50 and m/z 3776.89 respectively) are elevated in pancreatic cancer. More strikingly, tri-antennary fucosylated glycans with one and two sialic acids (m/z 3052.39 and m/z 3415.73) only appear in pancreatic cancer but not in normal samples. However, tetra-antennary glycans and bifucosylated glycans with molecular weight higher than 3800Da were not detected due to low abundances.

Since preliminary studies indicated that the fucosylation level is different between pancreatic cancer and normal controls, we cleaved the sialic acids from the glycans in order to: (1) merge glycans with sialic acid contents as the only difference into one peak so that sensitivity would be improved; and (2) eliminate the complicated heterogeneity of sialic acids so that the glycan spectrum and subsequent analysis would be simplified. However, it should be noted that glycan sialylation information which may serve as a potential cancer biomarker is missing with this approach.

A typical desialylated N-glycan profile of human haptoglobin from a normal control and a pancreatic cancer patient is shown in Figure 2.3b. The nonfucosylated biantennary complex type glycan (m/z 2070.10) was the most abundant structure as previously discussed. Eight glycan structures in total were identified as listed in Table 2.2.1. Compared with Figure 2.3a, the tri-antennary bifucosylated glycan (glycan 5) and three additional tetra-antennary glycans (glycan 6-8) were identified. We examined each glycan peak for structures which were increased or decreased in pancreatic cancer compared to healthy controls, type II diabetes patients and chronic pancreatitis patients. A zoom-in peak comparison illustrates that the fucosylated triantennary and tetrantennary structures

resulted in the best performance in discriminating pancreatic cancer from non-cancers. A representative zoom-in comparison of tri-antennary and tetra-antennary structures of normal control, type II diabetes, chronic pancreatitis and pancreatic cancer is shown in Figure 2.4. Both the fucosylated tri-antennary (m/z 2693.45) and tetra-antennary (m/z 3142.59) glycans were elevated in pancreatic cancer. The most striking finding was that bifucosylated tri-antennary (m/z 2867.46) and tetra-antennary (m/z 3316.60) glycans were present in 12 of 16 pancreatic cancer samples but never in non-cancerous samples. These two glycans may serve as possible markers for pancreatic cancer. The bifucosylated tetra-antennary glycan was reported to be unique at site N211 in pancreatic cancer in a site-specific study of haptoglobin by ESI-MS method.^{25,26} However, the tri-antennary glycan of human haptoglobin (m/z 2867.46) with both core and antennary fucosylation has not been reported previously with either the NP-HPLC or MALDI-TOF MS analysis^{16, 24} or ESI-MS approach for glycopeptide analysis²⁶. Our method detects this structure probably because desialylation reduces heterogeneity of glycans and improves sensitivity.

2.3.3 Fucosylation degree indices

In order to quantify the degree of fucosylation, Imre and coworkers developed a fucosylation index for α -1-acid glycoprotein N-glycans to discriminate between healthy, lymphoma and ovarian tumors.²⁸ The fucosylation index gives the average number of fucose units for a group of oligosaccharides. It is defined as:

$$\text{Fucosylation degree} = (1 \times \text{glycanF1} + 2 \times \text{glycanF2}) / \sum \text{glycans}$$

Where glycanF1 denotes the sum of abundances of singly-fucosylated glycans, glycanF2 denotes that of bifucosylated glycans, and $\sum \text{glycans}$ denotes the sum of abundances of all

glycans. In our study, we used this index to characterize the fucosylation level of tri-antennary and tetra-antennary glycans because they illustrated the greatest differences between pancreatic cancer and non-cancer in mass spectra depiction as discussed above. Hence the local fucosylation degree in our study is:

$$\text{Fucosylation degree} = (1 \times \text{glycan4} + 2 \times \text{glycan5} + 1 \times \text{glycan7} + 2 \times \text{glycan8}) / \sum \text{glycan3-8}$$

The abundance of each glycan was normalized by the sum of all glycan abundances. The overall fucosylation degree ranges from 0.06 to 0.95 for all samples. Non-cancer groups (normal, chronic pancreatitis and type II diabetes) have low fucosylation degree (mean is 0.27, 0.18 and 0.30 respectively), but fucosylation degrees are elevated in early stage pancreatic cancer (stage IIA and prior) with a mean of 0.40, and this index increases significantly for later stages (mean is 0.68, 0.63 and 0.64 for IIB, III and IV respectively). Fucosylation degree is generally lower in other pancreatic diseases such as chronic pancreatitis and in type II diabetes. A general comparison of fucosylation degrees between non-cancer and pancreatic cancer is shown in Figure 2.5. A student's t-test was also performed, indicating statistically significant differences between pancreatic cancer and non-cancer with a low p-value of 1.9×10^{-7} . If the cutoff value is set at 0.40, the sensitivity of the test is 94% while the specificity is 100% for this limited sample set. The result is that none of the benign samples was misidentified, however, one of the stage IIA cancer samples was below the index limit, as shown in Table 2.3 which summarizes numbers of correct classification with this fucosylation degree index. A high false positive rate is currently a major issue in diagnosis with the markers CA19-9 and CEA. With the use of haptoglobin fucosylation as a potential diagnostic marker, an individual with high fucosylation degree is likely to suffer from pancreatic cancer while pancreatic

cancer is unlikely for an individual with low fucosylation degree.

2.3.4 Reproducibility study and influence of freeze/thaw cycles

In order to utilize glycan structures as biomarkers for cancers and other malignancies, it is essential to develop a method for comprehensive, informative and quantitative glycan profile mapping (GPM). High-throughput methods such as lectin assay and antibody assay are able to quantify specific types of glycosylation patterns, but they are of low specificity for glycan epitopes, and detailed structural information is missing. HPLC or HPLC-ESI-MS methods provide both quantitative measurement and structure elucidation of glycans, but these existing approaches are time-consuming and do not have the potential for utilization in a high-throughput assay. MALDI (matrix-assisted laser desorption-ionization) MS has the attributes of simplicity in application and accuracy in composition assignment of glycans.²⁹ Moreover, application of permethylation procedure minimizes ion suppression by enabling glycans to have equivalent hydrophobic properties, thus quantitative power is enhanced. Permethylation also improves the reproducibility and sensitivity of glycan analysis,³⁰ and the spin-column based permethylation makes it more attractive in high-throughput analysis²⁷. According to a comparative study conducted by HUPO HGPI (Human Proteome Organization Human Disease Glycomics/Proteome Initiative) participated in by 20 research groups, MALDI-TOF MS analysis of permethylated glycans yields equivalent performance as the HPLC method with reductive amination.³¹ In this study, MALDI-TOF MS analysis of permethylated glycans was claimed to be highly repeatable and reproducible.

In order to evaluate the analytical reproducibility of our method and especially the repeatability of the fucosylation degree index, we processed four aliquots of a normal

control sample as described in the experimental section, and calculated the fucosylation degree of each aliquot as listed in Table 2.2a. The relative standard deviation (RSD) is only 6.3% for these four replicates, indicating high analytical reproducibility of this assay. Intra-assay variability has several possible sources: (1) losses or degradation of glycans during deglycosylation, desialylation or glycan purification; (2) peeling reactions at high pH during permethylation; (3) sample loss during liquid-liquid extraction in permethylation; (4) glycans were unevenly distributed on the MALDI plate due to uneven crystal formation of matrix. Although the reproducibility is excellent, special attention should be paid to sample handling to improve the stability of analysis.

In order to determine how repetitive freeze-thaw effects glycan stability, we performed one, two and four freeze-thaw cycles on three aliquots of the same serum sample from a pancreatic cancer Stage IIB patient. On the protein level, we did not identify any significant differences in the on-plate tryptic digestion spectrum. There were no missing peptides and there was no significant change in relative intensities of the peptides (data not shown). On the glycan level all 8 glycans were identified. Fucosylation degrees of three aliquots as displayed in Table 2.2b are consistent for the three aliquots, and the RSD is 4.6%. Our study indicates that freeze-thaw cycles did not exert significant modifications at either the peptide level or glycan level, up to four freeze-thaw cycles analyzed.

2.3.5 MS/MS study confirms glycan composition and location of fucosylation

While accurate masses can provide the oligosaccharide composition such as numbers of Hex (hexose), HexNAc (N-acetylhexosamine) and Fuc (fucose), MS² was performed to confirm oligosaccharide composition. At low energy CID, the predominant fragments

are y- ions resulting from the cleavage of labile GlcNAc-Hex glycosidic bond, thus the oligosaccharide composition can be inferred from mass differences of fragment ions.³² Permethylated also allows one to obtain branching information of a glycan because only unoccupied hydroxyl groups can be permethylated.³⁰ Fucosylation occurs during maturation of N-glycans and fucoses are usually attached to the N-glycan innermost core GlcNAc via α 1-6 linkage, or to subterminal GlcNAc via α 1-3 or α 1-4 linkage. While the fragmentation from CID could not provide detailed linkage information, one is able to discriminate between core fucosylation and antennary fucosylation based on diagnostic ions.

The CID spectrum of the newly identified bifucosylated triantennary glycan (m/z 2867.46) is shown in Figure 2.6a, demonstrating that it is a tri-antennary structure with both core and antennary fucosylation. The trimannosyl core fragment ion at m/z 852.24 carries the information that three antenna are originally attached.³⁰ The fragment ion at m/z 2404.05 is the product after loss of one non-reducing terminal Gal-GlcNAc. From the fact that the ion at m/z 2404.05 is able to lose another non-reducing terminal Gal-GlcNAc, producing an ion at m/z 1940.79 with high intensity, we know that for the majority of the glycans, one of the two fucoses must be attached to the core GlcNAc. Core fucosylation is also confirmed by the presence of a fragment ion at m/z 1303.46 corresponding to a fucosylated pentasaccharide core structure. A peak at m/z 1489.49 resulting from loss of core Fuc-GlcNAc unambiguously indicates antennary fucosylation. We cannot rule out the possibility that a minority of the glycan structures have both fucoses at the antenna as reported in Sarrats' work¹⁶ since there is a minor peak at m/z 2590.03 corresponding to loss of core GlcNAc only from the parent ion.

Similarly, we could conclude that glycan 2 (m/z 2244.20) as in Figure 2.6b is core fucosylated because of a diagnostic fragment at 1317.40 which corresponds to a fucosylated pentasaccharide core. This conclusion is also supported by an ion at 1329.40 resulting from loss of core Fuc-GlcNAc. Likewise, glycan 7 (m/z 3142.59) in Figure 2.6c is antennary fucosylated. A diagnostic peak at m/z 1939.08 is the cleavage product of peak at 2216.25 after loss of core GlcNAc, revealing that there is no core fucose attached originally. The glycan is only able to sequentially lose three antennary Gal-GlcNAc, resulting in fragments at m/z 2679.38, 2216.25 and 1753.14. The loss of the fourth antennary Gal-GlcNAc (theoretical m/z at 1290.01) is not observed, revealing that the fucose is located at this fourth antennary GlcNAc.

Thus, MS/MS analysis reveals that both core fucosylation and antennary fucosylation are present in the elevated haptoglobin fucosylation in pancreatic cancer.

2.4 Conclusion

In our study, a highly-sensitive analytical strategy is developed to elucidate N-glycan structures and evaluate fucosylation for target proteins in human cancer serum. A volume of 10 μ L human serum is sufficient for total analysis. This strategy consists of five steps: immunoprecipitation of target protein, deglycosylation and desialylation, purification of glycans, permethylation, and mass spectrometric analysis for structure elucidation of glycans. Glycans of haptoglobin from pancreatic cancer serum, control disease states, and normal controls were used to evaluate this method. Compared with the lectin microarray method, the mass spectrometric method on permethylated glycans yields more abundant structural information and has improved reproducibility. However, the large numbers of

glycans discovered and great overlap of glycans between cancer and non-cancer makes subsequent data analysis challenging. In our work, preliminary studies revealed fucosylated N-glycans as the indicative pancreatic cancer biomarker, hence desialylation was performed to simplify the glycan profile without losing fucosylation information. Due to the high sensitivity nature of our method, a tri-antennary complex type glycan with both core and antennary fucosylation was identified for the first time in haptoglobin in pancreatic cancer. This glycan as well as another bifucosylated tetra-antennary glycan are unique in pancreatic cancer, and do not appear in normal controls, chronic pancreatitis or type II diabetes. Besides, singly-fucosylated tri-antennary and tetra-antennary glycans were also found to be increased in pancreatic cancer.

Fucosylation degree index which measures the degree of fucosylation and hence activity of fucosyltransferase was utilized to provide a numerical depiction of haptoglobin glycomic differences between pancreatic cancer and other pancreas chronic diseases/normal controls. The predictive power of fucosylation degree index was evaluated with a sensitivity of 94% and a specificity of 100%. Elevation of haptoglobin fucosylation is present in all stages of pancreatic cancer including the early stage, revealing the potential of early detection, though this result has to be verified in a larger sample cohort. Because pancreatic cancer is a relatively rare disease, this method is most likely to be applied to high risk cohorts such as those with a genetic predisposition or those identified through risk stratification modeling. In the future, we plan to carry out this assay in an easily used and high-throughput format for larger sample size screening. Agarose bead-based immunoprecipitation can be performed in a 96-well filter plate. Sample incubation, purification with porous graphitized carbon resin and derivatization

can also be carried out in a 96-well filter plate platform in analogy to what is describe in previous work.³³

Figures

Figure 2. 1. Workflow to characterize N-glycan structures of haptoglobin and study fucosylation differences between pancreatic cancer and other benign pancreatic diseases/normal controls.

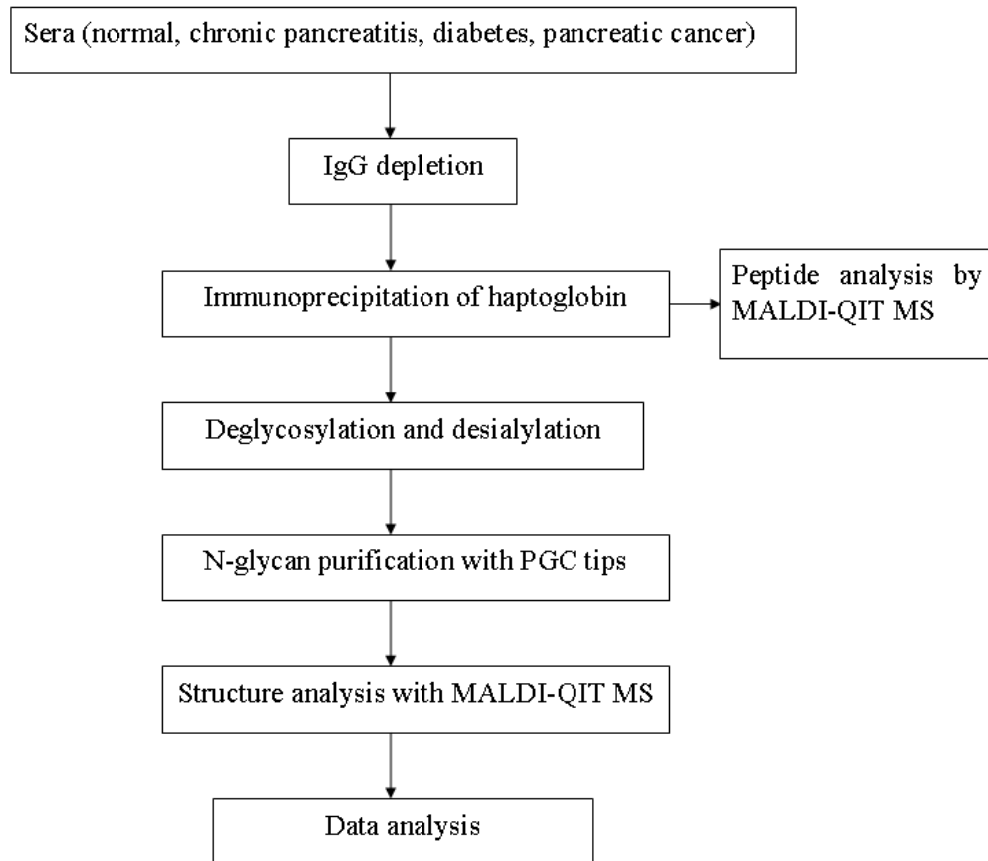
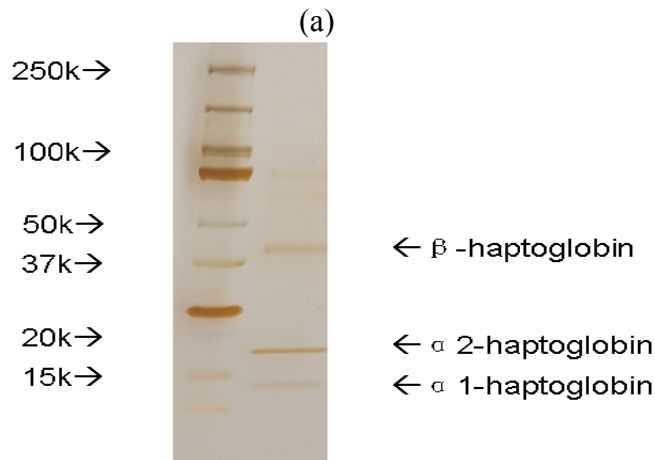
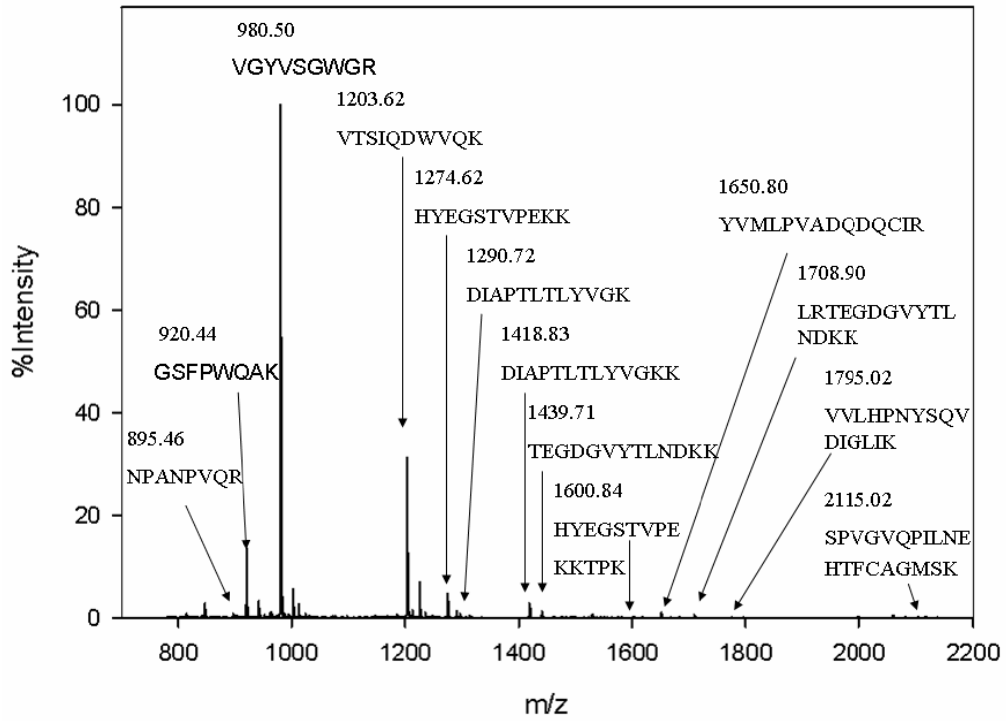
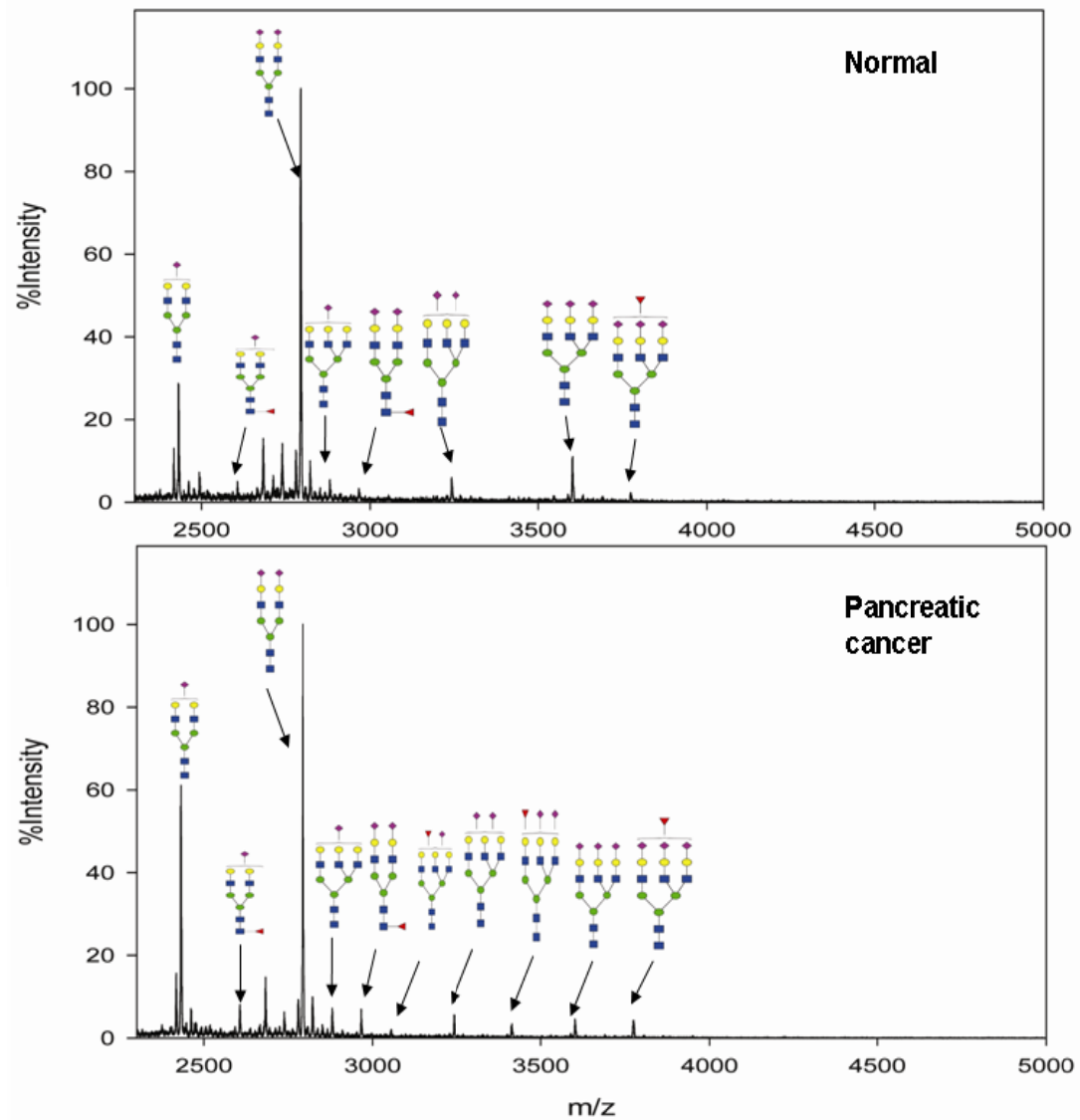


Figure 2. 2. Purification of haptoglobin from 10uL of human serum. (a) Tryptic peptide profile of haptoglobin resulted from 10-min on-plate digestion followed by MALDI-QIT-TOF MS analysis. (b) Patterns of haptoglobin α -1, α -2 and β chain on SDS-PAGE followed by silver staining.

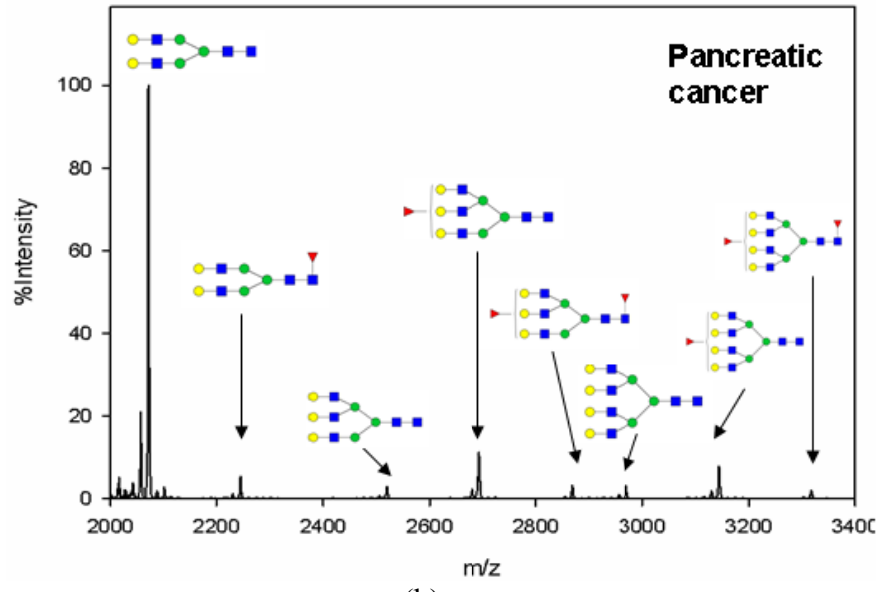
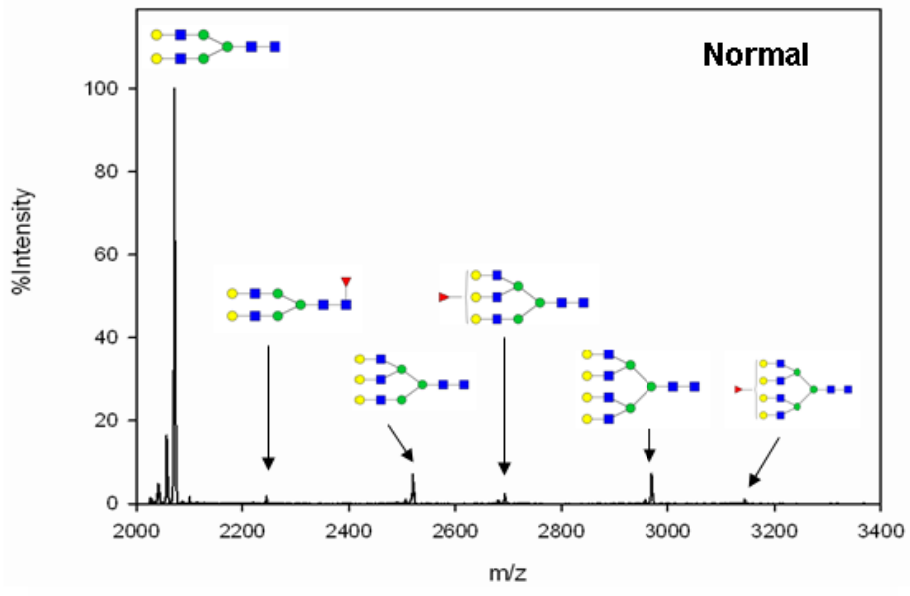


(b)

Figure 2. 3. (a) MALDI-QIT-TOF MS spectra of haptoglobin glycans from pooled serum of normal controls (upper panel) and pancreatic cancer patients (lower panel). (b) MALDI-QIT-TOF spectrum of desialylated haptoglobin glycans from serum of a normal control (upper panel) and a stage IV pancreatic cancer patient (lower panel). (red triangle-Fuc, blue square-GlcNAc, green circle-Man, yellow circle-Gal, purple diamond-NeuAc)



(a)



(b)

Figure 2. 4. Zoom-in MALDI-QIT-TOF mass spectra showing fucosylation difference in tri-antennary and tetra-antennary N-glycans of haptoglobin from serum of pancreatic cancer, normal control, chronic pancreatitis and type II diabetes.

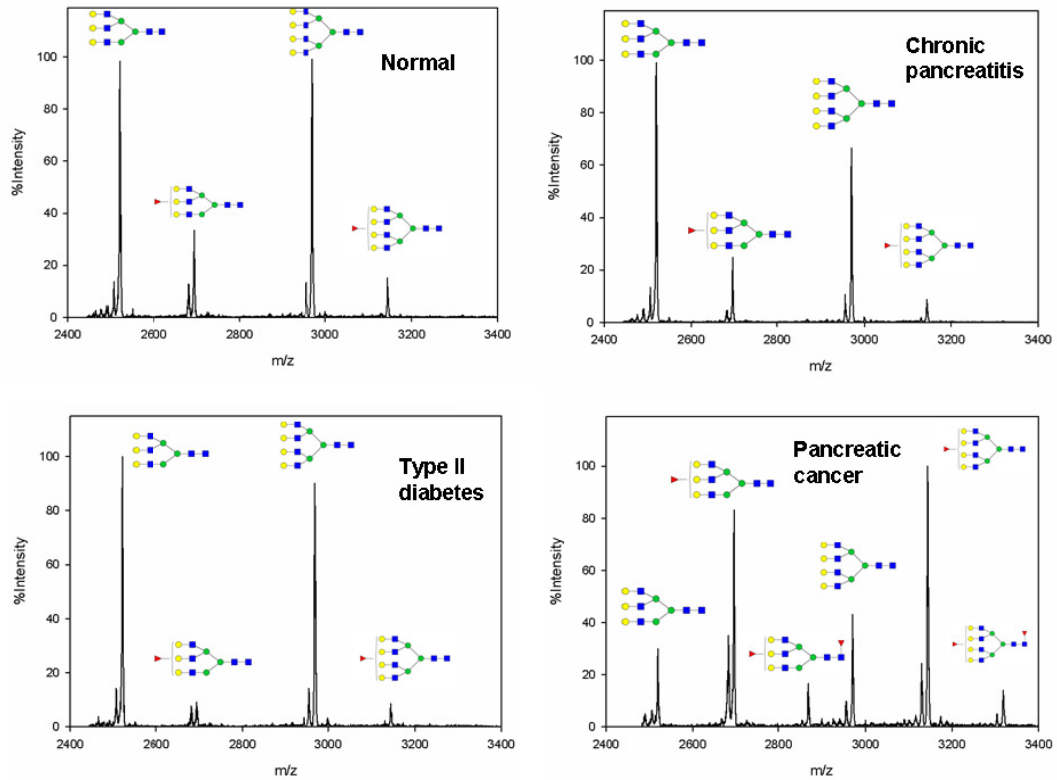


Figure 2. 5. Scatter plot of fucosylation degrees of haptoglobin N-glycans of pancreatic cancer patients and non-cancer samples. Fucosylation degrees are elevated in pancreatic cancer. P-value of pair-wise t-test is 1.9×10^{-7} .

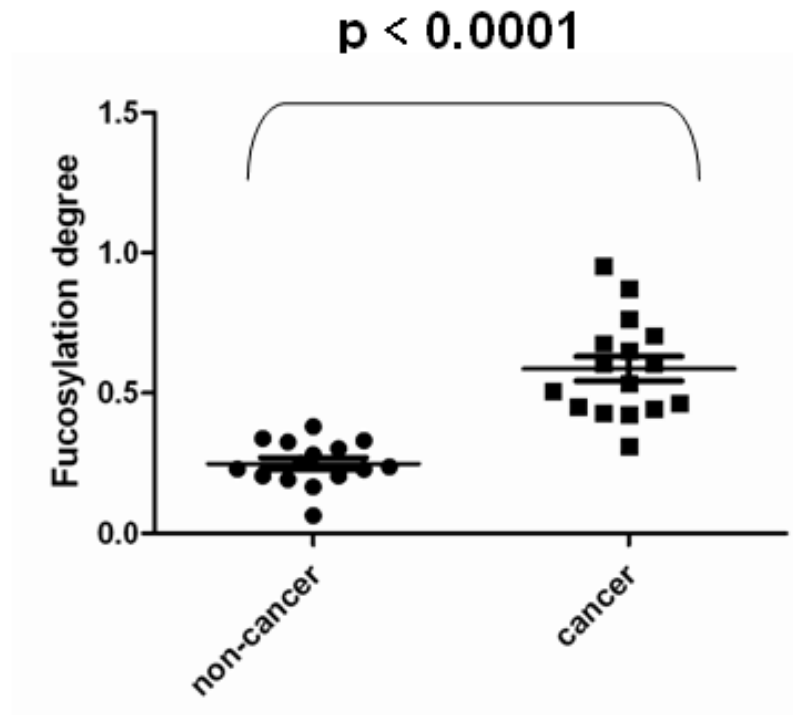
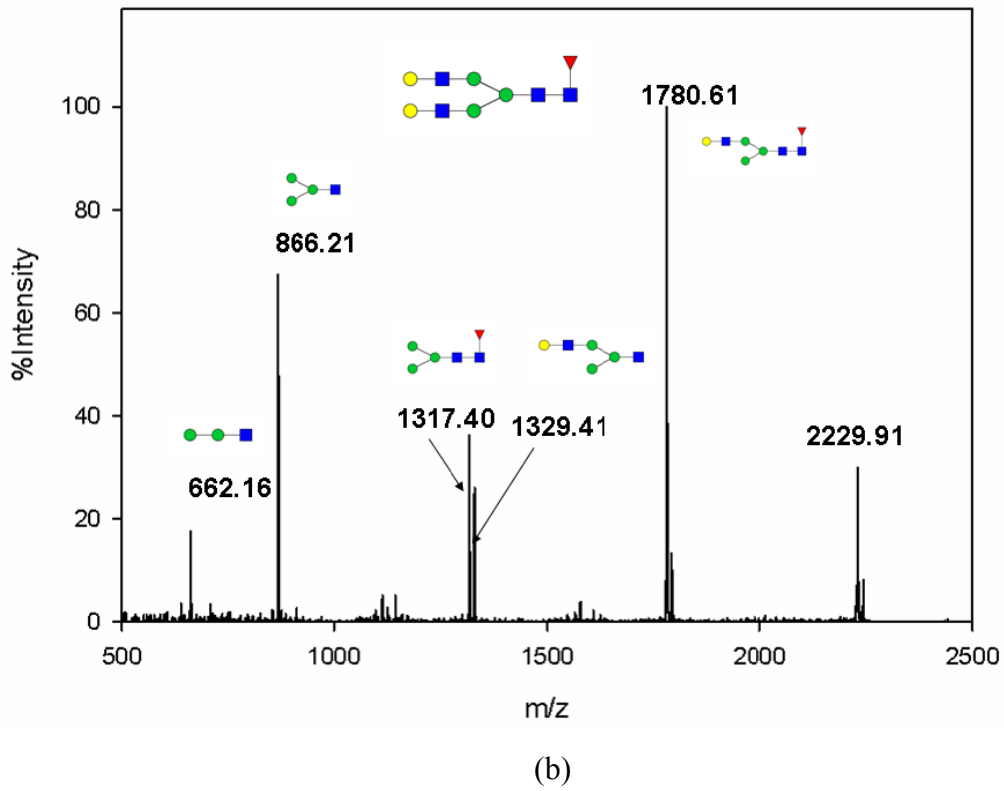
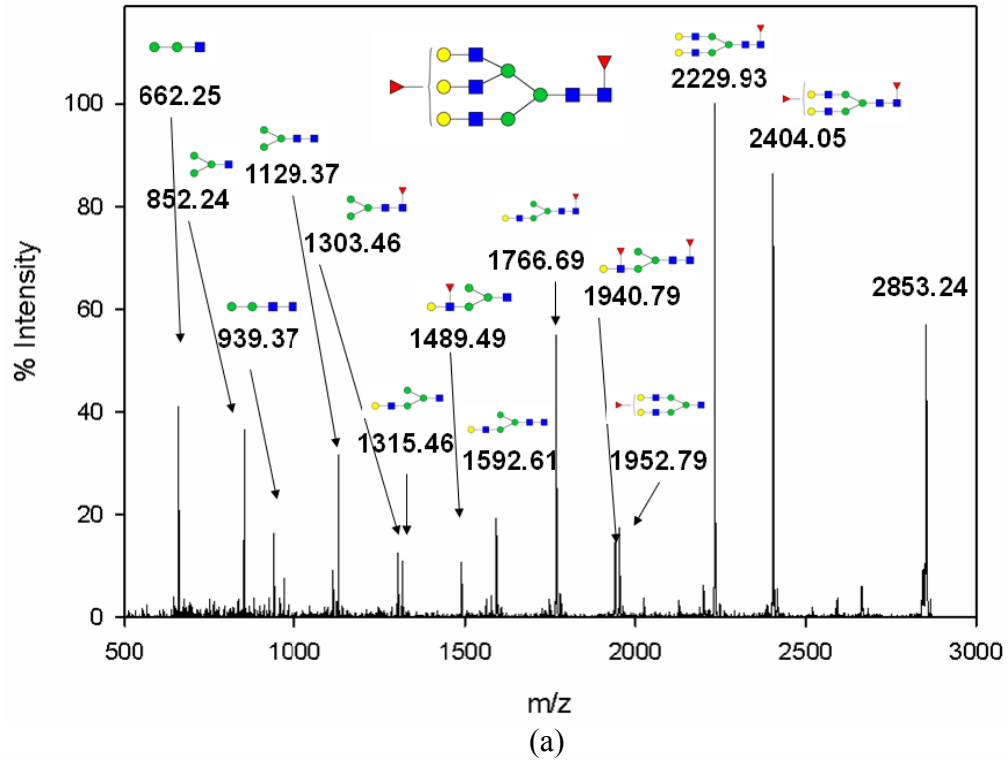
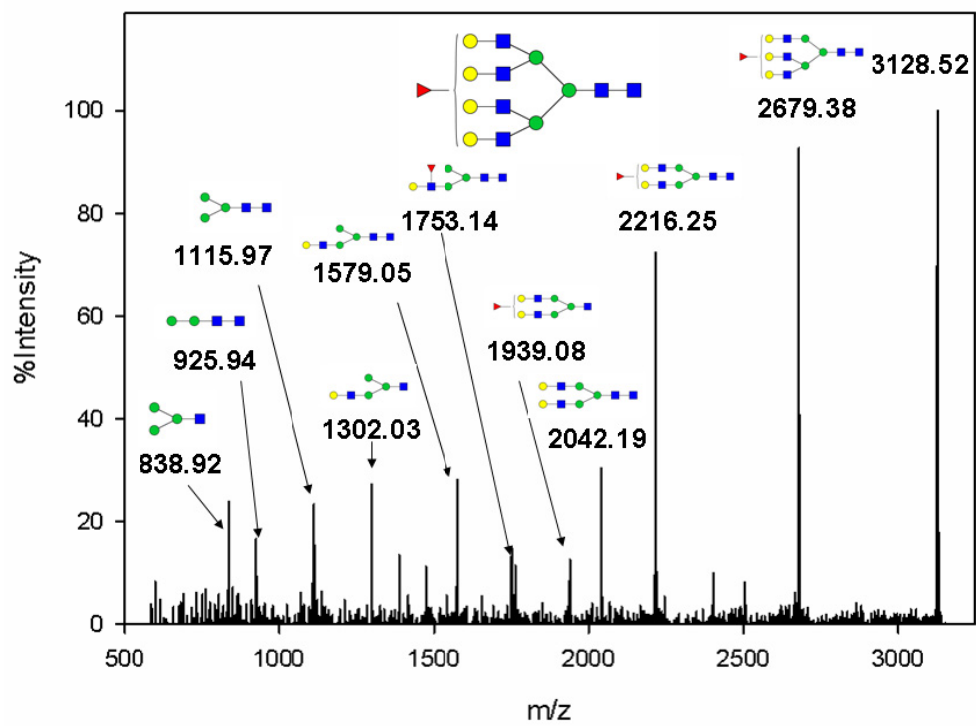


Figure 2. 6. MALDI-QIT-TOF MS/MS spectra of glycans at m/z 2867.46 (a), m/z 2244.20 (b), and m/z 3142.59 (c). Potential structures of fragment ions are labeled.





(c)

Table 2. 1. Eight desialylated N-glycans identified in human serum.









Peak NO.	m/z	structure
1	2070.10	
2	2244.20	
3	2519.36	
4	2693.45	
5	2867.46	
6	2968.53	
7	3142.59	
8	3316.60	

Table 2. 2. Summary of sample disease states and number of correct classification with fucosylation degree index

Disease states	Number of samples	Number of correct classifications
Normal	5	5
Chronic pancreatitis	5	5
Type II diabetes	5	5
Cancer stage IIA and prior	4	3
Cancer stage IIB	4	4
Cancer stage III	4	4
Cancer stage IV	4	4

Table 2. 3. (a) Reproducibility test of the assay for four aliquots of one normal serum sample processed following the workflow in Figure 1, (b) Influence of freeze/thaw cycles evaluated for three aliquots of one pancreatic cancer serum sample frozen/thawed once, twice and four times, and processed independently

	Aliquot1	Aliquot2	Aliquot3	Aliquot4	RSD
Fucosylation degree	0.3311	0.2876	0.2988	0.2949	6.3%

(a)

	Aliquot1	Aliquot2	Aliquot3	RSD
Fucosylation degree	0.7641	0.7240	0.6964	4.7%

(b)

Reference

- (1) Li, D. H.; Xie, K. P.; Wolff, R.; Abbruzzese, J. L., Pancreatic cancer. *Lancet* **2004**, 363, (9414), 1049-1057.
- (2) Rosty, C.; Goggins, M., Early detection of pancreatic carcinoma. *Hematology-Oncology Clinics of North America* **2002**, 16, (1), 37-+.
- (3) Hammoud, Z. T.; Mechref, Y.; Hussein, A.; Bekesova, S.; Zhang, M.; Kesler, K. A.; Novotny, M. V., Comparative glycomic profiling in esophageal adenocarcinoma. *J. Thorac. Cardiovasc. Surg.* **2010**, 139, (5), 1216-1223.
- (4) Kirmiz, C.; Li, B. S.; An, H. J.; Clowers, B. H.; Chew, H. K.; Lam, K. S.; Ferrige, A.; Alecio, R.; Borowsky, A. D.; Sulaimon, S.; Lebrilla, C. B.; Miyamoto, S., A serum glycomics approach to breast cancer biomarkers. *Mol. Cell. Proteomics* **2007**, 6, (1), 43-55.
- (5) Zhao, J.; Qiu, W. L.; Simeone, D. M.; Lubman, D. M., N-linked glycosylation profiling of pancreatic cancer serum using capillary liquid phase separation coupled with mass spectrometric analysis. *J. Proteome Res.* **2007**, 6, (3), 1126-1138.
- (6) Yang, Z. P.; Harris, L. E.; Palmer-Toy, D. E.; Hancock, W. S., Multilectin affinity chromatography for characterization of multiple glycoprotein biomarker candidates in serum from breast cancer patients. *Clin. Chem.* **2006**, 52, (10), 1897-1905.
- (7) Meany, D. L.; Zhang, Z.; Sokoll, L. J.; Zhang, H.; Chan, D. W., Glycoproteomics for Prostate Cancer Detection: Changes in Serum PSA Glycosylation Patterns. *J. Proteome Res.* **2009**, 8, (2), 613-619.
- (8) Dube, D. H.; Bertozzi, C. R., Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **2005**, 4, (6), 477-488.
- (9) Pucic, M.; Pinto, S.; Novokmet, M.; Knezevic, A.; Gornik, O.; Polasek, O.; Vlahovicek, K.; Wang, W.; Rudd, P. M.; Wright, A. F.; Campbell, H.; Rudan, I.; Lauc, G., Common aberrations from the normal human plasma N-glycan profile. *Glycobiology* **2010**, 20, (8), 970-975.
- (10) An, H. J.; Kronewitter, S. R.; de Leoz, M. L. A.; Lebrilla, C. B., Glycomics and disease markers. *Curr. Opin. Chem. Biol.* **2009**, 13, (5-6), 601-607.
- (11) Mas, E.; Pasqualini, E.; Caillol, N.; El Battari, A.; Crotte, C.; Lombardo, D.; Sadoulet, M. O., Fucosyltransferase activities in human pancreatic tissue: comparative study between cancer tissues and established tumoral cell lines. *Glycobiology* **1998**, 8, (6), 605-613.
- (12) Staudacher, E.; Altmann, F.; Wilson, I. B. H.; Marz, L., Fucose in N-glycans: from plant to man. *BBA-Gen. Subjects* **1999**, 1473, (1), 216-236.
- (13) Ogawa, J.; Inoue, H.; Koide, S., Expression of alpha-1,3-fucosyltransferase type IV and VII genes is related to poor prognosis in lung cancer. *Cancer Res.* **1996**, 56, (2), 325-329.
- (14) Sterling, R. K.; Jeffers, L.; Gordon, F.; Sherman, M.; Venook, A. P.; Reddy, K. R.; Satomura, S.; Schwartz, M. E., Clinical utility of AFP-L3% measurement in North American patients with HCV-related cirrhosis. *Am. J. Gastroenterol.* **2007**, 102, (10), 2196-2205.
- (15) Xiong, L.; Regnier, F. E., Use of a lectin affinity selector in the search for unusual glycosylation in proteomics. *J. Chromatogr. B* **2002**, 782, (1-2), 405-418.
- (16) Sarrats, A.; Saldova, R.; Pla, E.; Fort, E.; Harvey, D. J.; Struwe, W. B.; de Llorens, R.; Rudd, P. M.; Peracaula, R., Glycosylation of liver acute-phase proteins in pancreatic

- cancer and chronic pancreatitis. *Proteom. Clin. Appl.* **2010**, 4, (4), 432-448.
- (17) Dobryszczycka, W., Biological functions of haptoglobin - New pieces to an old puzzle. *Eur. J. Clin. Chem. Clin. Biochem.* **1997**, 35, (9), 647-654.
- (18) Ang, I. L.; Poon, T. C. W.; Lai, P. B. S.; Chan, A. T. C.; Ngai, S. M.; Hui, A. Y.; Johnson, P. J.; Sung, J. J. Y., Study of serum haptoglobin and its glycoforms in the diagnosis of hepatocellular carcinoma: A glycoproteomic approach. *J. Proteome Res.* **2006**, 5, (10), 2691-2700.
- (19) Fujimura, T.; Shinohara, Y.; Tissot, B.; Pang, P. C.; Kurogochi, M.; Saito, S.; Arai, Y.; Sadilek, M.; Murayama, K.; Dell, A.; Nishimura, S. T.; Hakomori, S. I., Glycosylation status of haptoglobin in sera of patients with prostate cancer vs. benign prostate disease or normal subjects. *Int. J. Cancer* **2008**, 122, (1), 39-49.
- (20) Park, S. Y.; Yoon, S. J.; Jeong, Y. T.; Kim, J. M.; Kim, J. Y.; Bernert, B.; Ullman, T.; Itzkowitz, S. H.; Kim, J. H.; Hakomori, S. I., N-glycosylation status of beta-haptoglobin in sera of patients with colon cancer, chronic inflammatory diseases and normal subjects. *Int. J. Cancer* **2010**, 126, (1), 142-155.
- (21) Thompson, S.; Dargan, E.; Turner, G. A., INCREASED FUCOSYLATION AND OTHER CARBOHYDRATE CHANGES IN HAPTOGLOBIN IN OVARIAN-CANCER. *Cancer Lett.* **1992**, 66, (1), 43-48.
- (22) Yoon, S. J.; Park, S. Y.; Pang, P. C.; Gallagher, J.; Gottesman, J. E.; Dell, A.; Kim, J. H.; Hakomori, S. I., N-glycosylation status of beta-haptoglobin in sera of patients with prostate cancer vs. benign prostate diseases. *Int. J. Oncol.* **2010**, 36, (1), 193-203.
- (23) Zhao, J.; Patwa, T. H.; Qiu, W. L.; Shedden, K.; Hinderer, R.; Misek, D. E.; Anderson, M. A.; Simeone, D. M.; Lubman, D. M., Glycoprotein microarrays with multi-lectin detection: Unique lectin binding patterns as a tool for classifying normal, chronic pancreatitis and pancreatic cancer sera. *J. Proteome Res.* **2007**, 6, (5), 1864-1874.
- (24) Okuyama, N.; Ide, Y.; Nakano, M.; Nakagawa, T.; Yamanaka, K.; Moriwaki, K.; Murata, K.; Ohigashi, H.; Yokoyama, S.; Eguchi, H.; Ishikawa, O.; Ito, T.; Kato, M.; Kasahara, A.; Kawano, S.; Gu, J. G.; Miyoshi, E., Fucosylated haptoglobin is a novel marker for pancreatic cancer: A detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. *Int. J. Cancer* **2006**, 118, (11), 2803-2808.
- (25) Matsumoto, H.; Shinzaki, S.; Narisada, M.; Kawamoto, S.; Kuwamoto, K.; Moriwaki, K.; Kanke, F.; Satomura, S.; Kumada, T.; Miyoshi, E., Clinical application of a lectin-antibody ELISA to measure fucosylated haptoglobin in sera of patients with pancreatic cancer. *Clin. Chem. Lab. Med.* **2010**, 48, (4), 505-512.
- (26) Nakano, M.; Nakagawa, T.; Ito, T.; Kitada, T.; Hijioka, T.; Kasahara, A.; Tajiri, M.; Wada, Y.; Taniguchi, N.; Miyoshi, E., Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: A novel approach for the development of tumor markers. *Int. J. Cancer* **2008**, 122, (10), 2301-2309.
- (27) Kang, P.; Mechref, Y.; Klouckova, I.; Novotny, M. V., Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun. Mass Spectrom.* **2005**, 19, (23), 3421-3428.
- (28) Imre, T.; Kremmer, T.; Heberger, K.; Molnar-Szollosi, E.; Ludanyi, K.; Pocsfalvi, G.; Malorni, A.; Drahos, L.; Vekey, K., Mass spectrometric and linear discriminant analysis of N-glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals. *J. Proteomics* **2008**, 71, (2), 186-197.
- (29) Harvey, D. J., Matrix-assisted laser desorption/ionization mass spectrometry of

- carbohydrates. *Mass Spectrom. Rev.* **1999**, 18, (6), 349-450.
- (30) Yu, S. Y.; Wu, S. W.; Khoo, K. H., Distinctive characteristics of MALDI-Q/TOF and TOF/TOF tandem mass spectrometry for sequencing of permethylated complex type N-glycans. *Glycoconjugate J.* **2006**, 23, (5-6), 355-369.
- (31) Wada, Y.; Azadi, P.; Costello, C. E.; Dell, A.; Dwek, R. A.; Geyer, H.; Geyer, R.; Kakehi, K.; Karlsson, N. G.; Kato, K.; Kawasaki, N.; Khoo, K. H.; Kim, S.; Kondo, A.; Lattova, E.; Mechref, Y.; Miyoshi, E.; Nakamura, K.; Narimatsu, H.; Novotny, M. V.; Packer, N. H.; Perreault, H.; Peter-Katalinic, J.; Pohlentz, G.; Reinhold, V. N.; Rudd, P. M.; Suzuki, A.; Taniguchi, N., Comparison of the methods for profiling glycoprotein glycans - HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **2007**, 17, (4), 411-422.
- (32) Harvey, D. J.; Martin, R. L.; Jackson, K. A.; Sutton, C. W., Fragmentation of N-linked glycans with a matrix-assisted laser desorption/ionization ion trap time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **2004**, 18, (24), 2997-3007.
- (33) Kim, Y. G.; Jeong, H. J.; Jang, K. S.; Yang, Y. H.; Song, Y. S.; Chung, J. H.; Kim, B. G., Rapid and high-throughput analysis of N-glycans from ovarian cancer serum using a 96-well plate platform. *Anal. Biochem.* **2009**, 391, (2), 151-153.

Chapter 3

An N-glycosylation Analysis of Human Alpha-2-Macroglobulin Using an Integrated Approach

3.1 Introduction

Glycosylation is one of the most prevalent post-translational modifications for proteins and plays an important role in cell recognition, signal transduction and cell proliferation. N-glycosylation is the most common type of glycosylation, where glycans attach to asparagine in the consensus sequence N-X-S/T, where X cannot be proline. Aberrant glycosylation, such as site-specific glycosylation abnormalities, has been found to be associated with various types of cancers or other malignancies.¹⁻³ Site-specific glycosylation information is of great importance in both clinical research and fundamental biology.

N-glycosylation analysis at the glycopeptide level currently remains challenging for many reasons. Proteins can have multiple glycosylation sites (glycosylation heterogeneity) and each site can be occupied by more than one glycan (glycosylation microheterogeneity), hence the concentrations of individual glycopeptides are usually very low and require highly sensitive methods. Glycopeptides have low ionization efficiency and often suffer from ion suppression from non-glycopeptides during mass spectrometric analysis. Furthermore, collision induced dissociation (CID) of glycopeptides mostly produces fragments from the glycan moiety and no peptide sequence or glycosylation site information is produced.⁴

Recently introduced electron transfer dissociation (ETD) overcomes some of the limitations of CID for glycopeptide analysis.⁵ ETD uses reagents, such as nitrobenzene or fluoranthene, to produce radical anions that interact with analyte cations (normally with charge +3 or above) to produce fragmentation mainly along the peptide backbone, generating c and z type ions without disrupting the glycan.⁶ By combining ETD, which mainly yields information on peptide sequence and glycosylation sites, with CID, which cleaves glycosidic bonds to reveal glycan composition, site-specific glycosylation can be identified in a single LC MS/MS run.^{5,7}

An alternative method to identify glycosylation sites uses endo- β -N-acetylglucosaminidases to partially deglycosylate glycopeptides.⁸⁻¹⁰ Endo H and Endo F1 cleave high-mannose and hybrid type glycans. Endo F2 and Endo F3 are able to cleave complex type glycans with two or three branches.¹¹ Endo- β -N-acetylglucosaminidase F3 cleaves between the two GlcNAc at the pentasaccharide core, leaving only the innermost GlcNAc and core fucose, if present, attached to the peptides.¹¹ The fact that fucose remains attached provides endo- β -N-acetylglucosaminidase with the unique ability to identify core fucosylation sites. Core fucosylation occurs during the maturation stage of glycosylation where fucoses are added to the innermost GlcNAc via α (1,6) linkage.¹² Endo- β -N-acetylglucosaminidase leaves a 203 Da (GlcNAc) or 349 Da (GlcNAc-Fuc) modification on the asparagine and such modifications allow for reliable identification of glycosylation sites by CID MS/MS and database searching. False positive identifications are less common because of the large mass increment introduced by GlcNAc or GlcNAc-Fuc attachment.

Alpha-2-macroglobulin is one of the largest and the most abundant proteins in human serum with a molecular weight around 720 kDa. Usually in a tetrameric form, alpha-2-macroglobulin is an acute phase protein that is mainly synthesized in the liver and is known as a proteinase inhibitor.¹³ Alpha-2-macroglobulin has eight potential N-glycosylation sites at N55, N70, N247, N396, N410, N869, N991 and N1424, and all the sites have been identified by previous work.¹⁴⁻¹⁸ However, the previous studies were performed on completely deglycosylated peptides where no information on glycans was obtained. In the current work, we have used several complementary approaches to study the site-specific glycosylation pattern of alpha-2-macroglobulin. Three N-glycosylation sites (N70, N396 and N1424) were identified in the CID/ETD MS/MS approach, where heterogeneity of glycans at each site was described. Five N-glycosylation sites (N396, N410, N869, N991 and N1424) were found with the Endo F3 partial deglycosylation method, which uniquely revealed core fucosylation at site N396, N410 and N1424. This integrated approach to studying N-glycosylation was performed with only 10 μ L of serum and could serve as a model for studies of other glycoproteins.

3.2 Experimental Section

3.2.1 Serum samples

Human normal sera were provided by the University Hospital, Ann Arbor, Michigan according to IRB approval. The samples were aliquoted and stored in a -80 °C freezer until further use. All samples were frozen and thawed only once.

3.2.2 Purification of alpha-2-macroglobulin from serum

Alpha-2-macroglobulin was purified from human serum as described in previous

work.¹⁹ Briefly, 10 μ L of human serum was depleted of immunoglobulin (IgG) with protein A/G agarose beads (Pierce Scientific, Rockford, IL) to avoid IgG interference during immunoprecipitation. Twenty micrograms of alpha-2-macroglobulin antibody (Abcam, Cambridge, MA) was immobilized on protein A/G agarose beads with disuccinimidyl suberate (DSS) crosslinker and then incubated with the depleted serum overnight. Alpha-2-macroglobulin was eluted with 100 mM glycine-HCl at pH 2.8 and desalted with 75 μ L Zeba desalting columns (Pierce Scientific, Rockford, IL).

On-plate tryptic digestion and mass spectrometric analysis were performed to verify the success of immunoprecipitation. Desalted alpha-2-macroglobulin was spotted on the MALDI plate and dried in air. Trypsin (Promega, Madison, WI) (0.4 μ g) was added to 10 μ L of 100 mM ammonium bicarbonate solution with 20% of acetonitrile and added onto the alpha-2-macroglobulin spot and incubated in a humid chamber at 37 °C for 10 min. 2,5-dihydroxybenzoic acid (DHB, 10mg/mL) (Laser Biolabs, France) was dissolved in 50% acetonitrile with 0.1% trifluoroacetic acid and added on top of the dried spot. The peptide peaks were searched against SWISS-PROT *Homo sapiens* and other mammalia protein database (2012_03) using Mascot, with methionine oxidation set as a variable modification. The tolerance for MS matching was set at 0.2 Da. Mass spectrometric analysis was performed on an Axima MALDI quadrupole ion trap-time of flight mass spectrometer (Shimadzu Biotech, Manchester, U.K.). A pulsed nitrogen laser (337 nm) at 5Hz was used for ionization. Helium was used to cool the trapped ions and argon was used for CID fragmentation. All spectra were acquired in the positive ion mode. Spectra were calibrated with an external peptide standard mixture (Bruker Daltonics, Billerica, MA) to a mass accuracy of 30 ppm.

3.2.3 Deglycosylation, purification, permethylation and identification of N-glycans

Alpha-2-macroglobulin was denatured in 10% denaturing solution (0.02% SDS, 10 mM 2-mercaptoethanol) at 60 °C for 30 min. Ammonium bicarbonate solution was added to a final concentration of 15 mM. N-glycosidase F (New England Biolabs, Ipswich, MA) was added to release N-glycans at 37 °C overnight. Ten microliter porous graphitized carbon tips (Sigma Aldrich, St. Louis, MO) were used to purify N-glycans from proteins and other impurities as described previously.¹⁹ In-solution permethylation was performed on dried purified N-glycans according to published procedures.¹⁹ Permethylated N-glycans were dissolved in 2.5 µL of 20% acetonitrile and 0.5 µL was spotted on a MALDI plate and 0.5 µL sodiated DHB (10 mg/mL in 50% acetonitrile with 100 mM sodium chloride) was spotted on top. N-glycans were analyzed by MALDI-QIT-TOF MS with the same parameters as described before. Glycomod (<http://www.expasy.org/tools/glycomod>) was utilized to predict the N-glycan compositions. Only N-glycans included in the GlycoSuite database were selected. The glycan compositions were further confirmed with CID MS/MS analysis.

3.2.4 LC-ESI-CID/ETD-MS analysis of chymotryptic glycopeptides

Purified alpha-2-macroglobulin was reduced with 10 mM of dithiothreitol (DTT) at 95 °C for 15 min and then alkylated with 22 mM of iodoacetamide (IAA) at room temperature in the dark for 15 min. Alpha-2-macroglobulin was diluted in 50 mM ammonium bicarbonate and incubated with 0.3 U of chymotrypsin (Promega, Madison, WI) at 37 °C for 16 h. Chymotrypsin was deactivated by boiling for 3 min. The digested mixture was dried in a SpeedVac and reconstituted in 10 µL 80% acetonitrile with 2% formic acid. ZIC-HILIC ziptips (Protea, Morgantown, WV) were used to enrich

glycopeptides. After the tips were equilibrated with 80% acetonitrile with 2% formic acid, the samples were loaded on the tips followed by washing with 80% acetonitrile with 2% formic acid to remove non-specific binding. The glycopeptides were eluted by 98% water with 2% formic acid.

Fused silica PicoTips (New Objectives, Woburn, MA) packed with C₁₈ material (5 µm particle size, 10 cm×75 µm i.d.) were used for both chromatographic separation and ionization spray. Gradient elution was performed on a Paradigm MG4 micropump system (Michrom Biosciences, Auburn, CA) at 300 nL/min with mobile phase A as 2% acetonitrile with 1% acetic acid in water and mobile phase B as 5% water with 1% acetic acid in acetonitrile. A 70 min gradient was used: (1) 5% B to 60% B in 35 min, (2) 60% B to 95% B in 1 min, (3) isocratic at 95% B for 4 min, (4) decrease from 95% B to 5% B in 0.1 min, (5) isocratic at 5% B for 30 min.

An LTQ-CID/ETD-MS (Thermo Fisher Scientific, San Jose, CA) operated in positive ion mode was used for all LC-MS experiments. The ESI spray voltage was set at 2.2 kV and capillary voltage at 45 V. The mass spectra were generated in a data-dependent manner. After a full scan from m/z 400 to m/z 1800, the three most intense ions were selected for ETD and CID fragmentation. For ETD MS/MS, the reactant temperature was 145 °C, the ionization energy was 70 V, the emission current was 15 µA, and the ion-ion reaction time with the reagent anion fluoranthene was set at 200 ms. For CID MS/MS, 35% of the normalized collision energy was used for fragmentation.

3.2.5 LC-ESI-CID-MS analysis of endo-β-N-acetylglucosaminidase F3 (Endo F3) treated chymotryptic glycopeptides

Chymotryptic glycopeptides of alpha-2-macroglobulin were purified as described in

section 2.4 and reconstituted in 50 mM sodium acetate. Ten mU Endo F3 (QAbio, Palm Desert, CA) was added and incubated with chymotryptic glycopeptides at 37 °C for 16 h. The resulting peptides were purified with 10 μ L C₁₈ ZipTips (Millipore, Billerica, MA). The tips were pre-wetted with 0.1% trifluoroacetic acid in 50% acetonitrile and equilibrated with 0.1% trifluoroacetic acid in water. The peptides were bound to the C₁₈ medium followed by washes to remove non-specific binding. Ten microliters of 50% acetonitrile with 0.1% trifluoroacetic acid were used for elution. The purified two-step digested glycopeptides were analyzed with LC-LTQ-CID-MS using the same LC method and MS parameters described in 2.4. After the MS survey scan, CID MS/MS was performed on the most intense ion, and the most intense fragment ion in MS/MS was selected for further MS³ fragmentation. CID MS/MS and MS³ were also performed on the second to the fourth most intense ions from the survey MS scan.

3.2.6 Data analysis

The MALDI data were acquired and processed in Launchpad software (Kratos, Manchester, U.K.), and the ESI data were acquired with Thermo Xcalibur software (Arlington, VA). The *m/z* values and intensities were exported as ASCII files and plotted in Sigmaplot (San Jose, CA).

For chymotryptic glycopeptide analysis, the oxonium ion (*m/z* 366) extracted ion chromatogram (XIC) was re-constructed to locate peptide elution times. CID MS/MS spectra were manually examined and only ions which generated both oxonium ions (*m/z* 204, 292, 366, 528 and 657) and b, y type glycosidic bond cleavages were considered as glycopeptides. Theoretical N-glycopeptide masses were calculated by adding the masses of theoretical chymotryptic peptides at N-glycosylation sites and the masses of N-glycans

obtained from MALDI MS analysis. Theoretical glycopeptide masses were matched against the masses obtained from LTQ-MS experiments. The peptide sequences were confirmed by matching the ETD MS/MS peaks with theoretical c, z type fragments listed in the Protein Prospector database (version 5.10.1) manually.

For glycopeptides resulting from Endo F3 partial deglycosylation, all CID MS/MS spectra were searched against SWISS-PROT *Homo sapiens* database (Release 2010_10, downloaded on Nov 2, 2010) for identification of glycosylation sites. Proteome Discoverer software (version 1.1, Thermo Fisher Scientific, San Jose, CA) incorporated with SEQUEST algorithm was used to perform searches. The following search parameters were used: (1) fixed modification: cysteine carbamidomethylation (+57.0 Da); (2) variable modification: methionine oxidation (+16.0 Da), and addition of N-acetylglucosamine (+203.1 Da) or N-acetylglucosamine-fucose (+349.1 Da) to asparagine; (3) missed cleavages allowed: three; (4) peptide ion tolerance: 1.4 Da; (5) fragmentation ion tolerance: 1.5 Da. All search results containing N-glycosylation sites were validated by manual examination of CID MS/MS spectra.

3.3 Results and Discussion

In our work, we sought to develop an integrated mass spectrometry-based workflow to identify site-specific N-glycosylation of human glycoproteins. Alpha-2-macroglobulin was selected for workflow development, because of the overall analytical complexity (8 possible glycosylation sites) and lack of previous studies detailing its glycosylation profile. The developed workflow for this study is outlined in Figure 3.1. The workflow systematically examines the glycoprotein at the glycan, glycopeptide, and peptide levels

to generate data that together provides a detailed description of the glycoprotein's N-glycosylation profile.

3.3.1 Purification of Alpha-2-macroglobulin from Human Serum

In this study, alpha-2-macroglobulin was immunoprecipitated from human serum. Ten-minute on-plate tryptic digestion followed by MALDI-QIT-TOF MS analysis was used to confirm the success of immunoprecipitation by peptide mass fingerprinting and MS/MS on the high intensity peaks. The mass spectrum is shown in Figure 3.2a with all the major peaks corresponding to alpha-2-macroglobulin tryptic peptides. The spectrum was searched against both the *Homo sapiens* and other mammalia protein database with Mascot, and returned alpha-2-macroglobulin as the only significant protein with 23 matched peptides. Among the 23 peptides, 21 are generated without miscleavages and 2 have 1 miss cleavage, reflecting the reasonable efficiency of on-plate digestion. On-plate fast trypsin digestion has high efficiency because of the high trypsin concentration, and the approximate protein-to-trypsin ratio is 1:1 rather than 50:1 which is normally used in traditional overnight digestion. The purity of alpha-2-macroglobulin was evaluated with SDS-PAGE followed by silver staining (Figure 3.2b). Alpha-2-macroglobulin monomer (~180 kDa) and two other cleavage fragments (~120 kDa and ~60 kDa) were observed with no other significant bands observed.

3.3.2 N-glycan Analysis of Alpha-2-macroglobulin

In-solution permethylation was performed on purified N-glycans to improve sensitivity during MS analysis and to stabilize the labile fucose and sialic acids.²⁰ A typical alpha-2-macroglobulin N-glycan profile by MALDI MS is shown in Figure 3.3a, which is dominated by the four most abundant N-glycans. These include bi-antennary

complex type glycans with one or two sialic acids and with or without fucosylation. CID MS/MS was performed on the four N-glycans to confirm the oligosaccharide compositions. At low energy CID, permethylated glycans typically generate y-ions resulting from cleavage of the labile GlcNAc-Gal bond or NeuAc-Gal bond. Fucoses are usually attached to the innermost GlcNAc via α 1-6 linkage or to subterminal GlcNAc via α 1-3 or α 1-4 linkage. Although CID does not help to determine linkage type, we utilized characteristic CID fragment ions to discriminate between terminal fucosylation and core fucosylation. The CID MS/MS spectrum of the singly sialylated biantennary fucosylated glycan at m/z 2966.44 is shown in Figure 3.3b. The signals at m/z 2230.03 and 2141.96 are the products after loss of the two terminal NeuAc and terminal GlcNAc-Gal-NeuAc from the parent ion, respectively. The core fucosylation was confirmed by the ion at m/z 1317.38 which corresponds to addition of GlcNAc-Fuc to the trimannosyl-GlcNAc core. The core fucosylation was further verified by glycopeptide analysis by chymotrypsin-Endo F3 two-step digestion.

3.3.3 MS/MS analysis of N-glycopeptides

Trypsin was initially used in our study, but very limited glycopeptide information was obtained, possibly because trypsin generated large glycopeptides which did not ionize well.⁴ Chymotrypsin cleaves the amide bonds C-terminal to hydrophobic amino acids, such as phenylalanine (F), tryptophan (W) and tyrosine (Y) and sometimes after methionine (M) and lysine (L). Chymotrypsin was found to produce more glycopeptides when compared to trypsin. While the lower cleavage specificity of chymotrypsin, when compared to trypsin, may limit the use of chymotrypsin in strict quantitative studies, the cleavage pattern was found to be very reproducible by precisely controlling the

incubation time.

A typical LC-MS base peak chromatogram of a chymotryptic glycopeptide is shown in Figure 3.4a. The glycan oxonium ion at m/z 366 (GlcNAc-Gal), which is a typical glycopeptide fragment, was used to locate the elution time of the glycopeptides. The extracted ion chromatogram of m/z 366 is shown in Figure 3.4b, revealing that most of the glycopeptides eluted between 26 min and 42 min. Summed mass spectra within a 1 min elution window around the peak maxima were obtained for all chromatographic peaks (27.15 min, 29.58 min, 37.35 min, 38.27 min and 41.77 min) between 26 min and 42 min in Figure 3.4a. Most of the mass spectrometric peaks were confirmed to be glycopeptides by manually inspecting the fragment ions from CID and ETD as discussed below. An example of an integrated mass spectrum between 37.60-38.60 min (Figure 3.4a) is shown in Figure 3.4c. It is noted that some peaks in Figure 3.4c were not identified as glycopeptides because there were few informative ETD fragments generated for determining peptide sequence and/or glycosylation site.

All glycopeptides were characterized by combining both CID and ETD fragment information. Since glycosidic bonds are more fragile than amide bonds and CID in the LTQ is a low-energy fragmentation method, the majority of product ions are from glycosidic bond cleavage, leaving the peptide backbone intact. A typical CID MS/MS spectrum of an N-glycopeptide (ESVRGNRSLF) is shown in Figure 3.5a. The lower mass range is dominated by three oxonium ions (m/z 366: GlcNAc-Gal, 528: Man-GlcNAc-Gal, and 657: GlcNAc-Gal-NeuAc). The higher mass range is dominated by glycan fragments with the intact peptide backbone. The Gal-NeuAc, GlcNAc-Man, Gal-GlcNAc, and Man-Man bonds were broken, resulting in ions at m/z 1397.7, 1357.1,

1130.7 and 1276.1, respectively. It is interesting to note that there are few glycosidic cleavages at the trimannose chitobiose core. Only one Man-Man bond cleavage (m/z 1276.1) was observed, while most of the glycopeptides fragments retain the core structure. Based on the partial glycopeptide fragment information, the glycan composition of glycopeptides was deduced.

It is well known that ETD generally produces *c* and *z* ions resulting from cleavage of the N-C_α bond and retains the post translational modifications such as glycosylation. The ETD spectrum of the same glycopeptide (ESVRGNRSLLF) is displayed in Figure 3.5b. ETD is believed to be less efficient and sensitive than CID²¹, and the ETD signals were generally lower than their CID counterparts based on our observation. The most abundant ion is the charge reduced species of the parent ion (m/z 1540.2, charge 2+). The presence of such charge reduced ions can be useful to determine the charge states of glycopeptides when using lower resolution instrumentation, such as the LTQ. In this case, 4 out of 9 *z* ions and 6 out of 9 *c* ions predicted were observed, as annotated in Figure 3.5b. The glycosylation site can be determined by the mass difference between *c*₅⁺ (m/z 546.3) and *c*₆²⁺ (m/z 1287.5). The mass difference is 2027.7, which is the addition of a glycan mass (1913.7) and the mass of asparagine (114.0). However, the useful signals are often less intense in ETD than in CID. It is reported that ETD is more efficient for charge states over 3+ and m/z lower than 1400.²¹ We also observed that for glycopeptides with a 2+ charge and m/z higher than 1600 (data not shown) that no ETD fragmentation occurred. A summary of all the N-glycopeptides identified for alpha-2-macroglobulin is shown in Table 3.1. Glycopeptides with the same peptide sequences, but slightly different glycans, eluted at approximately the same time during C₁₈ LC separation.⁷ Only glycopeptides that

met the following criteria are listed: (1) precursor masses must match with theoretical glycopeptide masses based on identified glycans and theoretical chymotryptic peptide masses, (2) CID spectra must contain both oxonium ions and glycan fragment ions with intact peptide backbones, and (3) ETD spectra must have a total of at least four matching c or z ions. Three sites (N70, N396 and N1424) were identified to be glycosylated with all four N-glycans as shown in Table 3.1. However, it should be noted that the other five sites may be glycosylated as well, but were not identified in this ETD/CID approach. This could be because the glycopeptides are of low abundance or these unidentified sites were associated with less abundant N-glycans which were not identified in the glycan analysis. It is also possible that these unidentified glycopeptides have lower ionization efficiencies or low ETD efficiencies and did not provide informative fragments.

3.3.4 MS/MS analysis of partially deglycosylated N-glycopeptides

Endoglycosidases have been used for the identification of glycosylation sites by partially deglycosylating glycopeptides and was recently used to identify 62 glycosylation sites from 37 serum glycoproteins.⁹ Endo- β -N-acetylglucosaminidase hydrolyzes the bonds between the two GlcNAc in chitobiose core linked to asparagine, so that only the innermost GlcNAc or GlcNAc-Fuc is retained on the asparagine side chain. Endo F3 is applicable to complex type biantennary or tri-antennary N-glycans, but is less efficient for tetra-antennary glycans. Compared to the ETD/CID combined approach for glycopeptides analysis, the endoglycosidase approach does not maintain glycosylation site microheterogeneity information. However, the advantage of this method is that signals from N-glycopeptides sharing the same peptide backbone are merged into two peaks (either core-fucosylated or non-core-fucosylated), so that mass spectral complexity

is reduced and overall signal intensity is increased. Furthermore, we found that CID of endoglycosidase treated peptides produced more peptide sequence ions²² which could be used for peptide sequencing and glycosylation site determination, as a complement to the ETD/CID approach.

A traditional approach for identification of N-glycosylation sites uses PNGase F to release N-glycans from proteins and depends on the mass shift (+0.98Da) of deamidation (asparagine to aspartic acid) during PNGase F hydrolysis for site determination.^{14-17, 23} However, this method requires high mass resolution mass spectrometers, making linear ion trap instruments less appropriate. Furthermore, deamidation can occur spontaneously as a sample artifact, rather than due to enzymatic action of PNGase F, increasing the chance of false positives.²⁴ The confidence of glycosylation site assignment can be improved by performing deglycosylation in H₂¹⁸O, introducing a mass shift of 2.98 Da.²⁵ However, partial ¹⁸O incorporation in the C-terminus may bring confusion in site identification. With the endoglycosidase approach, N-glycosylation site assignment is unambiguously confirmed by the residual GlcNAc or GlcNAc-Fuc with greater mass increments, which reduces false positive identification rate.

Based on the N-glycan analysis, most of alpha-2-macroglobulin N-glycans are biantennary complex type, thus we chose Endo F3 for partial deglycosylation. The search results of the acquired CID MS/MS and MS³ spectra revealed N-glycosylation sites as shown in Table 3.2. Ten partial glycopeptides associated with five N-glycosylation sites were identified with or without core fucosylation. It is interesting to note that the MS/MS spectra of glycopeptides with a GlcNAc-Fuc attached were dominated by the neutral loss of the core fucose²² where little peptide fragmentation was achieved. Further

fragmentation of the most abundant defucosylated ion (MS^3) generated fragment ions similar those from the MS/MS of their non-fucosylated counterparts. An example is shown in Figure 3.6 where the two glycopeptides selected include fucosylated (m/z 884.6, Figure 3.6a) and non-fucosylated (m/z 811.6, Figure 3.6b) with the same peptide backbone. The MS/MS spectrum of the fucosylated peptide at m/z 884.6 in Figure 3.6a does not provide much peptide backbone fragmentation and the most intense ion at m/z 811.6 results from neutral loss of the core fucose. MS^3 of this neutral loss peak at m/z 811.6 in Figure 3.6a yields extensive peptide backbone fragments (y_2 - y_{11} and b_2 - b_{11}), providing both the peptide sequence and glycosylation site information. The MS^3 spectrum of 884.6- \rightarrow 811.6 has great similarity with the MS/MS spectrum of the non-fucosylated counterpart (m/z 811.6) which is shown in Figure 3.6c. Thus, by combining Endo F3 partial deglycosylation with MS/MS of non-fucosylated glycopeptides or MS^3 of fucosylated glycopeptides, we can obtain the information on peptide sequence, glycosylation sites and attachment sites of core fucosylation. Combined with the ETD analysis of N-glycopeptides, six out of eight N-glycosylation sites were identified (N70, N396, N410, N869, N991 and N1424). However, site N55 and site N247 were not detected with the Endo F3 approach possibly because they were occupied by other N-glycans than the four most abundant complex type glycans identified in this study, and those less abundant glycans were not cleaved due to the substrate specificity of Endo F3. Expected masses of partial glycopeptides associated with site N70 were observed with the Endo F3 method, but Xcorr scores were not sufficient for confident assignment, even though some expected theoretical MS/MS peaks were found by manual inspection.

3.4 Conclusion

In this work, an integrated LC-MS/MS strategy was developed for comprehensive identification of both site-specific glycosylation and core fucosylation of glycoproteins. Using this workflow, a volume of only 10 μ L of human serum sufficed for two LC-MS/MS analyses on glycopeptides treated with and without Endo F3 treatment. There are three major aspects in this assay including N-glycan analysis, CID/ETD MS/MS analysis of intact glycopeptides, and CID MS/MS analysis of Endo F3 treated glycopeptides.

Glycopeptide CID/ETD MS/MS analysis identified three N-glycosylation sites, N70, N396 and N1424, with four glycoforms found for each site. Endo F3 cleaves the majority of glycan moieties with only the core GlcNAc or GlcNAc-Fuc attached to the peptide backbones, thus reducing the mass spectra complexity and provides the unique ability to identify site-specific core fucosylation. The advantages of Endo F3 assisted mass spectrometric analysis were successfully used to reveal five glycosylation sites at N396, N410, N869, N991 and N1424. With this combined approach, we identified a total of six N-glycosylation sites with site-specific glycosylation or core fucosylation patterns revealed.

Figures

Figure 3. 1. Strategy to characterize site-specific N-glycosylation of alpha-2-macroglobulin from human serum.

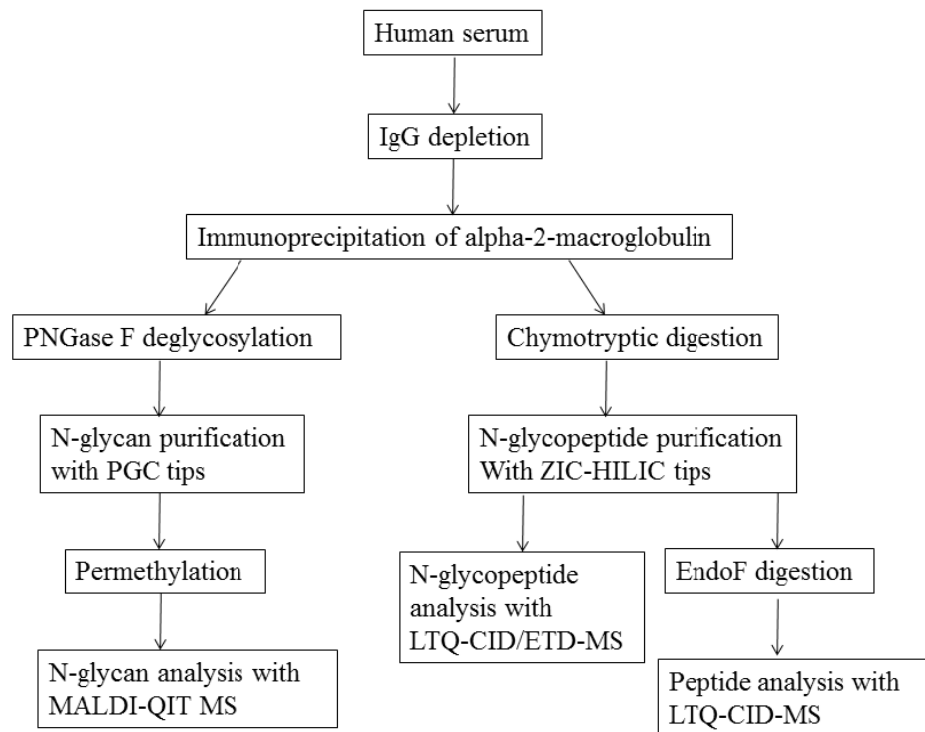
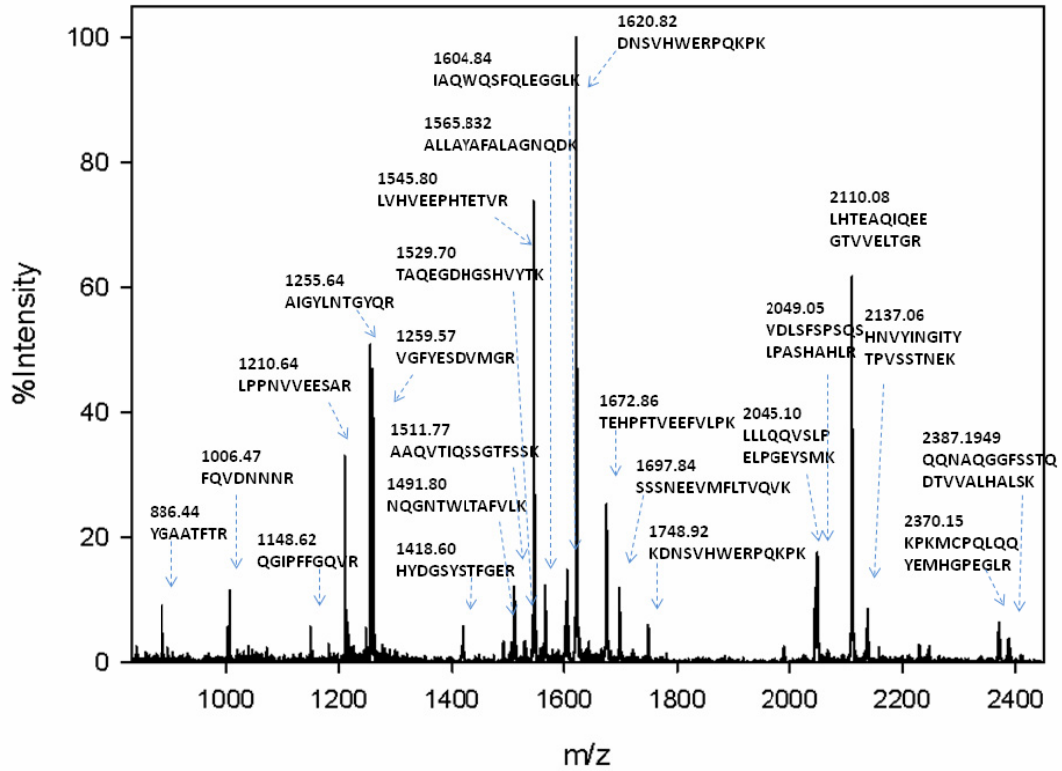
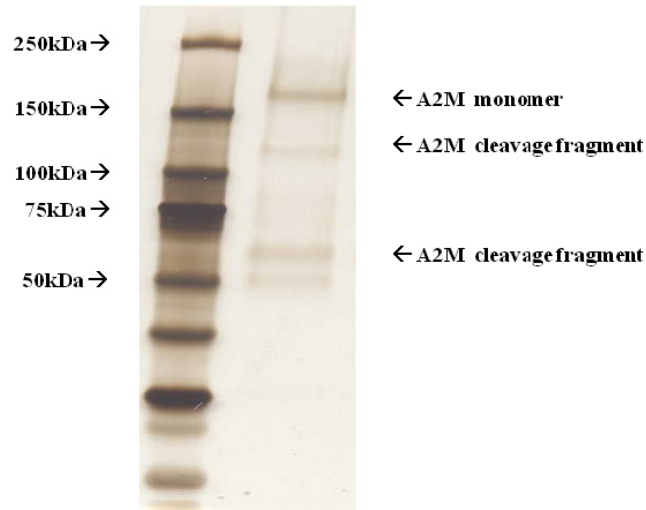


Figure 3. 2. (a) Peptide finger print of immunoprecipitated alpha-2-macroglobulin after on plate tryptic digestion followed by MALDI-QIT-TOF MS analysis. (b) SDS-PAGE followed by silver staining of 1/10 of immunoprecipitated alpha-2-macroglobulin.

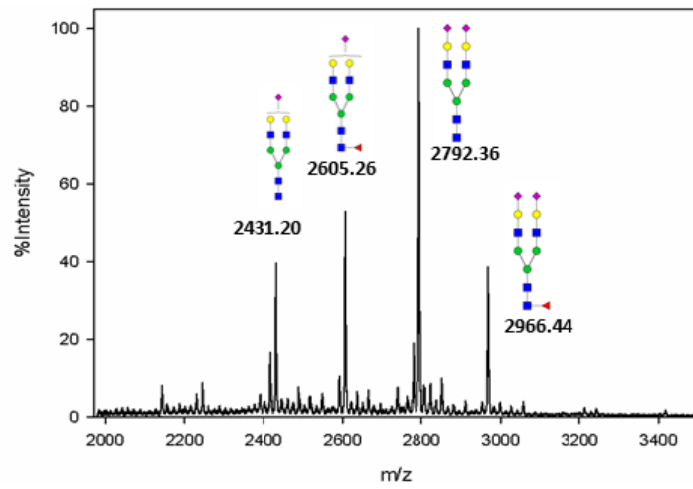


(a)

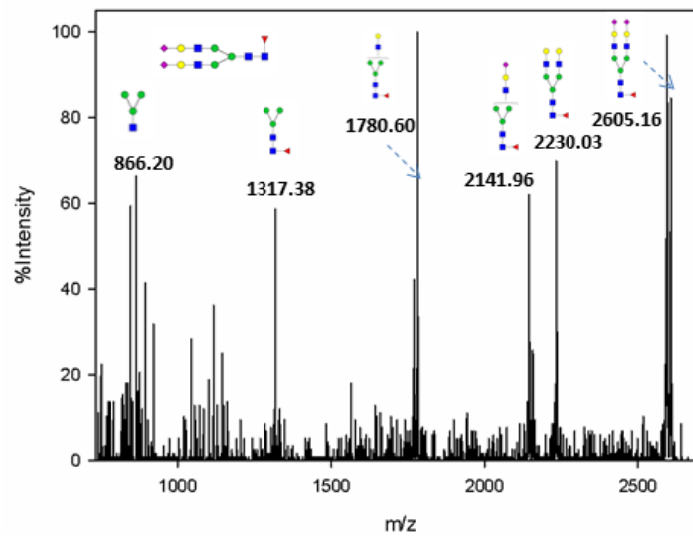


(b)

Figure 3. 3. (a) Representative mass spectrum of N-glycan profile of alpha-2-macroglobulin of a normal serum sample. Four biantennary complex type glycans were identified. (b) MS/MS spectrum of a fucosylated biantennary glycan at m/z 2966.44.

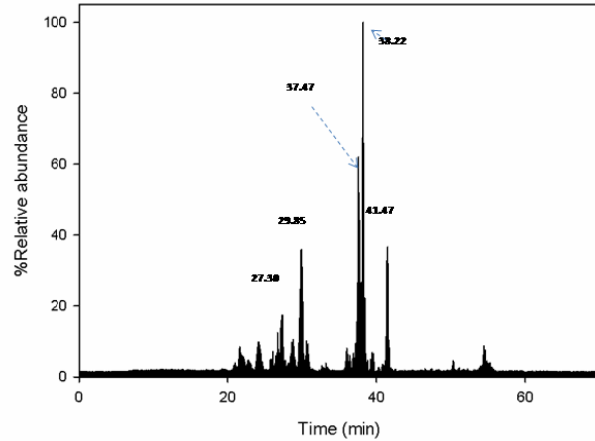


(a)

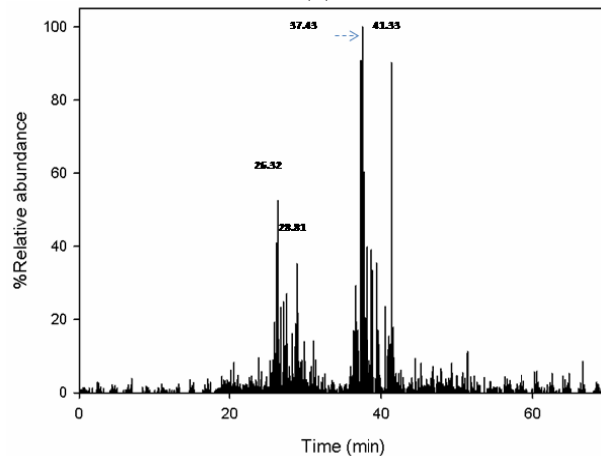


(b)

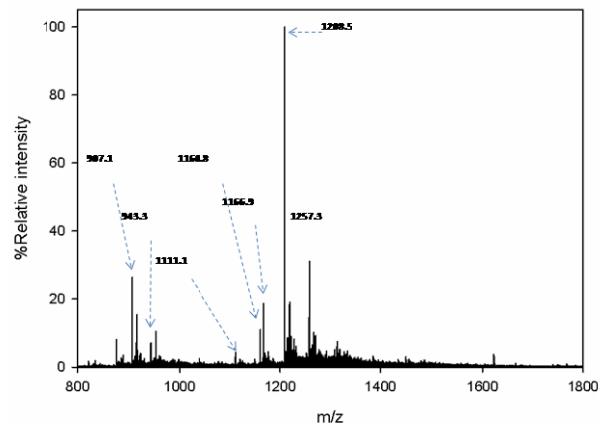
Figure 3. 4. (a) Base peak chromatogram of LC-MS analysis of chymotryptic N-glycopeptides purified by ZIC-HILIC. The peaks at 27.30 min, 29.85 min, 37.47 min, 38.22 min and 41.47 min are the primary glycopeptides. (b) Extracted ion chromatogram of oxonium fragment ion at m/z 366 (GlcNAc-Gal) from CID MS/MS spectra. (c) Summed mass spectrum at 37.60-38.60 min revealed some major N-glycopeptides.



(a)

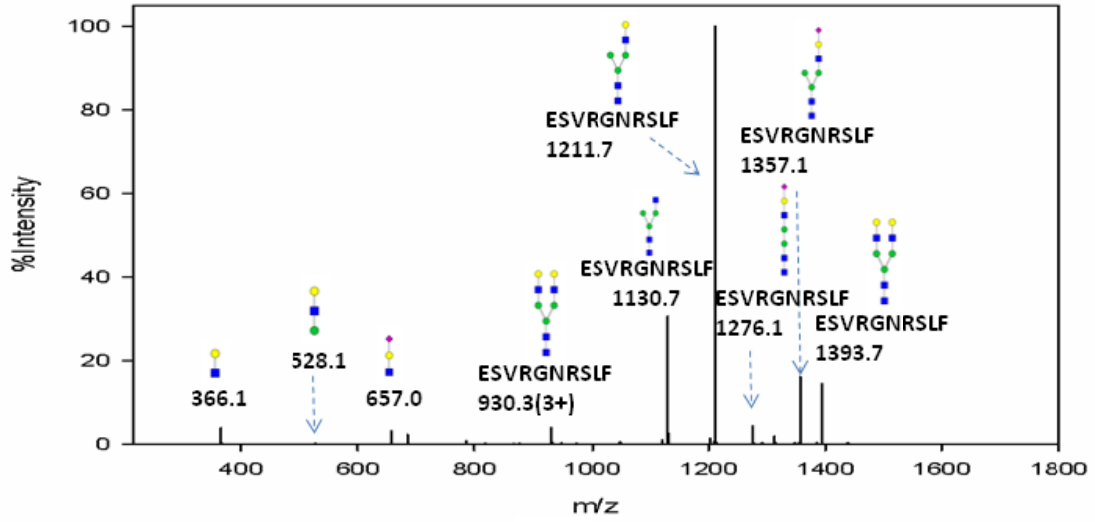


(b)

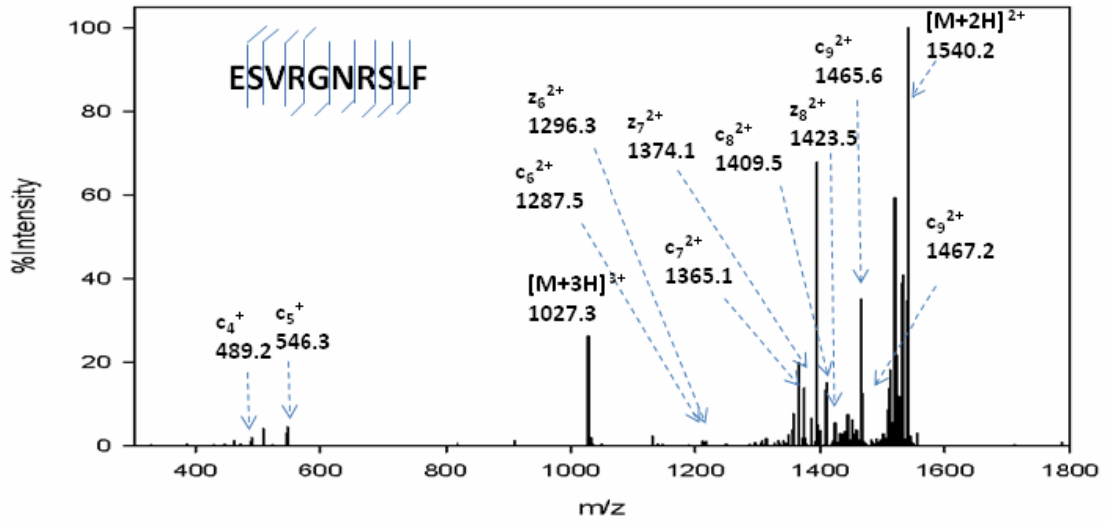


(c)

Figure 3. 5. (a) CID MS/MS spectrum of the glycopeptide at m/z 1026.7 (sequence ESVRGNRS LF). Glycosidic bond cleavages were observed, resulting in b,y type ions. (b) ETD MS/MS spectrum of the same glycopeptide. c,z type ions were observed with the intact glycan structure.



(a)



(b)

Figure 3. 6. (a) CID MS/MS of the fucosylated glycopeptide (SN(+GlcNAc-Fuc)ATTDEHGLVQF) at m/z 884.6. The major fragment at m/z 811.6 is the product ion after neutral loss of fucose from the precursor ion. Minimal peptide backbone fragmentation was observed. (b) CID MS3 of m/z 811.6 from (a) showed extensive fragmentation along the peptide backbone, providing both peptide sequence information and the glycosylation site. (c) CID MS/MS of the non-fucosylated glycopeptide (SN(+GlcNAc)ATTDEHGLVQF) at m/z 811.6. (c) has great similarity with (b), revealing presence of both fucosylation and non-fucosylation at the same site.

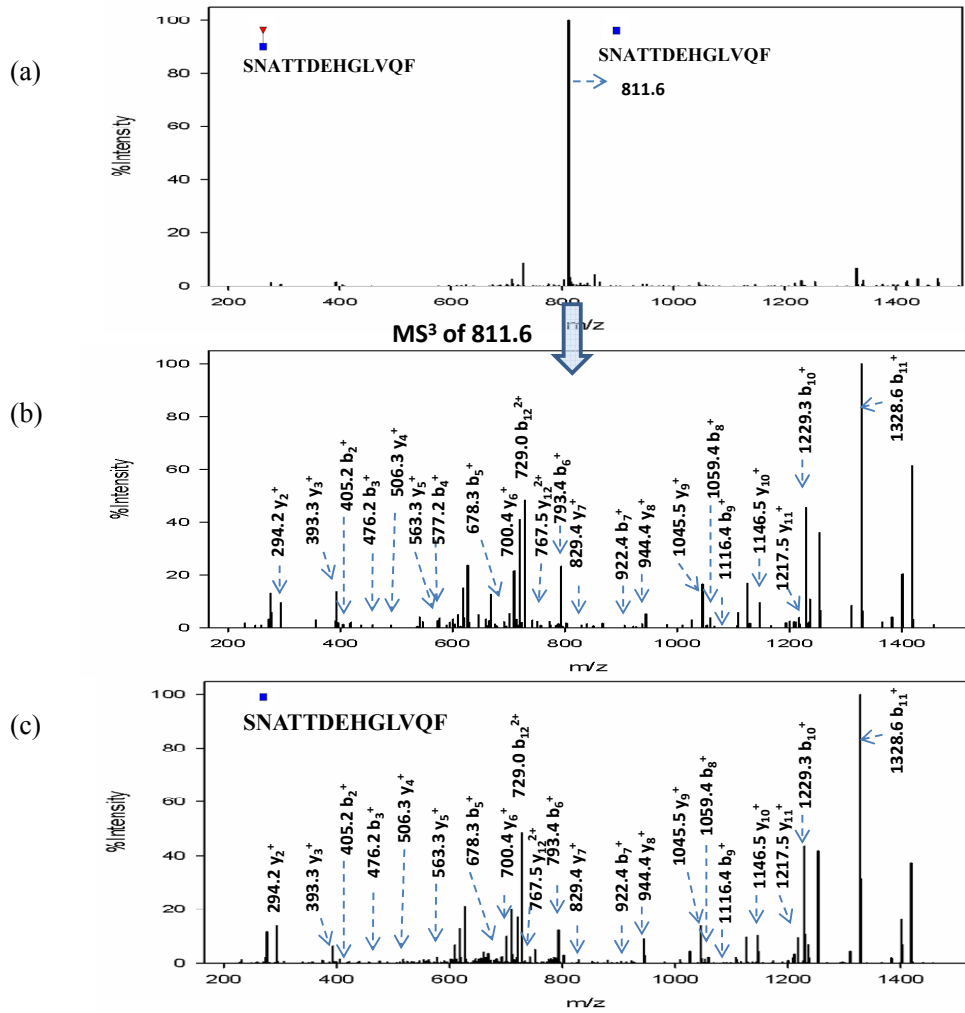


Table 3. 1. N-glycopeptides identified in LC-CID/ETD-MS/MS analysis.

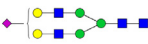



















Site	Peptide sequence	Glycan	<i>m/z</i>	Charge	r.t. (min)
70	ESVRGNRSLF		1026.7	+3	28.77
	ESVRGNRSLF		1075.3	+3	29.03
	ESVRGNRSLF		843.4	+4	30.02
	ESVRGNRSLF		1172.2	+3	29.03
	ESVRGNRSLFTDL		852.8	+4	30.51
396	SNATTDEHGLVQF		1111.1	+3	36.60
	SNATTDEHGLVQF		1160.8	+3	37.27
	SNATTDEHGLVQF		870.5	+4	37.43
	SNATTDEHGLVQF		1208.5	+3	38.82
	SNATTDEHGLVQF		907.1	+4	37.94
	SNATTDEHGLVQF		1257.3	+3	38.15
	SNATTDEHGLVQF		943.3	+4	38.51
	YSNATTDEHGL		1040.7	+3	38.70
	SNATTDEHGL		1083.8	+3	29.93
	YSNATTDEHGLVQF		947.6	+4	38.90
	1424	IYLDKVSNQTL		1069.5	+3
IYLDKVSNQTL			1118.0	+3	36.91
IYLDKVSNQTL			1166.9	+3	36.85
IYLDKVSNQTL			911.8	+4	37.02
LDKVSNQTL			1123.7	+3	37.58

Table 3. 2. Glycosylation sites identified by CID MS/MS of Endo F3 treated chymotryptic N-glycopeptides of alpha-2-macroglobulin.

Site	Sequence	Modification	m/z	r.t. (min)
396	S <u>N</u> ATTDEHGLVQF	GlcNAc	811.4 (+2)	25.76
		GlcNAc-Fuc	884.6 (+2)	25.60
410	S <u>I</u> NTTNVMGTSL	GlcNAc	720.9 (+2)	28.62
		GlcNAc-Fuc	794.0 (+2)	28.55
869	AVTPKSLG <u>N</u> VNF	GlcNAc	725.4 (+2)	26.89
991	DY <u>L</u> NETQQL	GlcNAc	664.1 (+2)	26.79
1424	LDKVS <u>N</u> QTL	GlcNAc	611.0 (+2)	19.12
		GlcNAc-Fuc	684.0 (+2)	19.07
1424	IYLDKVS <u>N</u> QTL	GlcNAc	749.0 (+2)	25.65
		GlcNAc-Fuc	822.0 (+2)	25.70

References

- (1) Dube, D. H.; Bertozzi, C. R., Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **2005**, 4, (6), 477-488.
- (2) Zhao, J.; Simeone, D. M.; Heidt, D.; Anderson, M. A.; Lubman, D. M., Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: Application to pancreatic cancer serum. *J Proteome Res* **2006**, 5, (7), 1792-1802.
- (3) Nakano, M.; Nakagawa, T.; Ito, T.; Kitada, T.; Hijioka, T.; Kasahara, A.; Tajiri, M.; Wada, Y.; Taniguchi, N.; Miyoshi, E., Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: A novel approach for the development of tumor markers. *Int J Cancer* **2008**, 122, (10), 2301-2309.
- (4) An, H. J.; Froehlich, J. W.; Lebrilla, C. B., Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr. Opin. Chem. Biol.* **2009**, 13, (4), 421-426.
- (5) Hogan, J. M.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A., Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J Proteome Res* **2005**, 4, (2), 628-632.
- (6) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F., The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta, Proteins Proteomics* **2006**, 1764, (12), 1811-1822.
- (7) Wang, D. D.; Hincapie, M.; Rejtar, T.; Karger, B. L., Ultrasensitive Characterization of Site-Specific Glycosylation of Affinity-Purified Haptoglobin from Lung Cancer Patient Plasma Using 10 μ m i.d. Porous Layer Open Tubular Liquid Chromatography-Linear Ion Trap Collision-Induced Dissociation/Electron Transfer Dissociation Mass Spectrometry. *Anal Chem* **2011**, 83, (6), 2029-2037.
- (8) Segu, Z. M.; Hussein, A.; Novotny, M. V.; Mechref, Y., Assigning N-Glycosylation Sites of Glycoproteins Using LC/MSMS in Conjunction with Endo-M/Exoglycosidase Mixture. *J Proteome Res* **2010**, 9, (7), 3598-3607.
- (9) Hagglund, P.; Bunkenborg, J.; Elortza, F.; Jensen, O. N.; Roepstorff, P., A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* **2004**, 3, (3), 556-566.
- (10) Zhang, W.; Wang, H.; Zhang, L.; Yao, J.; Yang, P., Large-scale assignment of N-glycosylation sites using complementary enzymatic deglycosylation. *Talanta* **2011**, 85, (1), 499-505.
- (11) Wang, H.; Zhang, W.; Zhao, J.; Zhang, L.; Liu, M.; Yan, G.; Yao, J.; Yu, H.; Yang, P., N-glycosylation pattern of recombinant human CD82 (KAI1), a tumor-associated membrane protein. *J Proteomics* **2012**, 75, (4), 1375-1385.
- (12) Miyoshi, E.; Moriwaki, K.; Nakagawa, T., Biological function of fucosylation in cancer biology. *J. Biochem. (Tokyo)*. **2008**, 143, (6), 725-729.
- (13) Barrett, A. J.; Starkey, P. M., INTERACTION OF ALPHA2-MACROGLOBULIN WITH PROTEINASES - CHARACTERISTICS AND SPECIFICITY OF REACTION, AND A HYPOTHESIS CONCERNING ITS MOLECULAR MECHANISM. *Biochem. J.* **1973**, 133, (4), 709-&.
- (14) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D., Human plasma N-glycoproteome analysis by immunoaffinity subtraction,

- hydrazide chemistry, and mass spectrometry. *J Proteome Res* **2005**, 4, (6), 2070-2080.
- (15) Bunkenborg, J.; Pilch, B. J.; Podtelejnikov, A. V.; Wisniewski, J. R., Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **2004**, 4, (2), 454-465.
- (16) Chen, R.; Jiang, X. N.; Sun, D. G.; Han, G. H.; Wang, F. J.; Ye, M. L.; Wang, L. M.; Zou, H. F., Glycoproteomics Analysis of Human Liver Tissue by Combination of Multiple Enzyme Digestion and Hydrazide Chemistry. *Journal of Proteome Research* **2009**, 8, (2), 651-661.
- (17) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **2003**, 21, (6), 660-666.
- (18) Zhao, J.; Qiu, W.; Simeone, D. M.; Lubman, D. M., N-linked glycosylation profiling of pancreatic cancer serum using capillary liquid phase separation coupled with mass spectrometric analysis. *J Proteome Res* **2007**, 6, (3), 1126-1138.
- (19) Lin, Z.; Simeone, D. M.; Anderson, M. A.; Brand, R. E.; Xie, X.; Shedden, K. A.; Ruffin, M. T.; Lubman, D. M., Mass Spectrometric Assay for Analysis of Haptoglobin Fucosylation in Pancreatic Cancer. *J Proteome Res* **2011**, 10, (5), 2602-2611.
- (20) Wada, Y.; Azadi, P.; Costello, C. E.; Dell, A.; Dwek, R. A.; Geyer, H.; Geyer, R.; Kakehi, K.; Karlsson, N. G.; Kato, K.; Kawasaki, N.; Khoo, K. H.; Kim, S.; Kondo, A.; Lattova, E.; Mechref, Y.; Miyoshi, E.; Nakamura, K.; Narimatsu, H.; Novotny, M. V.; Packer, N. H.; Perreault, H.; Peter-Katalinic, J.; Pohlentz, G.; Reinhold, V. N.; Rudd, P. M.; Suzuki, A.; Taniguchi, N., Comparison of the methods for profiling glycoprotein glycans - HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **2007**, 17, (4), 411-422.
- (21) Alley, W. R., Jr.; Mechref, Y.; Novotny, M. V., Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom* **2009**, 23, (1), 161-170.
- (22) Zhao, Y.; Jia, W.; Wang, J. F.; Ying, W. T.; Zhang, Y. J.; Qian, X. H., Fragmentation and Site-Specific Quantification of Core Fucosylated Glycoprotein by Multiple Reaction Monitoring-Mass Spectrometry. *Anal Chem* **2011**, 83, (22), 8802-8809.
- (23) Wang, Y. H.; Wu, S. L.; Hancock, W. S., Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap-Fourier transform mass spectrometry. *Glycobiology* **2006**, 16, (6), 514-523.
- (24) Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R., Chemical Deamidation: A Common Pitfall in Large-Scale N-Linked Glycoproteomic Mass Spectrometry-Based Analyses. *J Proteome Res* **2012**, 11, (3), 1949-1957.
- (25) Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T., Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* **2003**, 21, (6), 667-672.

Chapter 4

A Method for Label-free Relative Quantitation of site-specific Core-fucosylation by LC-MS/MS

4.1 Introduction

Glycosylation is one of the most prevalent protein posttranslational modifications and it is estimated that over 50% of proteins are glycosylated.¹ Protein glycosylation usually reflects the physiological and pathological environment of cells, and its alterations have been found to be involved in pathogenesis of multiple diseases including cancers.²⁻⁶ Given its importance in disease diagnostics, understanding of protein glycosylation has lagged behind achievements in genomics and proteomics mostly due to the complexity in glycosylation and lack of efficient and sensitive characterization or quantification methods. One protein may have several glycosylation sites (glycosylation heterogeneity) with multiple possible glycans at each site (glycosylation microheterogeneity), leading to an extremely low abundance of a particular glycopeptide. In most glycosylation studies, N-glycans are released from the peptide backbone by peptide N-glycosidase (PNGase F), followed by separate qualitative and quantitative analysis of glycans^{7, 8} and deglycosylated peptides^{9, 10}. This approach only provides aggregate glycosylation information, but cannot identify site-specific glycosylation patterns which may serve as potential disease markers. There are relatively few studies on intact glycopeptides and even fewer cases of quantification. Label free strategies based on intensities of precursor ions have been used often in the literature.^{11, 12}

Core-fucosylation, where the fucose attach to innermost GlcNAc via an α -1,6 linkage, is a subtype of N-glycosylation which has attracted much research interest due to its potential as a cancer biomarker^{6, 13, 14}. Notably, the core-fucosylation level of alpha-fetoprotein (AFP-L3) is an FDA-approved diagnostic tool for liver cancer¹⁵. To overcome the technical difficulty of intact glycopeptide analysis and to target site-specific core-fucosylation quantification, we developed a method which utilizes endoglycosidase F3 (Endo F3) to partially deglycosylate the glycopeptides and retain only the core GlcNAc and fucose. The qualitative aspect of this method has been used to identify glycosylation sites on both core-fucosylated individual proteins¹⁶ and from complex protein mixtures such as human plasma¹⁷⁻¹⁹. In terms of the quantitative aspect of this approach, analyses on serum protein core-fucosylation level change as potential hepatocellular carcinoma markers have been performed using precursor intensity-based quantification with differential dimethylation methods²⁰.

In this study, we developed a label-free LC-MS/MS methodology for relative quantification of core-fucosylation at specific glycosylation sites based on precursor ion intensities of Endo F3 treated glycopeptides. Using the property that partially deglycosylated glycopeptides with the same peptide backbone approximately co-elute by reverse phase liquid chromatography, the fucosylation ratio at a particular glycosylation site can be calculated through dividing the peak area of fucosylated peptide by the peak area of non-core-fucosylated counterpart with the same peptide sequence. This is a relative quantification method which eliminates the need of internal standards. We applied this assay in the quantitative study of alpha-2-macroglobulin (A2MG) site-specific core-fucosylation where A2MG was immunoprecipitated from human serum.

The fucosylation ratios at three sites N396, N410 and N1424 of human A2MG were calculated from 20 normal controls, 20 pancreatic cancer patients and 20 chronic pancreatitis patients. This assay could be utilized to monitor core-fucosylation changes in other proteins or protein mixtures and be used to identify aberrations in protein core-fucosylation on the onset of diseases.

4.2. Experimental Section

4.2.1 Serum samples

A total of 60 human serum samples (20 normal volunteers, 20 chronic pancreatitis patients and 20 pancreatic cancer patients) were included in this study. Demographic information and cancer stage information are shown in Table 4.1. All human normal serum samples were provided by the University Hospital, Ann Arbor, Michigan, and all chronic pancreatitis and pancreatic cancer samples were provided by the University of Pittsburgh. All samples were collected according to IRB approved protocols. The samples were aliquoted and stored in a -80 °C freezer. All samples were frozen and thawed only once.

4.2.2 Sample preparation

Sample preparation was detailed in previous work.¹⁶ Briefly, alpha-2-macroglobulin was immunoprecipitated (antibody from Abcam, Cambridge, MA) from 10 μ L human serum, reduced with dithiothreitol, alkylated with iodoacetamide, and digested with chymotrypsin (Promega, Madison, WI). Glycopeptides were enriched using ZIC-HILIC microtips (Protea, Morgantown, WV), and further partially deglycosylated with Endo F3 (QABio, Palm Desert, CA). The resulting peptides were desalted with C₁₈ ZipTips (Millipore, Billerica, MA) prior to LC-MS/MS analysis.

4.2.3 NanoLC-LTQ-MS analysis of partially deglycosylated peptides

NanoLC-MS/MS conditions were described in previous work.¹⁶ In summary, a C₁₈ capillary column (75 µm i.d. × 10 cm; 5 µm particles) was used for LC separation, and gradient elution was performed using a Paradigm MG4 micropump system (Michrom Biosciences, Auburn, CA) with a flow rate at 300 nL/min. Mobile phase A was 2% acetonitrile with 1% acetic acid in water and mobile phase B was 5% water with 1% acetic acid in acetonitrile. The analytical gradient lasted 70 min where composition of solvent B rose from 5% to 60% in 35 min, followed by a washing and equilibration step where solvent B increased to 95% in 1 min and held for 4 min, finally returned to 5% B in 0.1 min and held for 30 min.

An ESI-LTQ mass spectrometer (Thermo Fisher Scientific, San Jose, CA) operated in positive ion mode was used for analysis. The ESI spray voltage and capillary voltage were set at 2.2 kV and 45 V respectively. CID fragmentation was performed at 35% of the normalized collision energy. The mass spectra were acquired in a data-dependent manner. Following a full scan in the mass range of m/z 400 to 1800, CID MS/MS was performed on the most intense ion, and the most intense fragment ion in MS/MS was selected for MS³ fragmentation. CID MS/MS and MS³ were performed on the first to the fourth most intense ions from the survey MS scan.

4.2.4 Quantitative data analysis

Database search was performed using Proteome Discoverer (version 1.1, Thermo Fisher Scientific, San Jose, CA) with SEQUEST using the following search parameters: (1) fixed modification: cysteine carbamidomethylation (+57.0 Da); (2) variable modification: methionine oxidation (+16.0 Da), addition of N-acetylglucosamine (+203.1

Da) or N-acetylglucosamine-fucose (+349.1 Da) to asparagine; (3) missed cleavages allowed: three; (4) enzyme specificity: chymotrypsin at F, W, Y, and L; (5) peptide ion tolerance (average mass): 1.4 Da; (6) fragmentation ion tolerance (average mass): 0.8 Da. The MS/MS data was searched against SWISS-PROT *Homo sapiens* database (Release 2010_10, downloaded on Nov 2, 2010).

The partially deglycosylated glycopeptides identified were quantified using the peak areas from the extracted ion chromatogram (XIC). Peak area integration was performed manually using XCalibur Qual Browser (version 2.1) with the following parameters: (1) precursor peaks were extracted with a 1 Da (± 0.5 Da) mass window; (2) scan filter was set as full MS; (3) boxcar type of smoothing with 7 points was enabled; (4) peak detection algorithm was Genesis; (5) signal-over-noise ratio threshold was set at 3.

4.2.5 Statistical analysis of fucosylation ratios

The fucosylation ratio for each glycosylation site was calculated as:

$$\text{Fucosylation ratio} = \frac{A_{\text{XIC-Fucosylated}}}{A_{\text{XIC-Non-core-fucosylated}}}$$

where $A_{\text{XIC-Fucosylated}}$ and $A_{\text{XIC-Non-core-fucosylated}}$ are the peak areas of extracted precursor ion chromatograms of core-fucosylated peptide and non-core-fucosylated peptide with the same sequence respectively.

Five fucosylation ratios were obtained for each sample and log transformed. Each log ratio was compared across samples of different disease states using ANOVA with Prism 5 (GraphPad, La Jolla, CA). Correlation analysis of core-fucosylation patterns between different peptide sequences were performed using R.

4.3. Results and Discussion

In this report, we describe an endoglycosidase F assisted label-free LC-MS/MS assay for quantifying protein site-specific core-fucosylation level change. This assay as shown in Figure 4.1 is composed of the following steps: (1) identification of the partially deglycosylated peptide m/z 's and retention times; (2) peak area integration of the extracted ion chromatogram (XIC) of the partially deglycosylated peptides; (3) calculation of the core-fucosylation ratio of each peptide sequence (which is defined as the peak area of the fucosylated peptide divided by the peak area of the approximately coeluting non-core-fucosylated counterpart sharing the same peptide backbone); (4) comparison of core-fucosylation ratios between samples.

4.3.1 Selection of enzyme for proteolysis

To determine the proteolytic enzyme for this study, we compared the numbers of N-glycosylation sites identified using different enzymes including trypsin, chymotrypsin, trypsin/GluC and LysC/GluC. As shown in Table 4.2, we were able to identify the most glycosylation sites (N396, N410, N869, N991 and N1424) with chymotrypsin, while trypsin/GluC and LysC/GluC provided 3 sites respectively and only one site was identified with trypsin. In combining all four digestion methods, all eight potential N-glycosylation sites of A2MG were identified. It is interesting to note that sites N869 and N991 were not observed to have core-fucosylation, suggesting variation in the level of core-fucosylation between sites. In our study, we chose to use chymotrypsin because of the number of sites identified. The focus of this study is developing a LC-MS/MS assay for quantification of protein core-fucosylation rather than comprehensively quantifying alpha-2-macroglobulin core-fucosylation at every glycosylation site.

LTQ-Orbitrap MS, a high-resolution instrument was used for comparison of identification results, giving the same results (data not shown).

4.3.2 Label-free quantitative analysis

While most studies of site-specific glycosylation quantification are performed on intact glycopeptides, the high heterogeneity, the low concentration and low ionization efficiency of individual glycopeptides make quantification and data analysis more challenging. In this study, we focused on quantifying the extent of protein core-fucosylation, which is an important subtype of disease related glycosylation. By using endoglycosidase F, most of the glycan is removed, leaving only core GlcNAc or GlcNAc-Fuc (if there is core-fucosylation) attached to the peptide backbone. Hence the glycopeptides are divided into two categories – core fucosylated and non-core-fucosylated – and the proportion of core-fucosylation can be quantified as described in section 3.3. This approach has three major advantages compared to intact glycopeptide analysis: (1) ionization efficiency of partially deglycopeptides was significantly increased; (2) the sensitivity of the assay was improved due to the signal stacking effect; (3) data analysis was greatly simplified.

Label-free quantification strategy based on the area under curve (AUC) of the precursor ions has been used in quantification of both the protein abundance and protein post-translational modifications including glycosylation¹² and phosphorylation²¹. In this study, peak areas of 10 partially deglycosylated peptides (5 pairs, both fucosylated and non-core-fucosylated) from each of the 60 samples (20 normal, 20 chronic pancreatitis, and 20 pancreatic cancer) were extracted. A low-resolution instrument could be used in this case because of the low sample complexity due to effective isolation of a single

protein and purification of glycopeptides (only 16 peptides were identified in a typical database search). Over 90% of partially deglycosylated peptides quantified were identified using database search, and the peak areas of the very few unidentified peptides were extracted based on the protein correlation profiling method using match in both retention time and precursor mass.

We also attempted MRM for quantification purpose. There are two folds of technical difficulty in the MRM assay development for the core-fucosylation quantification. The first obstacle is the lack of standard peptides with core fucosylation. This makes it difficult to optimize the collision conditions and selection of transitions, which is crucial in the case of core-fucosylated peptides due to their high tendency to generate only neutral loss ions in CID MS/MS. Three transitions are normally used for a peptide to improve the specificity. However, only the neutral-loss fragment may be used for the core-fucosylated peptides, making the assay less specific.

The second obstacle is the differential isotopic labeling. We tried two isotopic labeling strategies, enzyme-catalyzed ^{18}O labeling and Tandem Mass Tag (TMT) labeling. Chymotrypsin was used to incorporate two ^{18}O in the C-terminus carboxylic groups. Due to the low-specificity of chymotrypsin, the incorporation was not consistently successful due to limitation in the kinetics. As shown in Figure 4.2, the labeling was efficient for peptides ending with F with most abundant ion species as the 2^{18}O labeled and very low non-labeled or singly-labeled signals. However, the labeling was not efficient for peptides ending with L. The most intense ions were singly labeled or nonlabeled. The low 2^{18}O incorporation rates of these peptides ending with L made quantification implausible. Unlike 2^{18}O which preserve the properties of the peptides,

TMT labeling which modifies the primary amine (N-terminal or K, R and H) changes their chemical physical characteristics. It was observed that the precursor ion intensities significantly decreased by around 10 folds after TMT labeling as demonstrated in Figure 4.3 for unclear reasons, which greatly lowers the sensitivity of the assay. Due to the failures in labeling strategies, we decided to use label-free method.

4.3.3 Fucosylation ratio indices

Fucosylation degree, the relative abundance of fucosylated glycans in a mixture of oligosaccharides, has been used to provide a numerical description of fucosylation changes of N-glycans in various disease states^{6,22}. In this study, core-fucosylation ratios were obtained on glycopeptides, and quantification of site-specific core-fucosylation was reported for the first time. The prior work quantified only the core-fucosylated peptides but not the non-core-fucosylated counterparts,²⁰ hence it was unclear if the quantitative change occurred in the protein amount or extent of core-fucosylation. This work quantifies the core-fucosylation only without influence of protein amount.

It was found that the core-fucosylated and non-core-fucosylated peptides with the same peptide sequence approximately coelute because the retention behavior was largely determined by the peptide backbone as shown in Figure 4.4. Among five pairs of deglycosylated peptides detected, the mean absolute differences in retention times of the four pairs of peptides (panel 1: 811.5/884.4, panel 2: 893.2/966.1, panel 4: 611.1/684.0, and panel 5: 749.3/822.3) are 6 seconds, while the mean absolute difference in retention times of one pair of peptides (panel 3: 720.9/793.9) is wider at 30 seconds. Similar chromatographic behavior is the base for fucosylation ratio construction.

Although chymotrypsin demonstrated reproducible cleavage in evaluating the

reproducibility, missed cleavages are common in chymotrypsin digestion, leading to multiple peptide sequences at the same glycosylation site as shown in Table 4.2. To compare the fucosylation degree at a particular site, it is important to include all the corresponding peptide sequences. It is further demonstrated in Section 3.4 that the core-fucosylation patterns of peptides corresponding to the same N-glycosylation site are highly correlated.

Reproducibility, an essential parameter in a quantitative assay is evaluated in order to demonstrate the ability of the assay to obtain the fucosylation ratios reproducibly in repeated analyses performed on different days. Four aliquots of the same serum sample were analyzed on different days and revealed a coefficient of variation (CV) less than 11% (Table 4.3).

4.3.4 Statistical analysis

Fucosylation ratios of 5 pairs of peptides originating from 3 glycosylation sites were compared between 60 samples categorized as three disease states (20 normal, 20 chronic pancreatitis, and 20 pancreatic cancer), generating a 5×60 data matrix as shown in. The fucosylation ratios range from 0.1 to 20, revealing the high diversity in distribution of core-fucosylation. To meet the normality assumption of ANOVA test, a log₁₀ transformation was performed to correct the deviation from normality for most populations except for the population of chronic pancreatitis with peptide sequence YSNATTDEHGLVQF (*m/z*: 893.0/966.0) due to one extremely low value. This value was considered as an outlier and was removed to ensure the normality of the data

To study the fucosylation ratio changes between disease groups, a one-way ANOVA F-test was applied to compare the means of three different populations, and multiple

comparisons were performed with Tukey's test. In general, a statistically significant fucosylation decrease was observed in all five pairs of peptides (m/z 811.5/884.5, m/z 893.0/966.0, m/z 720.9/794.0, m/z 749.2/822.2, and m/z 611.0/684.0) corresponding to three sites (site 396, 410 and 1424) of alpha-2-macroglobulin in pancreatic diseases (including pancreatic cancer and chronic pancreatitis) compared to normal controls, while the pancreatic cancer and chronic pancreatitis were undistinguishable. Scatterplots in Figure 4.5 provide a graphical comparison of the log fucosylation ratios at five pairs of peptides in the 60 samples, and the brackets show the statistically significant comparison (p -value<0.05). It is obvious that core-fucosylation decreases in both chronic pancreatitis and pancreatic cancer at all sites. The decrease of core-fucosylation was observed in global N-glycan analysis as shown in Figure 4.6 where N-glycans were purified, permethylated and analyzed as described in Chapter 2.

To quantify the fucosylation decrease between the normal control and the chronic pancreatitis, or the normal control and the pancreatic cancer, the effect sizes were calculated, and the decrease of fucosylation is demonstrated as the fold change in the unit of standard deviation. Cohen's d was used for effect size calculation:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{S}$$

where \bar{x}_1 and \bar{x}_2 are the means of fucosylation ratios at a particular peptide sequence for disease groups involved in comparison (normal vs. cancer, or normal vs. pancreatitis). As shown in Table 4, alpha-2-macroglobulin shows 1-2 standard deviation decrease in fucosylation ratio in chronic pancreatitis and pancreatic cancer patients for all the 3 glycosylation sites studied. A2MG core-fucosylation ratios of chronic pancreatitis and pancreatic cancer decreased to 6% to 65% of the normal volunteer values.

Furthermore, Pearson's r correlation analysis was performed in order to evaluate the correlation of core-fucosylation ratios between peptides corresponding to the same glycosylation site, and between peptides corresponding to different sites (Figure 4.7). Pearson's r compares the covariance of two populations again with the variances of both populations. Its value range from -1 to 1; a value between 0.5 and 1 indicates positive correlation, a value between -1 and -0.5 indicates negative correlation, and a value between -0.5 and 0.5 indicates low or zero correlation. As we expected, the core-fucosylation ratios of peptide sequences corresponding to the same glycosylation site are well correlated ($r \sim 0.8$), indicating high internal consistency of the quantification results. Site 396 and Site 1424 are highly correlated in core-fucosylation patterns while core-fucosylation ratios of Site 410 are not correlated with other two sites. It is unclear how the fucosyltransferase recognize specific N-glycosylation sites and modify them differently. It could recognize the amino sequences around the site or recognize the conformational structure of A2MG.

4.4. Conclusion

In this report, an endoglycosidase-assisted label-free LC-MS/MS assay which relatively quantifies site-specific core-fucosylation was developed and applied to human serum alpha-2-macroglobulin. The core-fucosylation ratio at each glycosylation site was calculated as the peak area of the fucosylated peptide divided by the peak area of the non-core-fucosylated peptide with the same peptide sequence. The assay was demonstrated to have inter-day reproducibility of less than 15%. The assay was utilized in a preliminary study of alpha-2-macroglobulin core-fucosylation changes in pancreatic

diseases including pancreatic cancer and chronic pancreatitis. Core-fucosylation levels were found to decrease at sites 396, 410 and 1424 in both chronic pancreatitis and pancreatic cancer compared to normal controls. Further exploration with larger sample cohort is needed for validation. This generic strategy could be effectively applied to monitor the aberration in site-specific core-fucosylation of other glycoproteins for cancer biomarker studies.

Figures

Figure 4. 1. Workflow of the study

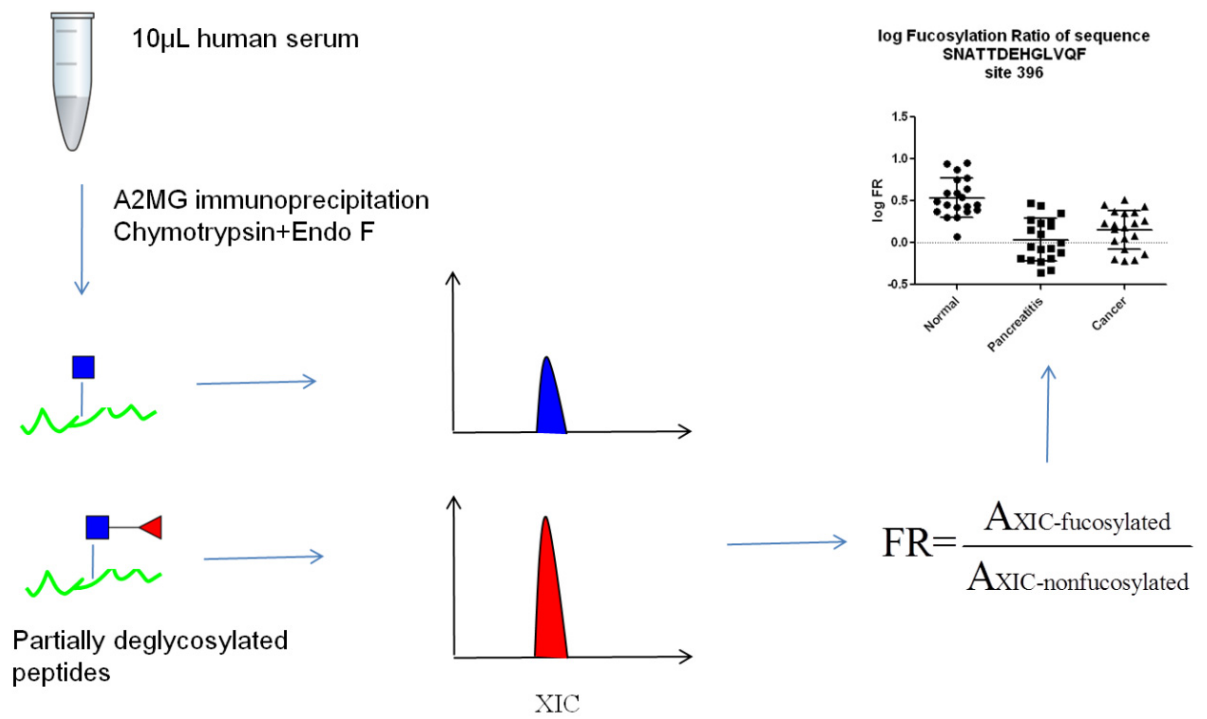


Figure 4. 2. Inconsistent labeling efficiency of chymotrypsin-catalyzed ^{18}O labeling. Top trace: labeling pattern of peptide SNATTDEHGLVQF. Bottom trace: labeling pattern of peptide IYLDKVSQTL.

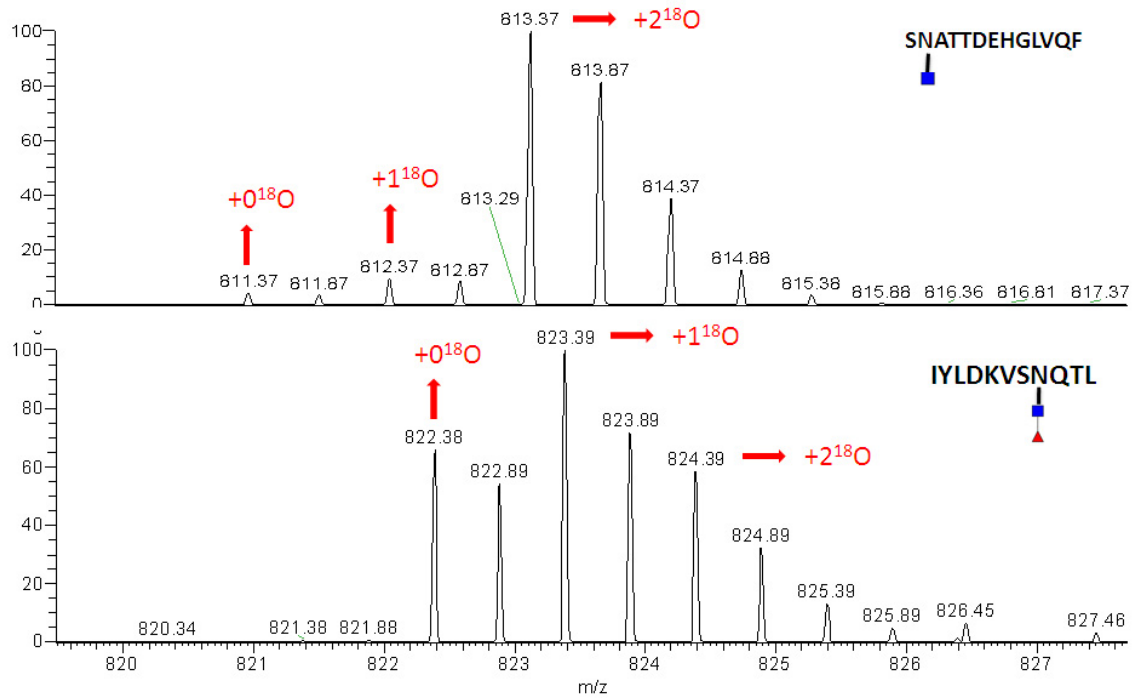


Figure 4. 3. XIC of the peptides before and after TMT labeling. Top: before TMT labeling; bottom: after TMT labeling. (RT: retention time, in mins, AA: peak area, SN: signal over noise ratio, BP: base peak m/z)

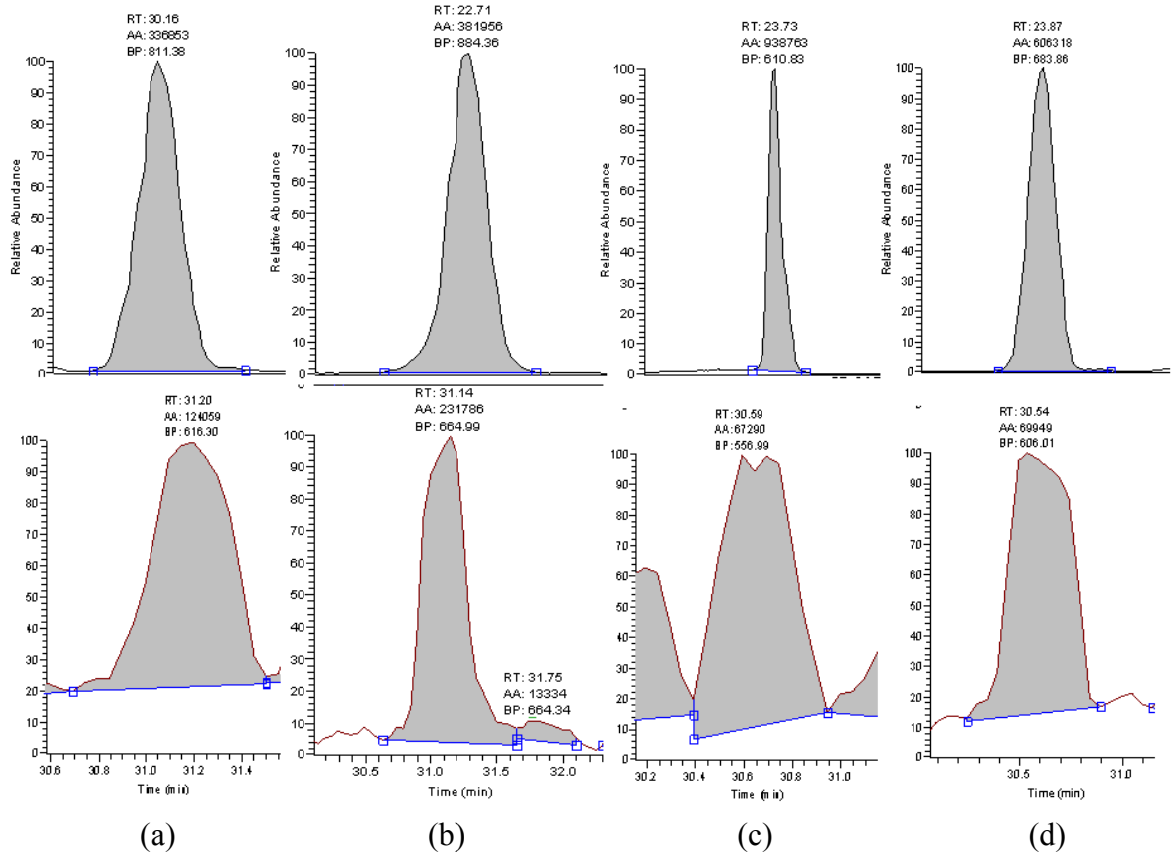


Figure 4. 4. Extracted ion chromatograms (XIC) of 5 pairs of partially deglycosylated peptides. The top chromatogram of each panel is the non-core-fucosylated peptide, and the bottom chromatogram in the same panel is the fucosylated counterpart with the same peptide backbone. (RT: retention time, in mins, AA: peak area, SN: signal over noise ratio, BP: base peak m/z).

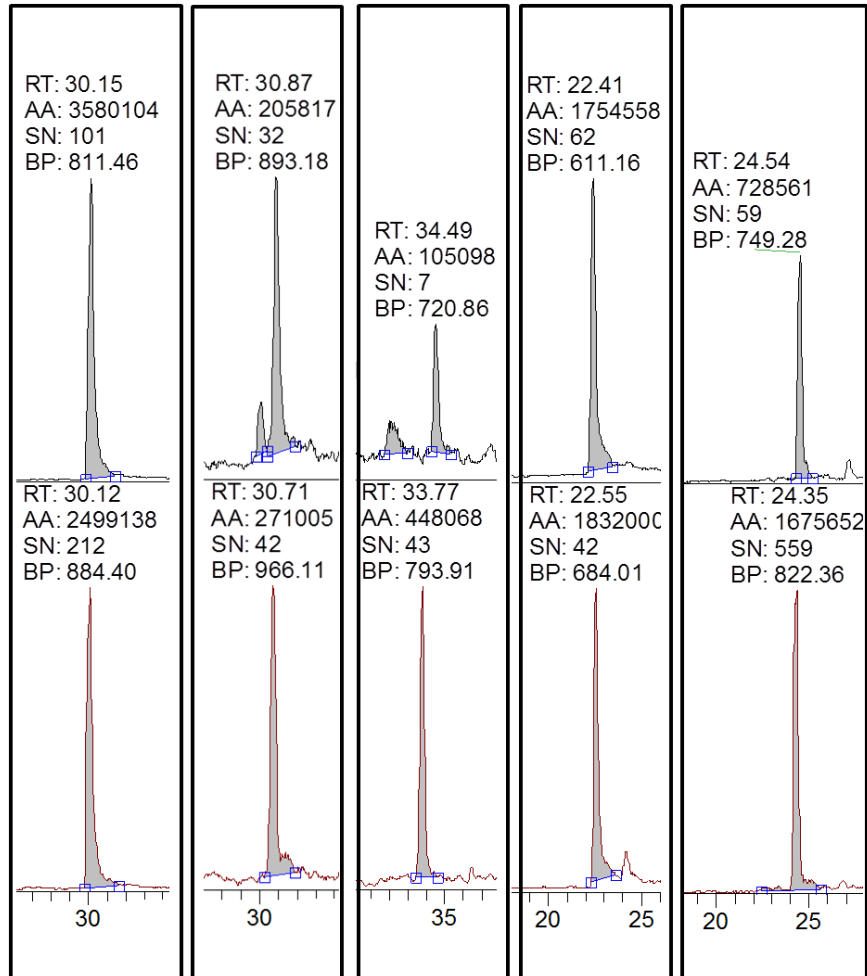


Figure 4. 5. Scatterplots of log fucosylation ratios at 5 peptide sequences corresponding to 3 glycosylation sites. Twenty samples per disease states (normal, chronic pancreatitis and pancreatic cancer) were included in comparison, except in Figure 4.3(b) where only 19 data points are shown in the pancreatitis group due to removal of an outlier.

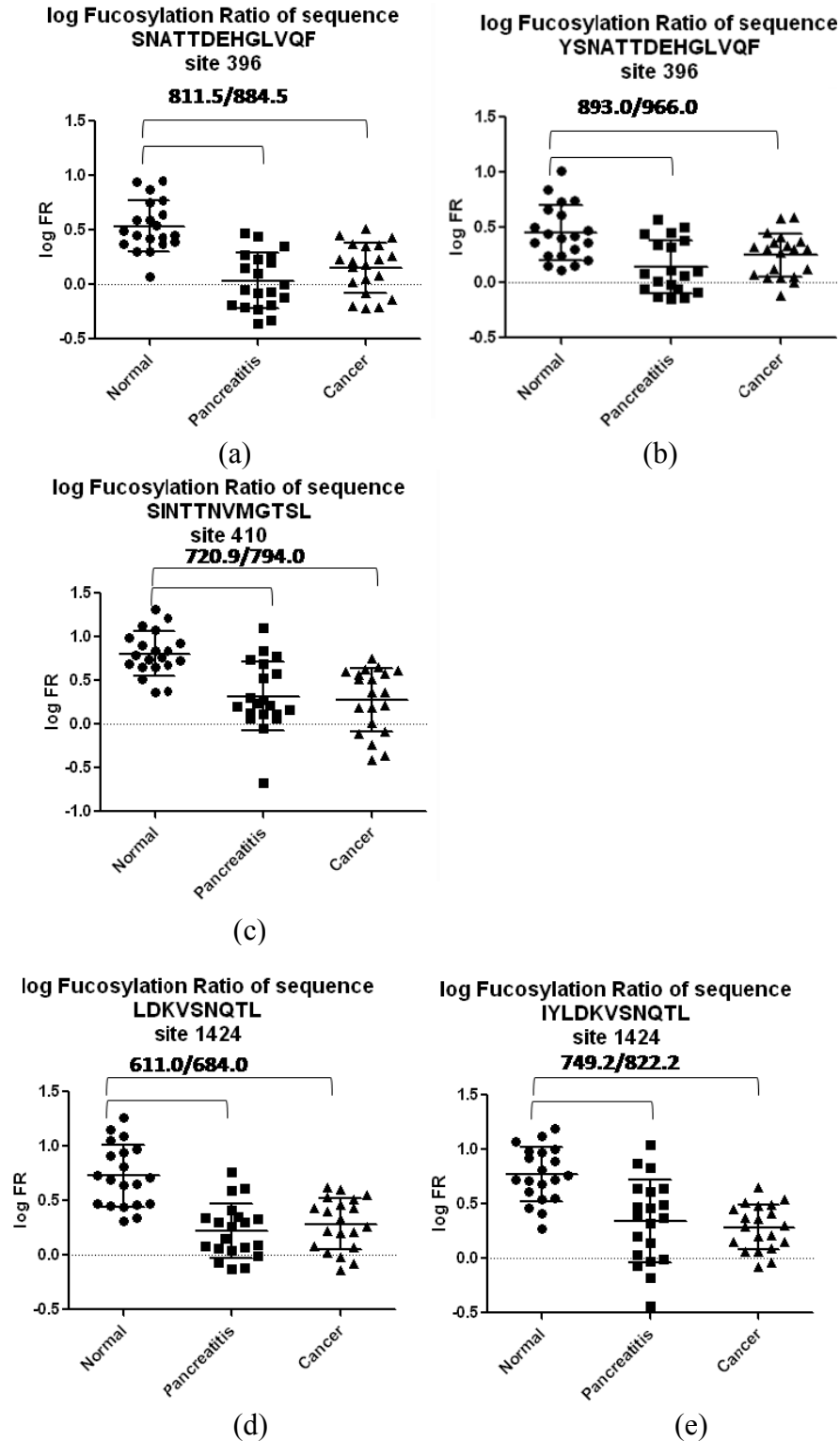


Figure 4. 6. Comparison of A2MG N-glycan profiles between normal control (top), chronic pancreatitis (middle) and pancreatic cancer (bottom).

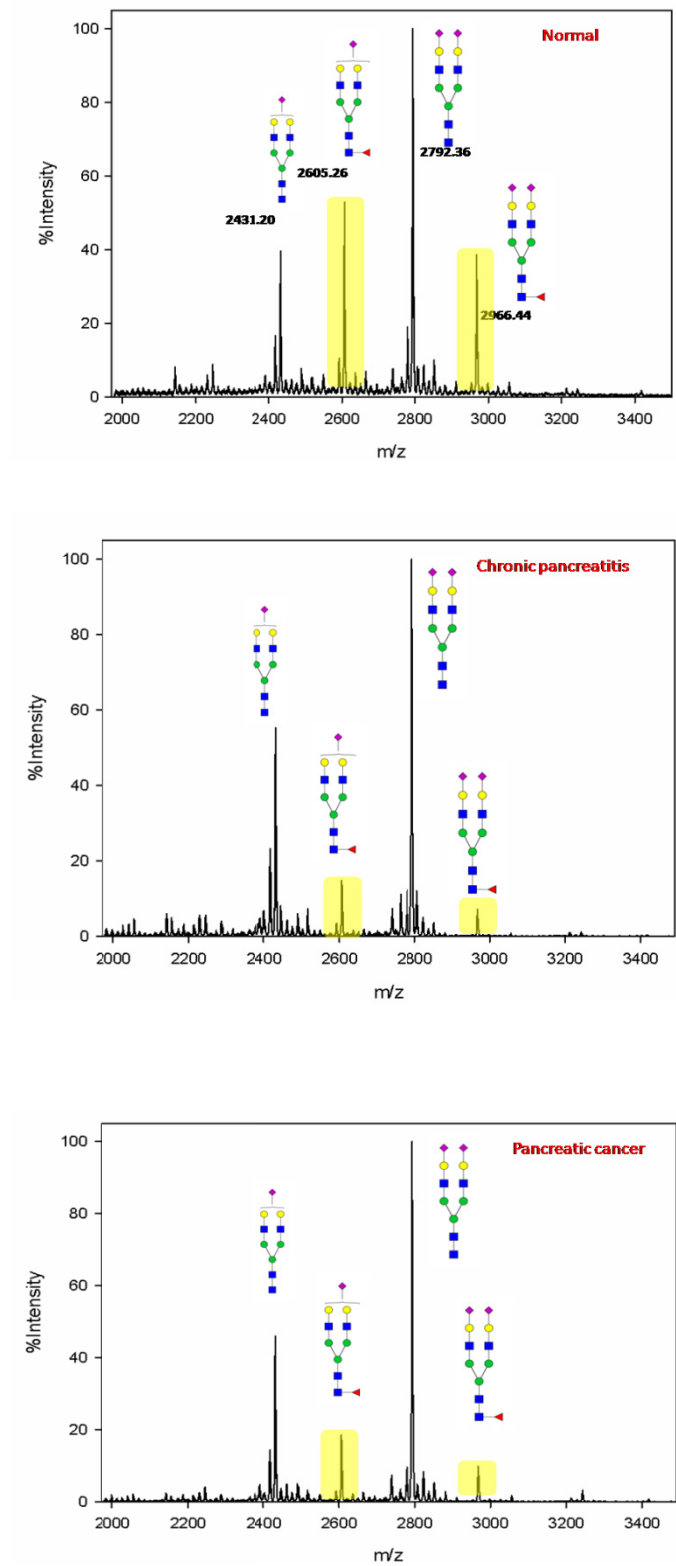
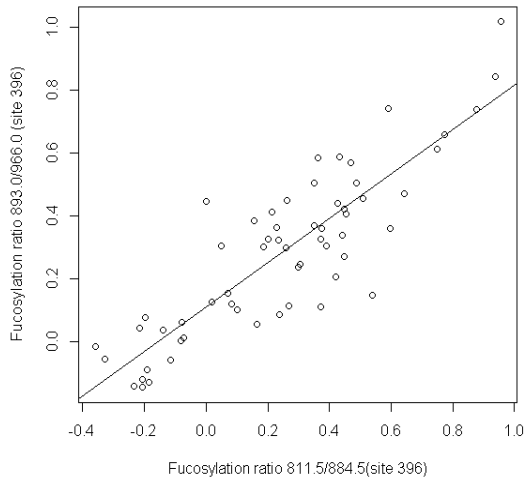
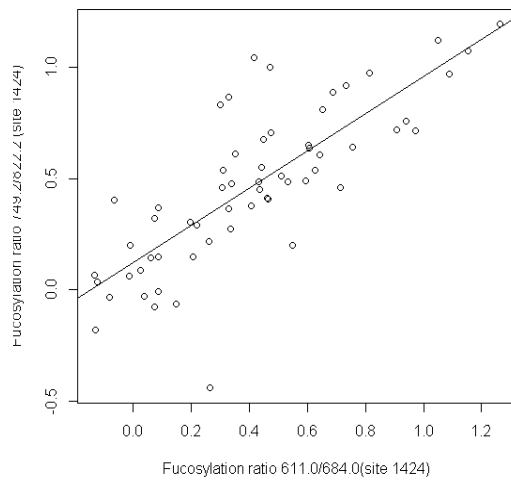


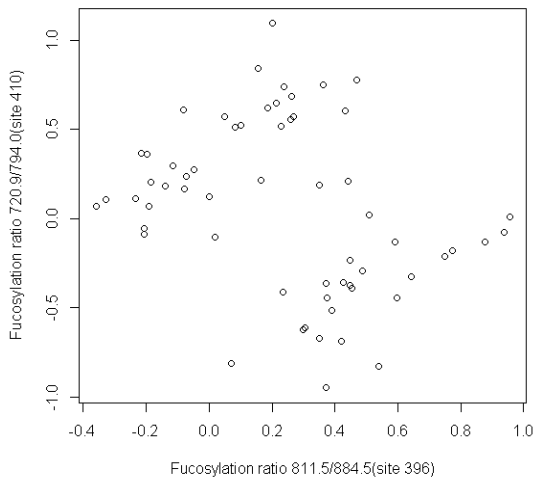
Figure 4. 7. Correlation analysis of fucosylation ratios between peptide sequences of the same site ((a) and (b)), and between peptide sequences of different sites ((c)-(j)). Pearson's r higher than 0.5 indicates strong correlation. Fucosylation ratios of peptide sequences corresponding to the same site (site 396: 811.5/884.5 and 893.0/966.0 in (a) and site 1424: 611.0/684.0 and 749.2/822.2 in (b)). Fucosylation ratios of site 396 and site 410 ((c),(d)), and site 1424 and site 410 ((e) and (f)) are not correlated. Fucosylation ratios of site 396 and site 1424 are correlated ((g)-(j)).



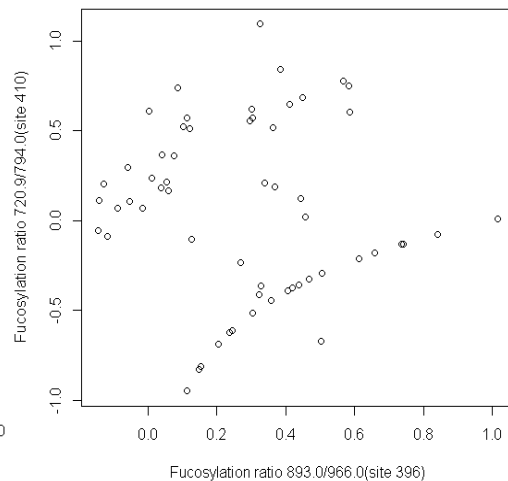
(a)



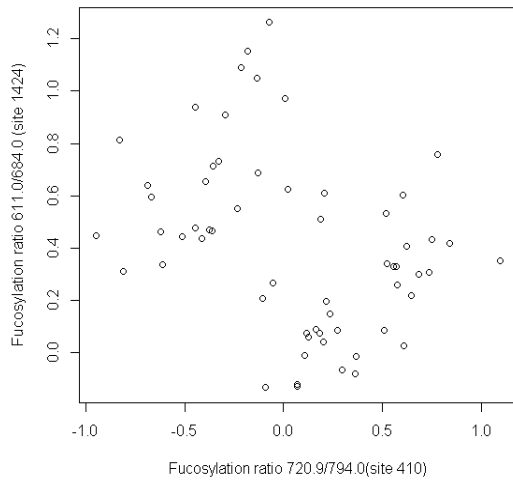
(b)



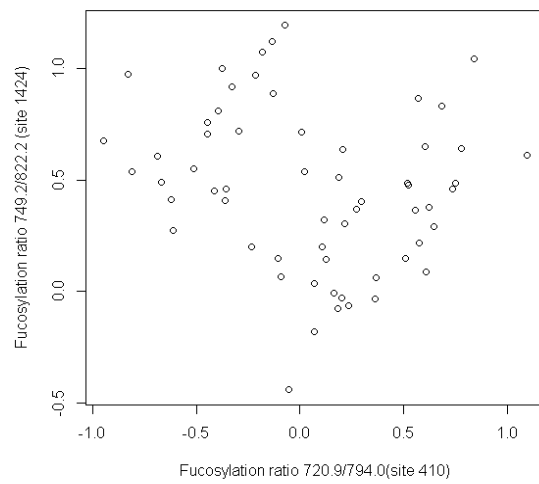
(c)



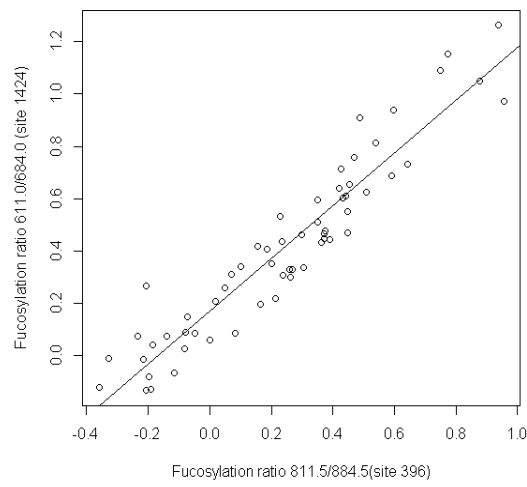
(d)



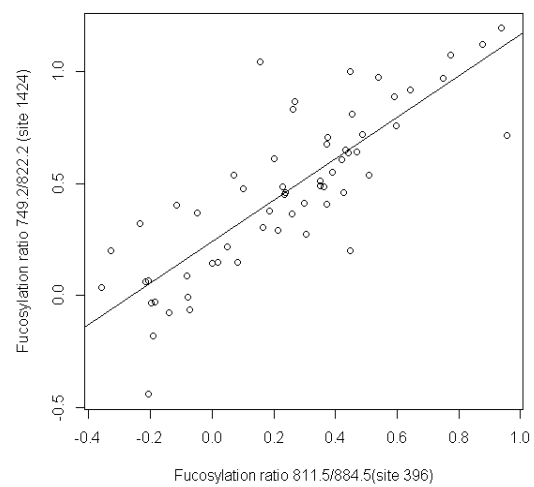
(e)



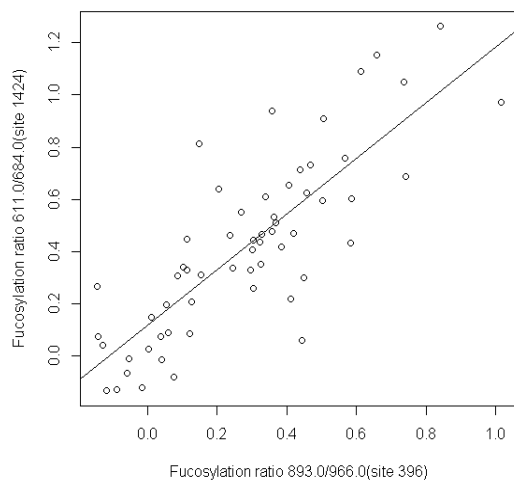
(f)



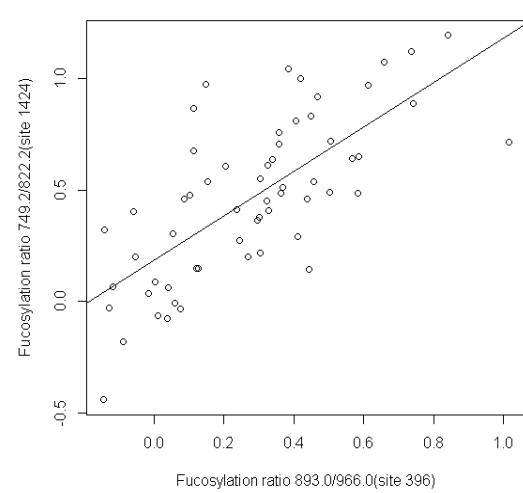
(g)



(h)



(i)



(j)

Table 4. 1. Demographic information and cancer stage information of human serum samples enrolled in this study.

	Normal	Chronic pancreatitis	Pancreatic cancer
Average(years)			
<50	5	6	1
51-60	9	7	5
61-70	6	6	7
>71	0	1	7
Gender			
Male	8	13	13
Female	12	7	7
Cancer stage			
IIA			2
IIB			6
III			6
IV			6

Table 4. 2. Glycosylation sites identified with different proteolysis enzymes or enzyme combinations. Glycosylation site is labeled as N. NF and F indicate fucosylated glycopeptide and non-core-fucosylated glycopeptide respectively. NA indicates that no fucosylated peptide was detected.

Enzyme	Site	Sequence	Peptide Mass (NF/F)
Trypsin	55	GCVLLSYL <u>N</u> ETVTVSASLESVR	867.7/916.2
Chymotrypsin	991	VLDYL <u>N</u> ETQQL	770/NA
		DYL <u>N</u> ETQQL	664.1/NA
	396	S <u>N</u> ATTDEHGLVQF	811.4/ 884.6
		YS <u>N</u> ATTDEHGLVQF	893.0/966.0
	410	S <u>I</u> NTTNVMGTSL	720.9/794.0
	869	AVTPKSLGNV <u>N</u> F	725.7/NA
1424	LDKVS <u>N</u> QTL	611.0/684.0	
	IYLDKVS <u>N</u> QTL	749.2/822.2	
Trypsin/GluC	247	M <u>N</u> VSVCGLYTYGK	848.2/921.4
	396	ANYYS <u>N</u> ATTDE	726.3/799.6
	869	SLGNV <u>N</u> FTVSAE	721.1/NA
LysC/GluC	396	VIFIRGNEANYYS <u>N</u> ATTDE	1189.5/1262.6
	70	SVRG <u>N</u> RSLFTDLEAE	633.5/682.2
	55	GCVLLSYL <u>N</u> E	685.8/758.8

Table 4. 3. Reproducibility test of the assay for four aliquots of the same normal serum sample processed on four different days. The values in row 2-5 are fucosylation ratios.

	Site 396	Site 396	Site 410	Site 1424	Site 1424
	811.5/884.5	893.0/996.0	720.9/794.0	611.0/684.0	749.2/822.2
Replicate1	5.94	4.54	16.42	11.89	14.18
Replicate2	5.76	5.15	15.09	9.66	14.04
Replicate3	4.67	4.85	14.70	10.22	11.64
Replicate4	5.20	4.52	13.27	9.52	12.05
RSD	10.68%	6.22%	8.71%	10.52%	10.15%

Table 4. 4. Statistical summary of core-fucosylation ratios in different disease states. SD stands for standard deviation

	Site 396	Site 396	Site 410	Site 1424	Site 1424
	811.5/884.5	893.0/996.0	720.9/794.0	611.0/684.0	749.2/822.2
SD(normal)	0.2316	0.2507	0.2529	0.2853	0.2502
SD(pancreatitis)	0.2531	0.2418	0.3927	0.2483	0.3831
SD(Cancer)	0.2309	0.1955	0.3639	0.2365	0.2058
Normal mean	0.5353	0.4484	0.8078	0.7280	0.7694
Pancreatitis mean	0.0338	0.1447	0.3187	0.2861	0.3406
Cancer mean	0.1540	0.2468	0.2768	0.4742	0.2843
Effect size(N-CP)	2.0693	1.2332	1.5152	1.6563	1.3542
Effect size(N-C)	1.5756	0.9220	1.4036	1.0470	1.6475
Percentage(CP/N)	6.31%	32.27%	39.45%	39.30%	44.27%
Percentage(C/N)	28.77%	55.04%	34.27%	65.14%	36.95%

References

- (1) Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica Et Biophysica Acta-General Subjects* **1999**, 1473, (1), 4-8.
- (2) Tousi, F.; Hancock, W. S.; Hincapie, M., Technologies and strategies for glycoproteomics and glycomics and their application to clinical biomarker research. *Analytical Methods* **2011**, 3, (1), 20-32.
- (3) Zhao, J.; Patwa, T. H.; Qiu, W.; Shedden, K.; Hinderer, R.; Misek, D. E.; Anderson, M. A.; Simeone, D. M.; Lubman, D. M., Glycoprotein microarrays with multi-lectin detection: Unique lectin binding patterns as a tool for classifying normal, chronic pancreatitis and pancreatic cancer sera. *J Proteome Res* **2007**, 6, (5), 1864-1874.
- (4) Drake, P. M.; Cho, W.; Li, B.; Prakobphol, A.; Johansen, E.; Anderson, N. L.; Regnier, F. E.; Gibson, B. W.; Fisher, S. J., Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin. Chem.* **2010**, 56, (2), 223-36.
- (5) Meany, D. L.; Zhang, Z.; Sokoll, L. J.; Zhang, H.; Chan, D. W., Glycoproteomics for prostate cancer detection: changes in serum PSA glycosylation patterns. *J Proteome Res* **2009**, 8, (2), 613-9.
- (6) Lin, Z. X.; Simeone, D. M.; Anderson, M. A.; Brand, R. E.; Xie, X. L.; Shedden, K. A.; Ruffin, M. T.; Lubman, D. M., Mass Spectrometric Assay for Analysis of Haptoglobin Fucosylation in Pancreatic Cancer. *J Proteome Res* **2011**, 10, (5), 2602-2611.
- (7) Adamczyk, B.; Struwe, W. B.; Ercan, A.; Nigrovic, P. A.; Rudd, P. M., Characterization of fibrinogen glycosylation and its importance for serum/plasma N-glycome analysis. *J Proteome Res* **2013**, 12, (1), 444-54.
- (8) Barkauskas, D. A.; An, H. J.; Kronewitter, S. R.; de Leoz, M. L.; Chew, H. K.; de Vere White, R. W.; Leiserowitz, G. S.; Miyamoto, S.; Lebrilla, C. B.; Rocke, D. M., Detecting glycan cancer biomarkers in serum samples using MALDI FT-ICR mass spectrometry data. *Bioinformatics* **2009**, 25, (2), 251-7.
- (9) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **2003**, 21, (6), 660-666.
- (10) Zhao, J.; Qiu, W.; Simeone, D. M.; Lubman, D. M., N-linked glycosylation profiling of pancreatic cancer serum using capillary liquid phase separation coupled with mass spectrometric analysis. *J Proteome Res* **2007**, 6, (3), 1126-1138.
- (11) Ivancic, M. M.; Gadgil, H. S.; Halsall, H. B.; Treuheit, M. J., LC/MS analysis of complex multiglycosylated human alpha(1)-acid glycoprotein as a model for developing identification and quantitation methods for intact glycopeptide analysis. *Anal. Biochem.* **2010**, 400, (1), 25-32.
- (12) Ozohanics, O.; Turiak, L.; Puerta, A.; Vekey, K.; Drahos, L., High-performance liquid chromatography coupled to mass spectrometry methodology for analyzing site-specific N-glycosylation patterns. *J. Chromatogr.* **2012**, 1259, 200-212.
- (13) Wen, C. L.; Chen, K. Y.; Chen, C. T.; Chuang, J. G.; Yang, P. C.; Chow, L. P., Development of an AlphaLISA assay to quantify serum core-fucosylated E-cadherin as a metastatic lung adenocarcinoma biomarker. *J Proteomics* **2012**, 75, (13), 3963-3976.
- (14) Saldova, R.; Fan, Y.; Fitzpatrick, J. M.; Watson, R. W. G.; Rudd, P. M., Core fucosylation and alpha 2-3 sialylation in serum N-glycome is significantly increased in

- prostate cancer comparing to benign prostate hyperplasia. *Glycobiology* **2011**, 21, (2), 195-205.
- (15) Li, D.; Mallory, T.; Satomura, S., AFP-L3: a new generation of tumor marker for hepatocellular carcinoma. *Clin. Chim. Acta* **2001**, 313, (1-2), 15-19.
- (16) Lin, Z.; Lo, A.; Simeone, D. M.; Ruffin, M. T.; Lubman, D. M., An N-glycosylation Analysis of Human Alpha-2-Macroglobulin Using an Integrated Approach. *J Proteomics Bioinform* **2012**, 5, 127-134.
- (17) Zhang, W.; Wang, H.; Zhang, L.; Yao, J.; Yang, P. Y., Large-scale assignment of N-glycosylation sites using complementary enzymatic deglycosylation. *Talanta* **2011**, 85, (1), 499-505.
- (18) Jia, W.; Lu, Z.; Fu, Y.; Wang, H. P.; Wang, L. H.; Chi, H.; Yuan, Z. F.; Zheng, Z. B.; Song, L. N.; Han, H. H.; Liang, Y. M.; Wang, J. L.; Cai, Y.; Zhang, Y. K.; Deng, Y. L.; Ying, W. T.; He, S. M.; Qian, X. H., A Strategy for Precise and Large Scale Identification of Core Fucosylated Glycoproteins. *Molecular & Cellular Proteomics* **2009**, 8, (5), 913-923.
- (19) Hagglund, P.; Matthiesen, R.; Elortza, F.; Hojrup, P.; Roepstorff, P.; Jensen, O. N.; Bunkenborg, J., An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins. *J Proteome Res* **2007**, 6, (8), 3021-31.
- (20) Chen, R.; Wang, F.; Tan, Y.; Sun, Z.; Song, C.; Ye, M.; Wang, H.; Zou, H., Development of a combined chemical and enzymatic approach for the mass spectrometric identification and quantification of aberrant N-glycosylation. *J Proteomics* **2012**, 75, (5), 1666-74.
- (21) Steen, H.; Jeбанathirajah, J. A.; Springer, M.; Kirschner, M. W., Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proc Natl Acad Sci U S A* **2005**, 102, (11), 3948-53.
- (22) Imre, T.; Kremmer, T.; Heberger, K.; Molnar-Szollosi, E.; Ludanyi, K.; Pocsfalvi, G.; Malorni, A.; Drahos, L.; Vekey, K., Mass spectrometric and linear discriminant analysis of N-glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals. *J Proteomics* **2008**, 71, (2), 186-197.

Chapter 5

A Strategy for Profiling of Core-fucosylation in Human Serum Using Lectin Peptide Enrichment and HCD-MS/MS

5.1 Introduction

Protein N-glycosylation, where N-glycans attach to Asn in the sequon of Asn-Xaa-Ser/Thr (Xaa cannot be Pro) via an amide bond, is one of the most prevalent post-translational modifications. It is estimated that over 60% of human proteins are N-glycosylated¹. N-glycosylation regulates a variety of protein and cell functions including protein folding, signal transduction, cell recognition, cell metastasis and immunogenicity²⁻⁴. Aberrations of protein N-glycosylation have been found to associate with various types of cancer⁵⁻⁹, which may be used as cancer biomarkers or target for cancer immunotherapy. Such cancer-specific aberrations include increased branching of glycans⁶, increased or decreased fucosylation¹⁰, sialylation¹¹, and mannosylation¹². In this study, we focused on protein core-fucosylation, where the fucose attaches to the innermost N-acetylglucosamine (GlcNAc) via an α 1,6 linkage. Core-fucosylation has been found to be altered in different cancers^{11, 13-15}, and more notably, core-fucosylation level of alpha fetoprotein (AFP-L3 fraction) is an established FDA-approved liver cancer clinical diagnosis^{16, 17}, which demonstrates higher sensitivity and specificity compared to the level of total AFP. While the systematic studies on core-fucosylation are relatively few at present, a highly sensitive method for large-scale profiling and quantifying of

serum protein core-fucosylation is urgently desired for cancer biomarker discovery.

The first challenge in large-scale serum glycosylation study is the wide dynamic range of serum proteins. The 14 most abundant proteins such as albumin, immunoglobulin, transferrin and fibrinogen comprise approximately 95% of the protein mass, and many of them are glycoproteins which greatly mask the signals of the mid-low abundance glycoproteins. Hence, it is essential to remove these background proteins and focus the efforts on glycosylation analysis of mid-low abundance proteins where the studies are lacking. The second challenge is the low concentrations of glycopeptides relative to the nonglycopeptides, which leads to the necessity of efficient fractionation strategy. Hydrazide chemistry has been commonly used to enrich the glycoproteins or glycopeptides by converting the cis-diol groups of glycans to aldehydes and reacting with the hydrazide groups on a solid support. The immobilized glycopeptides or glycoproteins may be released by peptide N-glycosidase F digestion (PNGase F) which breaks the linkage between glycan moieties and peptide/protein backbones and subject to LC-MS/MS analysis¹⁸. However, this strategy is not specific for core-fucosylation, and furthermore irreversibly destroys the glycans, including the core fucose, making the site-specific core-fucosylation identification impossible. Another common strategy utilizes the lectins, a family of glycan binding proteins (GBP) which specifically recognize and bind to different carbohydrate epitopes^{19, 20}. The lectin *Lens culinaris* (LCA) which binds to Fuc α 1,6 GlcNAc, α -Man and α -Glc²¹ is the choice for selective isolation of core-fucosylated glycopeptides or glycoproteins. Due to the high heterogeneity of glycoforms per glycosylation site, deglycosylation is usually performed for profiling of glycosylation sites. However, the general strategy using PNGase F

removes the glycan moieties entirely from the peptides, leaving the core-fucosylation sites undistinguishable from the non-core-fucosylation sites. In this work, we performed partial deglycosylation with endoglycosidase F, which breaks the linkage between innermost GlcNAc, leaving a recognizable core fucose tag at the glycosylation sites that are core-fucosylated.

The recent developments of mass spectrometry have allowed more efficient and confident identification of glycosylation sites in intact glycopeptides. The new generation of Orbitrap mass analyzers has ultrahigh resolution over 240,000 FWHM and low-ppm mass accuracy. The Higher Energy C-trap Dissociation (HCD) has been implemented in recent glycopeptide analysis. HCD fragment measurement occurs in the Orbitrap, and overcomes the 1/3 cutoff rule in the traditional Collision Induced Dissociation (CID) method, which enables the detection of glycan oxonium ions in the low mass range for direct evidence of glycosylation²²⁻²⁵. In the particular field of core-fucosylation studies, researchers have reported CID MS³ method triggered by neutral loss of core fucose^{14, 26} for identification of core-fucosylation sites. However, the CID fragments are measured in a low-resolution manner, and no oxonium ions were detected, lowering the confidence of identification. In this work, we compared the performances of neutral loss-triggered CID MS³, HCD-only, and neutral-loss-triggered HCD, and identified HCD-only as the best strategy. Another merit of HCD is that quantitative analysis by iTRAQ labeling is made possible for relative quantification of the core-fucosylation changes in the onset of diseases.

In this study, we developed a workflow for profiling of serum protein core-fucosylation which combines removal of serum high-abundance proteins, LCA

lectin enrichment of glycopeptides, endoglycosidase partial deglycosylation, and LC-HCD-MS/MS analysis. The efficacy of this workflow was evaluated in 250 µg of serum proteins (~50 µL serum) with 14 most abundant proteins removed. 135 unique core-fucosylation sites were identified in 92 proteins. The quantitative aspect of this workflow was evaluated by 4-plex iTRAQ labeling which identified 81 core-fucosylation sites in 55 proteins and suggested satisfying quantification results

5.2 Materials and methods

All chemical reagents, if not noted, were purchased from Sigma Aldrich (St. Louis, MO).

5.2.1 Serum immunoaffinity depletion

Human serum samples of healthy people were obtained from the University Hospital, Ann Arbor, Michigan according to IRB approval. The samples were aliquoted and stored in a -80 °C freezer until use, with only one freeze/thaw cycle. Fourteen of the most abundant proteins were depleted from 250 µL of the serum sample using the Seppro® IgY-14 LC 10 column (Sigma Aldrich, St. Louis, MO) according to the manufacturer's protocol. The flow-through fraction was collected between 0-30 min of an LC run. The depleted serum was buffer-exchanged using Amicon® Ultra 15 mL (3 kDa MWCO) filter units (Millipore, Bradford, MA) by centrifugation at 7,500×g and washing three times with 4 mL of HPLC grade water before being concentrated to a final volume of approximately 250 µL. The final protein concentration was measured by Bradford protein assay (Bio-Rad, Hercules, CA) using bovine serum albumin as the protein standard, and aliquoted to 250 µg (for label-free) or 100 µg (for iTRAQ labeling) protein amount per

vial.

5. 2.2 Peptide level LCA enrichment

Two hundred and fifty microgram of serum protein was reduced with 10 mM dithiothreitol at 50 °C, alkylated with 22 mM iodoacetamide at room temperature in dark, digested with trypsin (Promega, Madison, WI) with protein/trypsin ratio of 50/1 at 37 °C overnight, dried down using SpeedVac (Thermo Scientific, San Jose, CA), and reconstituted in 500 µL lectin binding buffer (20mM Tris, 150 mM NaCl, 100 mM CaCl₂ and 100 mM MnCl₂, pH=7.6). Lectin affinity chromatography was performed in a 2 mL centrifugal column (Pierce, Rockford, IL). Prior to use, 400 µL of agarose-bond *Lens culinaris* agglutinin LCA (50% slurry, 3 mg/mL lectin/gel volume) purchased from Vector Labs (Burlingame, CA) was added to the centrifugal column, spun down, and prepared by rinsing with 500 µL binding buffer three times.

The peptide samples were incubated with LCA at room temperature with gentle agitation for 1 hour. Following five washes with binding buffer to remove nonspecific binding, the core-fucosylated proteins were eluted with 500 µL of elution buffer (200 mM α -methyl mannoside and 200 mM α -methyl glucoside in the binding buffer) for three times. Two strategies were compared for the following removal of the salts and monosaccharides in enriched glycopeptides. The first strategy used C₁₈ spin column (Pierce Scientific, Rockford, IL) where the glycopeptide samples were reconstituted in 5% acetonitrile (ACN) with 0.1% trifluoroacetic acid (TFA), loaded on the spin column, eluted with 70% ACN with 0.1% TFA, and dried down. The second strategy used Amicon® Ultra 4 mL (3 KDa MWCO) filter units to exchange the lectin elution buffer by centrifugation at 5,000×g and washing three times with 4 mL of 250 mM sodium

acetate. One microliter (5mU) of Endoglycosidase F3 (QAbio, Palm Desert, CA) was added and incubated for 20h at 37 °C. The partially deglycosylated peptides obtained from both methods were dried down, reconstituted in 10 μ L of water with 0.1% TFA, and subsequently desalted using C₁₈ ZipTip (Millipore, Bradford, MA).

5. 2.3 Protein level LCA enrichment

Two hundred and fifty microgram of serum protein was reconstituted in 500 μ L lectin binding buffer, and the lectin enrichment was performed as described in Section 2.2. The enriched core-fucosylated proteins were buffer-exchanged to 50 mM ammonium bicarbonate using Amicon® Ultra 4 mL (3 kDa MWCO) filter units by centrifugation at 5,000 \times g and washing three times with 4 mL of 50 mM ammonium bicarbonate. The protein digestion was performed using the aforementioned protocol. Molecular weight cut-off was used for removal of the eluting monosaccharides and salts. N-glycopeptides were concentrated with Amicon® Ultra 4 mL (3 kDa MWCO) filter units by centrifugation at 5,000 \times g and washing three times with 4 mL of 250 mM sodium acetate. Subsequent Endo F3 deglycosylation and desalting were described as in Section 2.2.

5. 2.4 Isobaric iTRAQ labeling

Four aliquots of 100 μ g depleted serum sample were digested with trypsin in 50 mM triethylammonium bicarbonate (TEAB) and dried down as previously described. The four peptide samples were labeled with iTRAQ isobaric 4-plex reagents (AB Sciex, Foster City, CA) according to the manufacturer's protocol. Briefly, 20 μ L of 50 mM TEAB was added to each of the peptide sample, and 50 μ L of ethanol was added to each of the iTRAQ reagent vial. The reconstituted iTRAQ reagents were transfer to respective peptide samples, mixed, and incubated at room temperature for 1 h. A volume of 5 μ L 5%

hydroxylamine was added to react with excessive iTRAQ reagents and quench the reaction for 15 min. The four differentially labeled samples were combined and core-fucosylated peptides were enriched using LCA lectin affinity chromatography as described in Section 2.2. The following workflow was similar to non-labeled strategy.

5. 2.5 LC-ESI-MS/MS analysis

Orbitrap Elite (Thermo Scientific) coupled with Proxeon EASY nLC II system (Thermo Scientific) was used for LC-ESI-MS/MS analysis. The samples were reconstituted in 0.1% formic acid and loaded to in a 25 cm self-packed LC column (75 μm i.d., Magic C₁₈ AQ, 5 μm particle size). A 71-min linear gradient from 5% solvent B to 35% solvent B was used at flow rate 400 nL/min where the solvent A is 2% ACN in water with 0.1% FA and the solvent B is 2% water in ACN with 0.1% FA. The gradient was then ramped to 95% B in 2 min and stayed isocratic for 5 min, and returned to 5% B in 1 min and equilibrated for 9 min.

For HCD analysis, a top 15 method was used. The mass spectrometer performed a FT-MS full scan (m/z range 400-1800, resolution 120,000), followed by HCD MS/MS activation of the 15 most abundant ions with ion signal intensity above 5000. The normalized collision energy (NCE) for HCD was set at 35%, the resolution was set at 15,000, the isolation window was ± 1.5 Da, and the activation time was 10 ms.

For neutral loss-triggered HCD analysis, FT-MS full scan was first acquired (m/z range 400-1800, resolution 120,000), followed by a CID MS/MS (ion intensity threshold 5000, resolution 5000, isolation width ± 2 Da, NCE 35%, activation time 10 ms) of the most abundant precursor ion. Upon observation of neutral loss corresponding to core fucose loss (m/z 36.51 for +4 charge, m/z 48.69 for +3 charge and m/z 73.02 for +2

charge), an additional HCD fragmentation was triggered on the original precursor (precursor of CID MS²) with the aforementioned conditions. The CID and HCD scans were repeated for the 15 most abundant precursor ions.

For neutral loss-triggered CID analysis, the full scan and the first CID were set in the same way as the neutral loss-triggered HCD top-15 method. The only difference is neutral loss triggered CID MS³ was performed on the neutral-loss fragment in CID MS² (precursor ion of MS²-fucose) (ion intensity threshold 5000, resolution 5000, isolation width ± 4 Da, NCE 35%, activation time 10 ms). CID MS³ fragments were measured in the ion trap.

For all the three mass spectrometric methods, the ESI spray voltage was 2.25 kV, the capillary temperature was 300 °C. Dynamic exclusion was enabled at 1 repeat count over 30 s within a 10 ppm exclusion window. The automatic gain control (AGC) was used to accumulate sufficient ions for analysis. For full MS scan, AGC target was set at 1×10^6 , the AGC target for MSⁿ in the Orbitrap was 5×10^4 , and 2×10^3 for MSⁿ in the ion trap. The maximum ion injection time was 250 ms for both full MS and MSⁿ in the Orbitrap, and 150 ms for MSⁿ in the ion trap. Only +1 charge state was rejected. All data acquisition was controlled by XCalibur 2.1.

5. 2.6 Data analysis

All MS/MS data was searched against a *Homo Sapiens* database from SwissProt (Release 2010_10, downloaded on Nov 2, 2010) for core-fucosylation site identification. Proteome Discoverer 1.1 (Thermo Scientific) with Sequest algorithm was used for database search. The following search parameters were used: (1) fixed modification: cysteine carbamidomethylation (+57.022 Da); (2) variable modification: methionine

oxidation (+15.996 Da), and addition of N-acetylglucosamine (+203.080 Da) or N-acetylglucosamine-fucose (+349.138 Da) to asparagines. For iTRAQ labeled sample search, an additional iTRAQ modification of +144.102 Da at N-terminus or Lys; (3) missed cleavages allowed: three; (4) precursor ion tolerance: 10 ppm; (5) fragmentation ion tolerance: 0.03 Da for HCD data, and 0.8 Da for CID data; and (6) precursor ion order is MS(n-1). All search results containing core-fucosylation sites were verified using the oxonium ions (m/z 204.08 Da and 126.06 Da) by manual examination of HCD spectra.

5.3 Results and Discussion

5.3.1 Analytical strategy

Both the solid phase extraction (SPE) method for enrichment of core-fucosylated peptides and MS/MS method were optimized in this study for comprehensive profiling of serum core-fucosylation. Figure 5.1 shows the overall SPE enrichment workflows compared in this study. All experiments were performed in triplicates to assess the performance of the protocols. Top-14 proteins were removed from human serum using an IgY-14 column to focus the effort on analysis of mid-low abundance glycoproteins. An aliquot of 250 μg (~50 μL of starting serum) or 100 μg (~20 μL of starting serum for each iTRAQ channel) was used for label-free or iTRAQ labeling strategy respectively. Compared to prior reports on serum core-fucosylation which used larger volumes of serum (200 μL ¹⁴ or 800 μL ²⁶) but obtained fewer core-fucosylation sites, part of the improvement of this assay lies in the high abundance protein depletion which reduces the complexity and dynamic range of the analytes, allowing deeper interrogation of

low-abundance proteins.

Both peptide (Strategy A and B) and protein level (Strategy C) lectin enrichment protocols have been applied and compared. Lectin enrichment at the protein level is believed to be more efficient because proteins usually carry multiple glycan epitopes at different sites that can be captured by the lectin, generating a stronger binding, but the binding is weaker for individual glycopeptides which normally have only one glycan. However, the specificity of peptide level enrichment is much higher than protein level enrichment. While the LCA enriched peptides yield relatively pure core-fucosylated glycopeptides, the LCA enriched proteins yield a mixture of core-fucosylated glycopeptides, non-core-fucosylated glycopeptides and nonglycopeptides after trypsin digestion, which requires further fractionation. Some studies used an additional lectin enrichment at the peptide level, but it was not successful (data not shown) in this study due to the low starting amount of proteins and further sample loss in the extra cleaning steps. As shown in Figure 5.2, peptide level enrichment (strategy A and B) is superior to protein level enrichment (strategy C) with significantly increased number of core-fucosylated peptides (Figure 5.2(a)) and core-fucosylated proteins (Figure 5.2(b)) identified, mainly due to the higher concentrating effect of lectin enrichment.

Following peptide level lectin enrichment, two methods were compared to remove the high concentration of monosaccharides used in the elution step of lectin affinity chromatography. Strategy A used a cellulose membrane at molecular weight cut-off of 3000 Da based on the fact that the glycopeptides normally have higher molecular weights than nonglycopeptides^{26, 27}. Strategy B used C₁₈ spin column based on the approximate non-retention of monosaccharides on the reverse phase²⁸. It is shown in Figure 5.2 that

the C₁₈ SPE (strategy B) provided worse performance than molecular weight cut-off (strate A) probably owing to the sample loss during multiple steps of drying, and the low retention of some glycopeptides with hydrophilic peptide backbone. Comparatively, the molecular weight cut-off is more efficient by incorporating enrichment, buffer exchange and concentration in one step. In summary, strategy A which integrates peptide level LCA enrichment, molecular weight cut-off for desalting and concentration was chosen as the workflow prior to LC-MS/MS analysis.

5.3.2 MS/MS methods

A sample prepared with strategy A was divided to nine aliquots and analyzed in triplicates with three MS/MS methods namely HCD, neutral loss-triggered HCD and neutral loss-triggered CID-MS³. In the previous reports of core-fucosylation mapping, people used neutral loss-triggered CID-MS³^{14, 26, 29} based on the fact that ion trap type CID-MS² of core-fucosylated peptides almost only generates neutral loss of core fucose (Figure 5.3 (a)) and an additional CID-MS³ on the neutral loss product provides further fragmentation on the peptide backbone (Figure 5.3 (b)). Recently, a novel MS/MS method HCD is used in post-translational modification studies. HCD is a more energetic beam-type CID performed in an octapole collision cell in a LTQ-Orbitrap instrument, which keeps the high sensitivity merit of the ion trap, but overcomes the limitation of low mass cut-off, enabling full-mass-range fragmentation acquisition. Furthermore, the fragment ions of HCD are measured in the Orbitrap with high resolution and high mass accuracy³⁰. In glycopeptide analysis, these properties are very useful in the unambiguous assignment of glycan oxonium ions (eg. [HexNAc+H]⁺ m/z 204.087, or [Hex-HexNAc+H]⁺ m/z 366.138)³¹ for location of glycopeptides or distinct Y1 ions

[peptide+HexNAc+H]⁺ for peptide sequence identification^{22, 31}. The oxonium ions are considered as diagnostic ions for glycopeptides and can be employed to further trigger ETD for peptide backbone fragmentation. This HCD product ion triggered ETD has been used in multiple studies for streamline analysis with significant reduction in duty cycle^{23, 24}.

In this study, we evaluated the performance of HCD in analysis of the partially deglycosylated core-fucosylated glycopeptides. As shown in Figure 5.3 (c), HCD MS/MS generates b and y ions similar to the neutral loss-triggered CID MS³ (Figure 5.3 (b)) but with significantly higher mass accuracy (data not shown), and enables detection of low-mass range sequence ions (y_1^+) and glycan oxonium ions (m/z 204.09 and 126.05). Furthermore, the neutral loss-triggered HCD was tested where a CID was used followed by HCD performed on the original precursor ion rather than the neutral loss product. HCD-only method provided the best performance among the three MS/MS method in the number of unique core-fucosylated peptides and core-fucosylated proteins as shown in Figure 5.4. HCD method is superior to CID method mainly due to the high mass accuracy and wide mass range, and also likely due to the more efficient fragmentation. Neutral loss-triggered HCD is theoretically more productive and specific than HCD because only ions which generate neutral loss equivalent to m/z 's of fucose loss would be further fragmented by HCD, and thus reduces the interference of irrelevant signals. However, the number of ID's was slightly lower in neutral loss-triggered HCD compared to HCD probably because some core-fucosylated peptides generate low-intensity neutral-loss peaks which were not selected for HCD analysis. It is obvious in Figure 5.4 that HCD is the most effective MS/MS method for core-fucosylated peptide analysis and was used

throughout this study.

5.3.3 Assignment of core-fucosylation sites

By using Strategy A as shown in Figure 5.1, a total of 140 unique core-fucosylated peptides corresponding to 138 unique core-fucosylated sites and 92 proteins were identified with high mass accuracy. A histogram of mass errors is shown in Figure 5.5, where over 90% of the peptides were identified within ± 1 ppm mass error. The utilization of mass analyzer with high resolution and high mass accuracy significantly enhances the confidence of identification. Among the 92 proteins, 67% carry only one core-fucosylation site, 21.3% carry two core-fucosylation sites, and 11.7% carry three or more core fucosylation sites as illustrated in Figure 5.6 (a). Among the 140 core-fucosylated peptides identified, approximately 98% contain the motif of N-X-S/T as shown in Figure 5.6 (b), which is consistent with other glycosylation site profiling studies^{32, 33}, and suggests the high confidence of site identification. Among the three non-sequon peptides, two of them contain N-X-C motif, which is atypical but scatterly reported in various studies³³⁻³⁵. Due to rigidity consideration, only peptides containing N-X-S/T sequon were reported in this study. Table 5.1 shows 137 peptides corresponding to 135 unique core-fucosylation sites and 90 proteins. All identifications were manually validated by observation of both oxonium ions (m/z 204.08 Da and 126.06 Da). Notably, 25 novel glycosylation sites were reported in this study. An example of HCD-MS/MS spectrum of the newly identified core-fucosylation sites in Sushi, nidogen and EGF-like domain-containing protein 1 is shown in Figure 5.7. This study is by far the most comprehensive serum core-fucosylation profiling, which complements the understanding of serum glycoproteomics. Furthermore, this assay could be performed in a quantitative

manner using iTRAQ labeling for further interrogation of core-fucosylation as potential cancer biomarker as detailed in Section 5.3.5.

5.3.4 Bioinformatic concerns due to neutral loss

There are three ion series in the HCD-MS/MS analysis of core-fucosylated peptides, the oxonium ions, the b, y ions, and the b, y ions with neutral loss of GlcNAc-Fuc as illustrated in Figure 5.7. It should be noted that the peptide-GlcNAc bond cleavage occurs readily in HCD analysis so that the b, y ions with neutral loss are the most intense peaks in most cases. However, current implementation of Sequest algorithm in Proteome Discoverer has troubles in assigning these fragments with neutral loss to the original core-fucosylated peptides, which lowers the number of matched ions and therefore lowers the score of identification. This problem is more severe for peptides with glycosylation sites toward the C terminus. It is known that y ions are usually more intense than b ions due to the guaranteed basicity of the C-terminal Lys and Arg. For the peptides with core-fucosylation sites toward the C terminus, approximately half or more y ions would suffer from the neutral loss problems and will not be correctly identified. As shown in Figure 5.7, only half of the y ions present as labeled blue were identified by Proteome Discoverer in the newly identified glycopeptide, while the other half y ions containing the core fucosylation as labeled red were not assigned due to the neutral loss. Comparatively, the missed assignment is not that severe for peptides with glycosylation site toward the C terminus as shown in Figure 5.3 (c) with almost all y ions correctly assigned.

It is speculated that the number of core-fucosylation sites will increase if the neutral loss assignment problem can be solved, and more bioinformatic work is underway.

5.3.5 Quantitative study of the assay by iTRAQ labeling

The promise of using this assay for quantitative analysis was evaluated. While it is obvious that the label-free quantification based on the extracted peak area of the peptide precursor ions can be used to quantify the change of site-specific core-fucosylation, the HCD-MS/MS employed in this work enables isobaric chemical labeling methods such as iTRAQ labeling. Four-plex iTRAQ labeling reagents were used in this study which allowed simultaneous comparison of three samples (one iTRAQ channel is the internal standard which remains constant) in a single LC-MS/MS run. iTRAQ reagents contain three functional groups: the reporter ion group which is labeled with isotopic variants, the amine reaction group which links to the free amine at the N-terminus or lysine, and the mass normalization group which balances the mass differences of the reporter ion group to ensure the same mass for the different isotopic variants of the iTRAQ tag.

For implementation of the iTRAQ labeling in discovery of site-specific core-fucosylation as potential cancer biomarkers, four samples with different disease states (*eg*, one normal control sample, one pancreatic cancer sample, one chronic pancreatitis sample and one internal standard sample) are digested, labeled with four isotopic iTRAQ variants respectively, and combined for following LCA enrichment *etc*. In the precursor scan, due to the isobaric nature of iTRAQ reagent, the same peptides labeled with different iTRAQ variants have the same mass hence the complexity of full scan remain the same. The HCD method enables the detection of reporter ions in the low mass range for quantification purpose. The HCD-only configuration allows both core-fucosylation identification and quantification with lower duty-cycle compared to CID-HCD dual run.

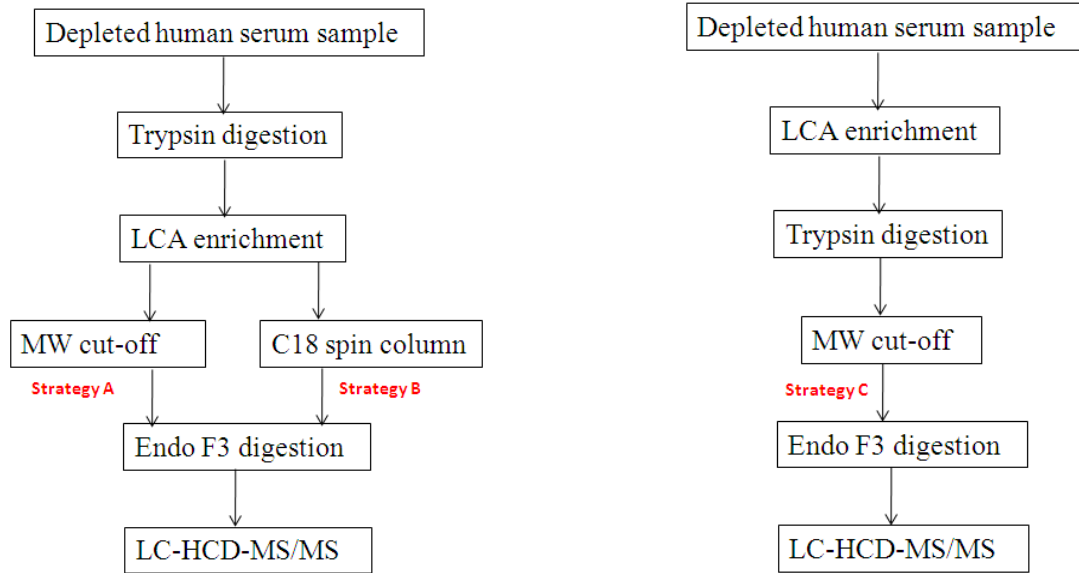
For proof of principle, four identical serum samples with 100 µg protein amounts were labeled with 4-plex iTRAQ reagents respectively and the quantification results are shown in Figure 5.8. The ratios of three reporter ions relative to one reporter ions average at 1 with little spread as expected. Over 95% of identified peptides were correctly labeled (data not shown). However, the number of unique core-fucosylated peptide identified is reasonable but lower than the label-free strategy (81 versus 140) for unclear reasons. It is expected that the iTRAQ labeling strategy combining with the described workflow will be valuable in future study of quantifying core-fucosylation change in different cancers.

5.4 Conclusion

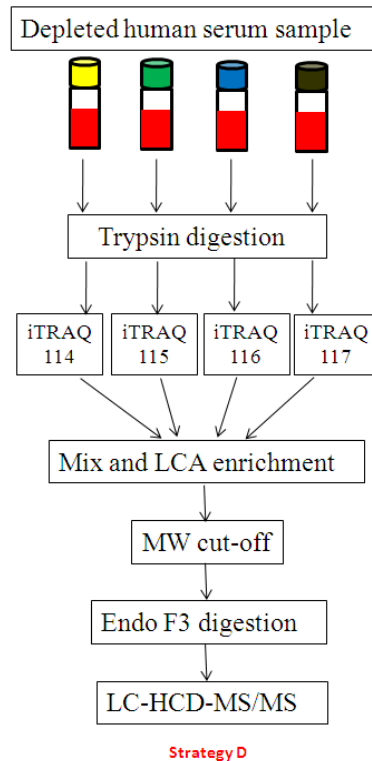
In this report, an LC-MS/MS assay was developed and optimized to identify and quantify large-scale core-fucosylation in serum proteins with high confidence. This workflow integrates high-abundance protein removal, peptide level lectin enrichment, partial deglycosylation and HCD-MS/MS analysis. 135 core-fucosylation sites corresponding to 90 proteins were reported in this study, where 25 sites were novel identifications of glycosylation. All the reported sites were verified by the presence of consensus N-X-S/T sequon and oxonium ions. This study is the most comprehensive mapping of serum core-fucose up to now, and represents a significant improvement over the prior studies that utilized conventional CID which is less efficient in identification and less compatible for quantification. The quantitative aspect of this assay was successfully evaluated by iTRAQ labeling, which proves promising in the future discovery of core-fucosylation as cancer biomarker.

Figures

Figure 5. 1. Enrichment workflow comparisons for protocol optimization. (a) label-free strategies, where Strategy A and B are peptide level LCA enrichment, and Strategy C is protein level LCA enrichment. Two hundred and fifty micrograms of serum proteins were used.. (b) iTRAQ labeling strategy, where four aliquots of 100 μ g serum proteins tryptic digests were differentially labeled and combined for LCA enrichment

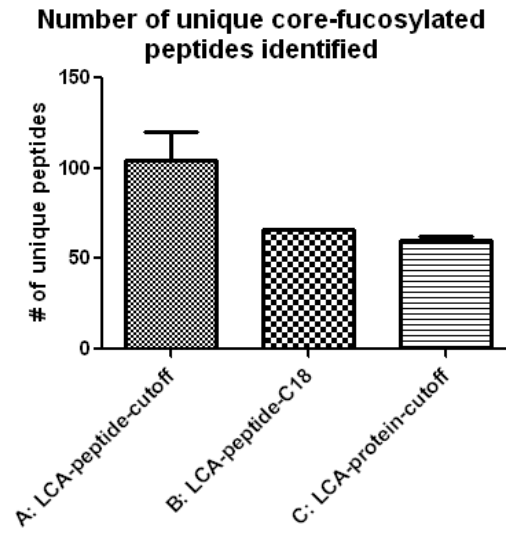


(a)

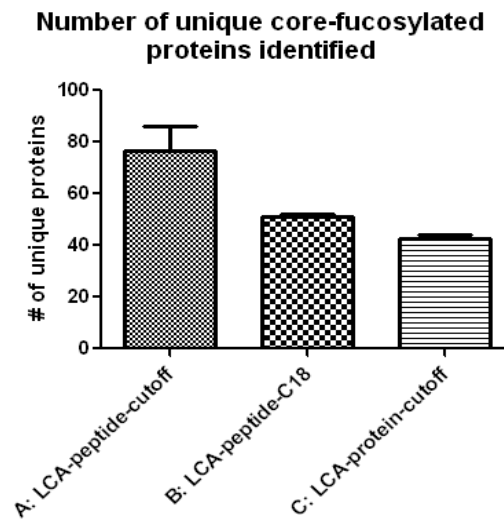


(b)

Figure 5. 2. Comparison of numbers of identified unique core-fucosylated peptides (a) and core-fucosylated proteins (b) using different workflows as shown in Figure 5.1(a).

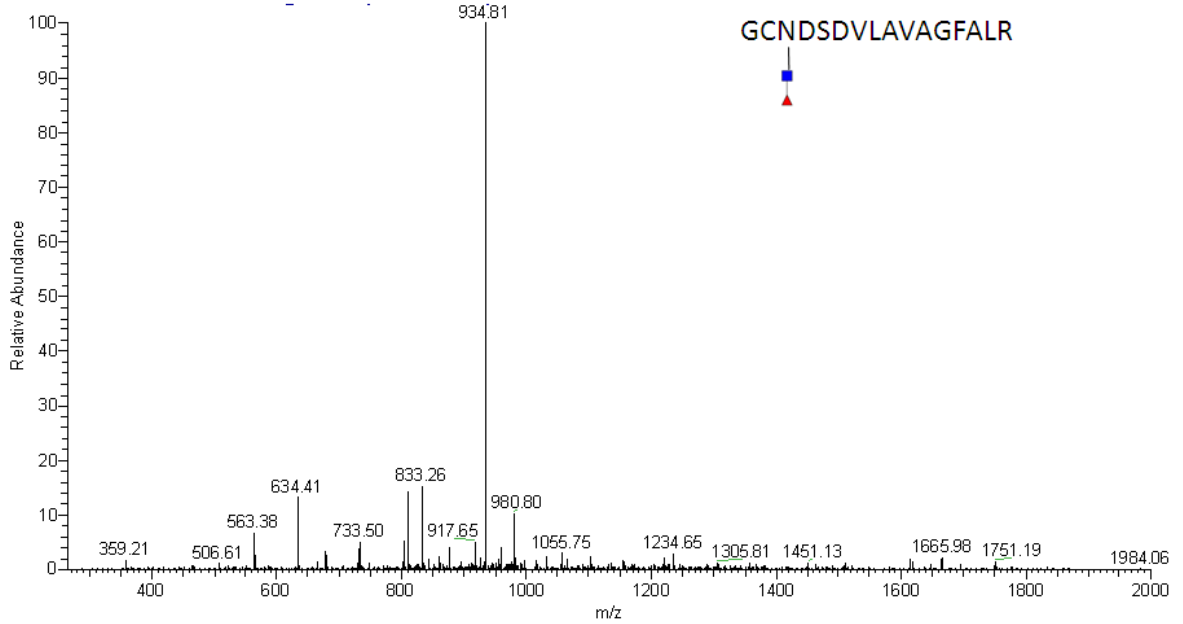


(a)

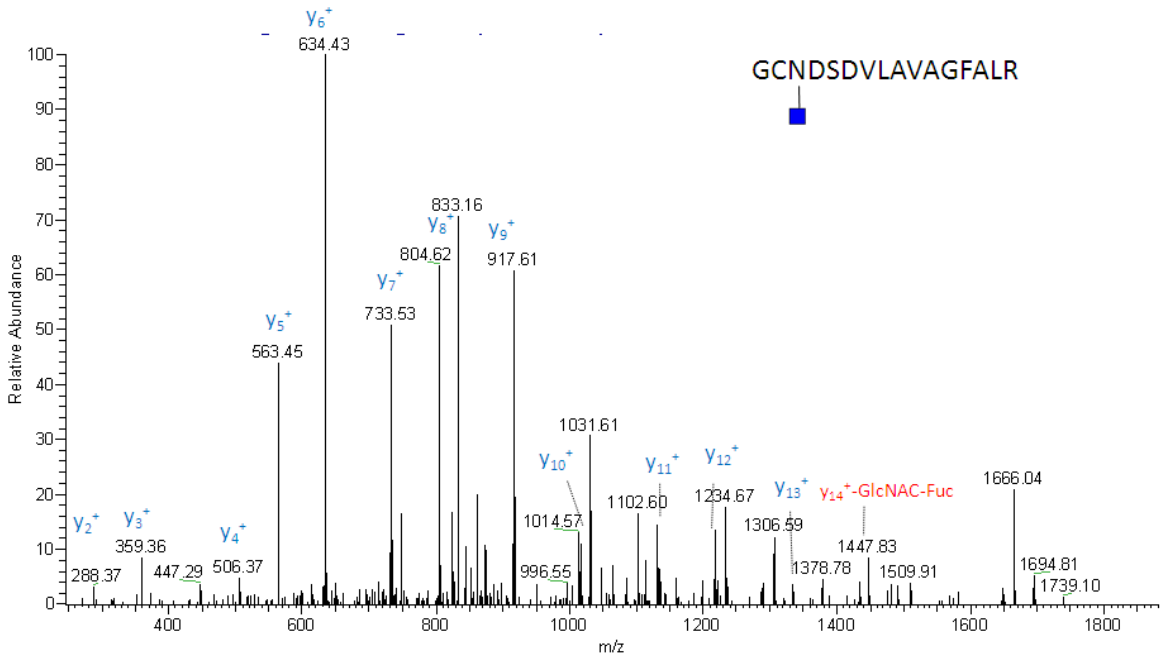


(b)

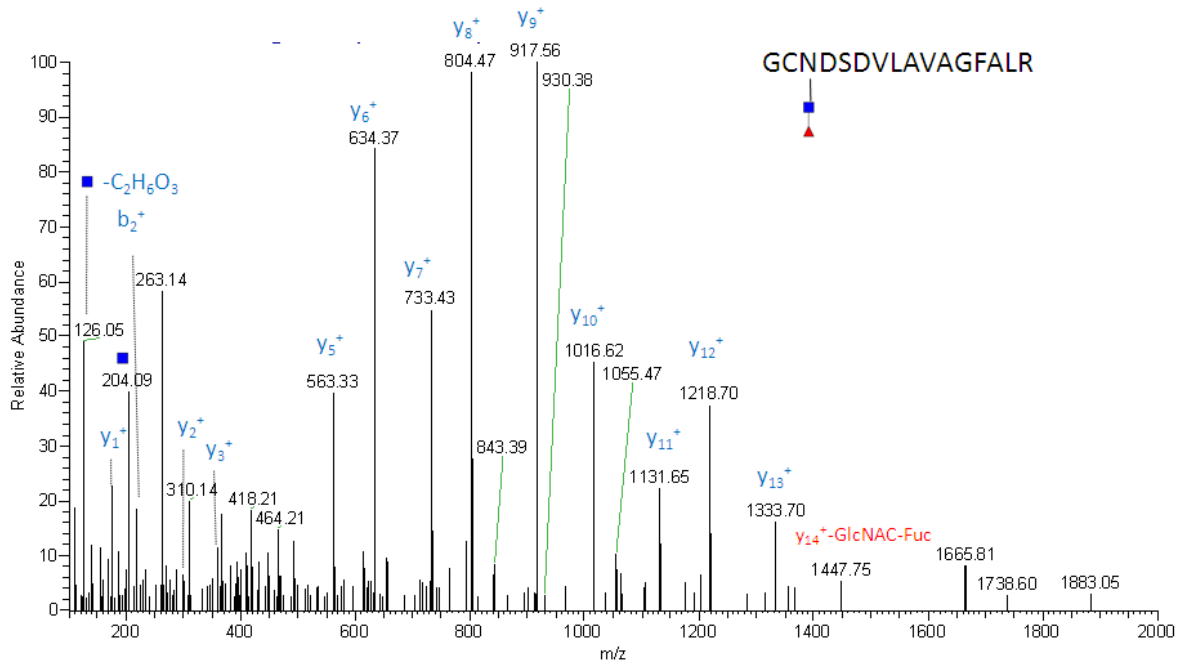
Figure 5. 3. Comparison of neutral loss-triggered CID-MS3 and HCD MS/MS of the same core-fucosylated peptide. (a) CID MS/MS of the core-fucosylated peptide which is dominated by the neutral loss of core fucose (m/z 934.81). (b) Neutral loss triggered-CID MS3 of fragment m/z 934.81. (c) HCD MS/MS of the same core-fucosylated peptide



(a)

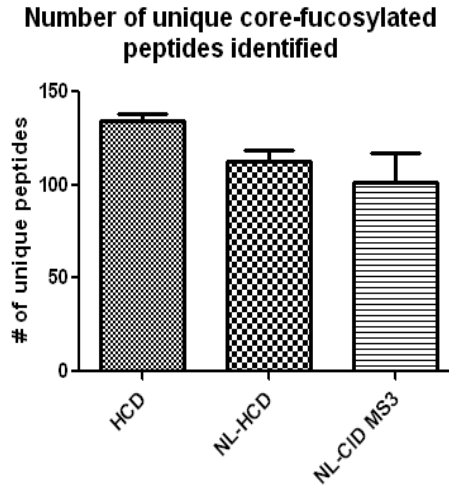


(b)

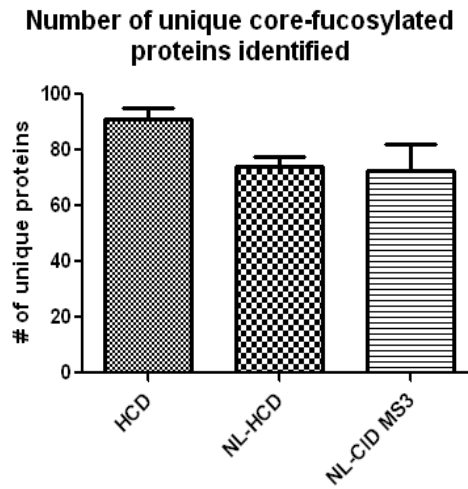


(c)

Figure 5. 4. Comparison of numbers of identified unique core-fucosylated peptides (a) and core-fucosylated proteins (b) using HCD, neutral loss-triggered HCD and neutral loss-triggered CID.



(a)



(b)

Figure 5. 5. Histogram of mass accuracy distribution of identified core-fucosylated peptides by HCD-MS/MS.

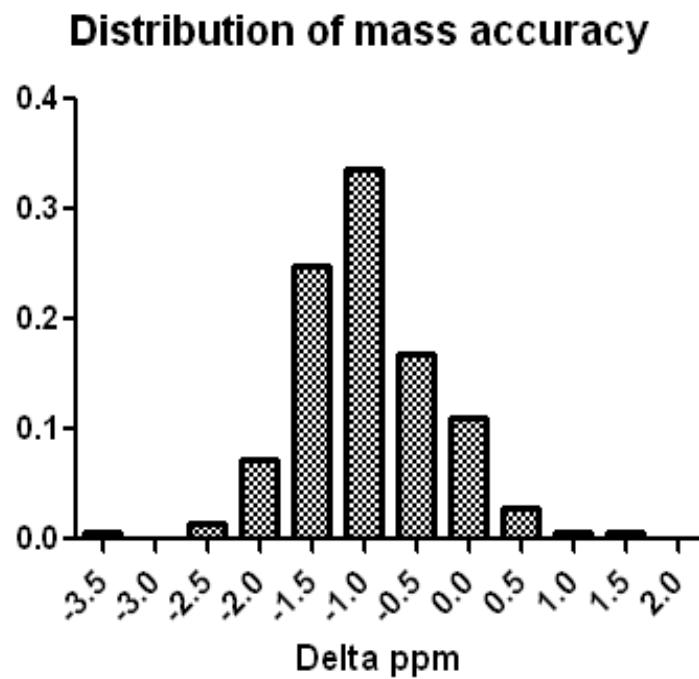
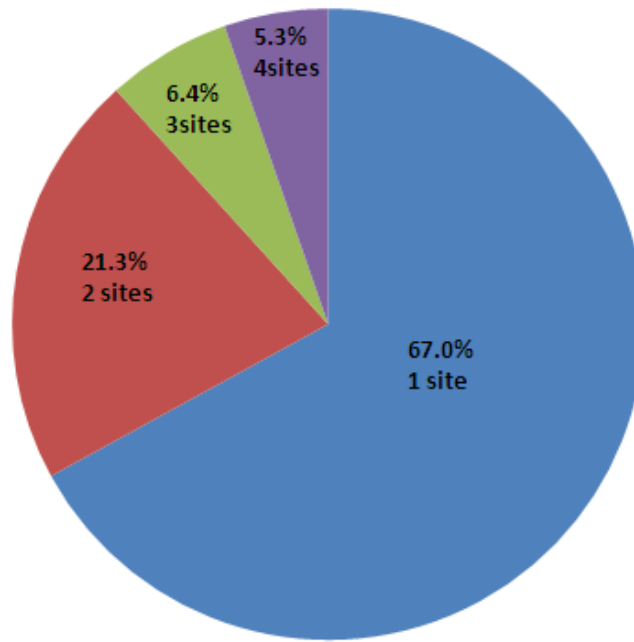
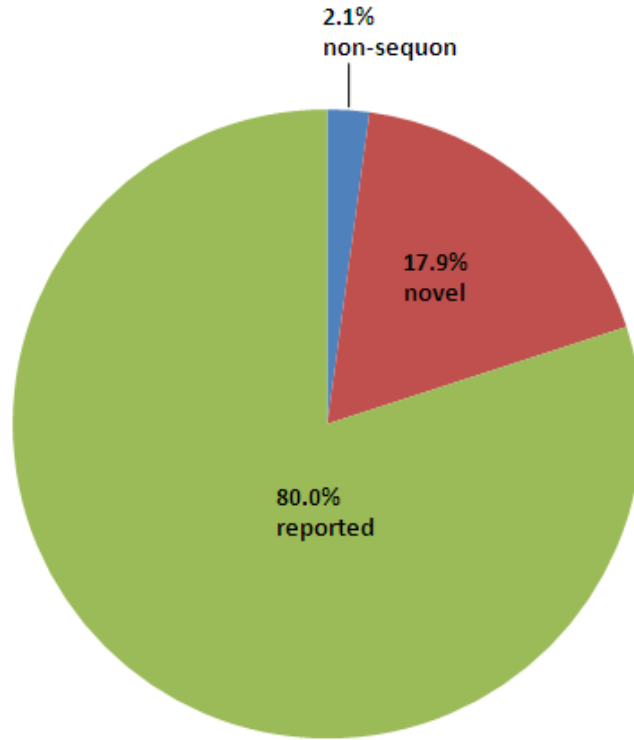


Figure 5. 6. (a) Distribution of singly and multiply core-fucosylated proteins identified and (b) proportion of novel identifications, reported identifications and non-motif identifications.



(a)



(b)

Figure 5. 7. HCD-MS/MS spectrum of a newly identified glycosylation site of Sushi, nidogen and EGF-like domain-containing protein 1. The blue annotations show identified fragments, and the red annotations show unidentified but manually assigned fragments.

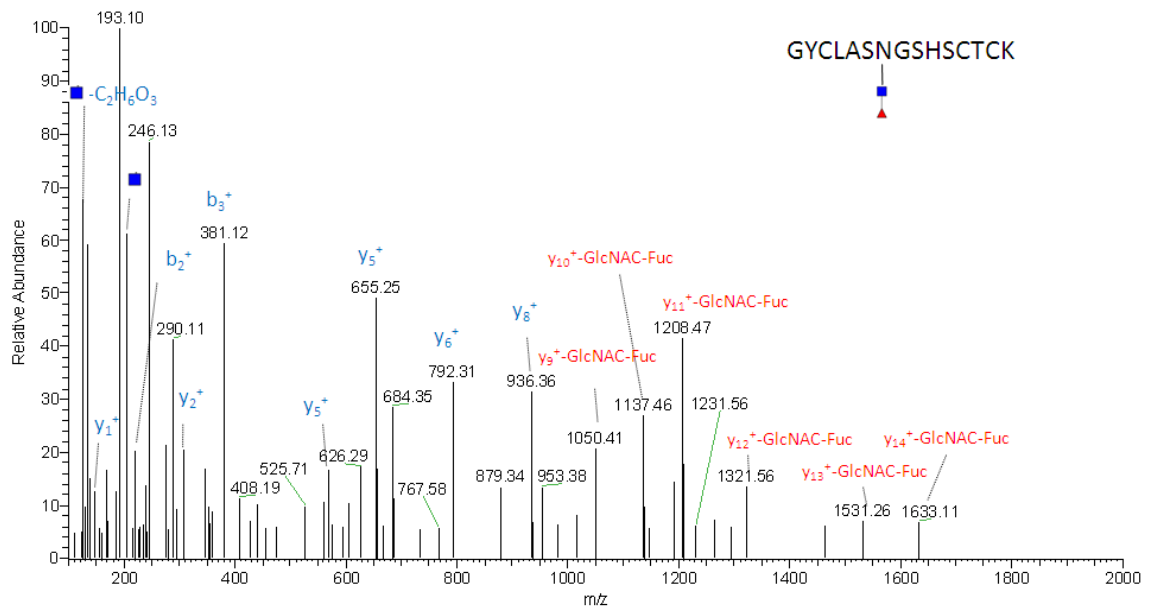


Figure 5. 8. Summary of ratios of reporter ion abundances in an isobaric iTRAQ experiment where four aliquots of 100 μ g of serum protein digests were labeled with 4-plex reagents (reporter ion masses 114, 115, 116 and 117) and mixed for subsequent analysis as shown in Figure 5.1 (b). Ratios were constructed against the abundances of reporter ions 114

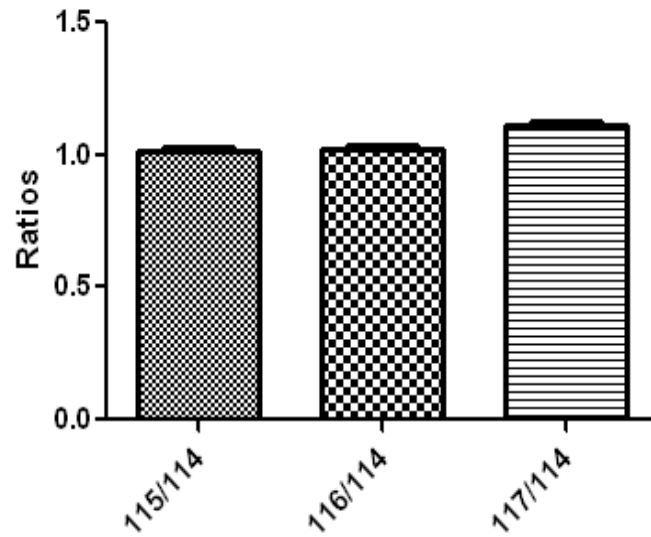


Table 5. 1. List of core-fucosylated peptide containing the N-X-S/T motif in this study.

Protein and Gene Accessions	Protein name	Sequence	Charge	m/z [Da]	ΔM [ppm]
P0C0L5 (CO4B_HUMAN)	Complement C4-B	NTTCQDLQIEVTVK	2	999.48138	-1.46
P23142 (FBLN1_HUMAN)	Fibulin-1	CATPHGDNASLEATFVK	3	723.00024	0.07
P17936 (IBP3_HUMAN)	Insulin-like growth factor binding protein 3	GLCVNASAVSR	2	741.85822	-0.91
		YKVDYESQSTDTQNFSSSEK	3	898.05811	-1.29
P05155 (IC1_HUMAN)	Plasma protease C1 inhibitor	GVTSVSQIFHSPDLAIRDTFVNASR	3	1022.84967	-1.14
		VLSNNSDANLELINTWVAKNTNNK	3	1007.83868	0.09
Q14515 (SPRL1_HUMAN)	SPARC-like protein 1	NYSHHQLNR	3	506.56992	-1.8
Q9NZK5 (CECR1_HUMAN)	Adenosine deaminase CECR1	VQNVTEFDDSLR	2	942.95667	-0.69
Q6UX71 (PXDC2_HUMAN)	Plexin domain-containing protein 2	VNLSFDFPFYGHFLR	3	736.69214	-1.36
P55290 (CAD13_HUMAN)	Cadherin-13	DPAGWLNINPINGTVDTTAVLDR	2	1401.19592	0.17
		NLSVVILGASDK	2	782.91858	-1.14
P43251 (BTD_HUMAN)	Biotinidase	FNDTEVLQR	2	735.85083	-1.03
P05090 (APOD_HUMAN)	Apolipoprotein D	ADGTVNQIEGEATPVNLTEPAKLEVK	3	1024.85278	-1.11
P49908 (SEPP1_HUMAN)	Selenoprotein P	CGNCSLTTLKDEDFCK	3	766.32483	-0.84
		EGYSNISYIVVNHQGISSR	3	824.73358	-1.11
P06681 (CO2_HUMAN)	C2 protein	LGSYPVGGNVSEFECEDGFILR	2	1333.12268	1.57
P24821 (TENA_HUMAN)	Tenascin	GNPCEPECPGNCHLR	3	744.96967	-0.48
		LNWTAADQAYEHFIQVQEANK	3	980.14044	-1.3
		NTTSYVLR	2	651.82336	-2.24
		VEAAQNLTLPGLSR	2	909.47699	-1.36
P00734 (THRB_HUMAN)	Thrombin light chain	YYNQSEAGSHIQR	3	672.31635	-0.39
		YPHKPEINSTTHPGADLQENFCR	4	765.85486	-0.93
Q8IZF2 (GP116_HUMAN)	Probable G-protein-coupled receptor 116	LNLVPGENITCQDPVIGVGEPEGK	2	1378.19019	0.85
Q13740 (CD166_HUMAN)	CD166 antigen	LNLSENYTSLISNAR	2	1022.50787	0.17
Q99784 (NOE1_HUMAN)	Noelin	VQNMSQSIEVLDR	2	934.44928	-2.22
Q86VB7 (C163A_HUMAN)	Scavenger receptor cysteine-rich type 1 protein M130	APGWANSSAGSGR	2	783.85388	-1.68
		CKGNESSLWDPCPAR	3	676.95929	-0.24
		GNESALWDCK	2	764.82672	-0.7
		GNESLWDPCPAR	2	870.87201	-0.83
Q9UGM5 (FETUB_HUMAN)	Fetuin-B	GCNDSVDVLAVAGFALR	2	1007.47449	-0.87
Q12884 (SEPR_HUMAN)	Seprase	SVNASNYGLSPDR	2	864.89905	-0.86
P05156 (CFAI_HUMAN)	Complement factor I light chain	FKLSDLSINSTECLHVHCR	4	667.06976	-0.93
		FLNNGTCTAEGK	2	830.87115	-1.25
		LSDLINSTECLHVHCR	4	598.27911	-0.71
		NGTAVCATNR	2	706.81866	-1.55
P49747 (COMP_HUMAN)	Cartilage oligomeric matrix protein	CGPCPAGFTGNGSHCTDVNECNAHP CFPR	4	907.11292	0.08
P13473 (LAMP2_HUMAN)	Lysosome-associated membrane glycoprotein 2	VASVININPNTHSTGSCR	3	793.04846	0.32
P00736 (C1R_HUMAN)	Complement C1r subcomponent	EHEAQSNASLDVFLGHTNVEELMK	3	1016.47559	-0.96
P11597 (CETP_HUMAN)	Cholesteryl ester transfer protein	SIDVSIQNVSVVFK	2	942.49469	-1.48
Q9H4A9 (DPEP2_HUMAN)	Dipeptidase 2	NFSYGGTSLDR	2	818.86975	-0.91
P36955 (PEDF_HUMAN)	Pigment epithelium-derived factor	VTQNLTLIEESLTSEFIHDIDR	3	974.82007	-0.37
P04278 (SHBG_HUMAN)	Sex hormone-binding globulin	SHEIWTWSCPQSPGNGTDASH	3	885.37274	-0.89
P13598 (ICAM2_HUMAN)	Intercellular adhesion molecule 2	QESMNSNVSVYQPPR	2	1042.97485	-0.83
O00391 (QSOX1_HUMAN)	Sulfhydryl oxidase 1	NGSGAVFPVAGADVQTLR	2	1054.52795	-0.98
O14498 (ISLR_HUMAN)	Immunoglobulin superfamily containing leucine-rich repeat protein	FQAFANGSLLIPDFGK	2	1037.52209	-0.53
O15204 (ADEC1_HUMAN)	ADAM DEC1	EHAVFTSNQEEQDPANHTCGVK	4	712.56519	0.51

P00450 (CERU_HUMAN)	Ceruloplasmin	EHEGAIYPDNTTDFQR	3	747.99756	-0.08
		ELHHLQEQNVSNAFLDK	3	791.05109	-1.09
		ELHHLQEQNVSNAFLDKGEFYIGSK	4	813.89655	-1.62
P00748 (FA12_HUMAN)	Coagulation factor XII	NHSCEPCQTLAVR	3	640.95142	-1.54
		RNHSCEPCQTLAVR	3	692.9856	-0.74
P00751 (CFAB_HUMAN)	Complement factor B	SPYYNVSDIEISFHCYDGYTLR	3	979.09088	-1.16
P01008 (ANT3_HUMAN)	Antithrombin-III	AAINKWVSNKTEGR	4	481.50049	-0.91
P01009 (A1AT_HUMAN)	Alpha-1-antitrypsin	YLGNAIAIFFLPDEGK	2	1053.01917	-0.72
P01011 (AACT_HUMAN)	Alpha-1-antichymotrypsin	FNLTESEAEIHQSFQHLR	4	688.08728	-1.57
		YTGNASALFILPDQDK	2	1051.51172	-0.7
P01023 (A2MG_HUMAN)	Alpha-2-macroglobulin	GCVLLSYLNETVTVSASLESVR	3	916.12732	-1.08
		VSNQTLSLFFTVLQDVPVR	2	1256.66028	-1.96
P01042 (KNG1_HUMAN)	Kininogen-1	LNAENNATFYFK	2	890.91663	-0.93
		YNSQNSNNQFVLYR	2	1112.51025	-0.94
P01833 (PIGR_HUMAN)	Polymeric immunoglobulin receptor	VPGNVAVLGETLK	2	873.9715	-0.86
P01857 (IGHG1_HUMAN)	Ig gamma-1 chain C region	EEQYNSTYR	2	769.82758	-0.93
P01859 (IGHG2_HUMAN)	Ig gamma-2 chain C region	GLTFQQNASSMCPDQDTAIR	2	1352.60571	-1.34
P01871 (IGHM_HUMAN)	Ig mu chain C region	YKNNSDISSTR	3	545.25562	-1.77
P01877 (IGHA2_HUMAN)	Ig alpha-2 chain C region	TPLTANITK	2	654.34979	-1.36
P02671 (FIBA_HUMAN)	Fibrinogen alpha chain	GLIDEVNDQFTNRINK	3	810.05865	-2.63
P02745 (C1QA_HUMAN)	Complement C1q subcomponent subunit A	RNPPMGNGVVFVFDTVITNQEEPPQN	3	1206.90515	-3.36
		HSGR			
P02749 (APOH_HUMAN)	Beta-2-glycoprotein 1	LGNWSAMPSCCK	2	800.354	-1.41
		VYKPSAGNNSLYR	2	909.44855	-0.99
P02765 (FETUA_HUMAN)	Alpha-2-HS-glycoprotein	AALAAFNAQNNGSNFQLEEISR	2	1357.64709	-1.32
		KVCQDCPLLAPLNDTR	3	750.36334	-1.52
P02787 (TRFE_HUMAN)	Serotransferrin	QQQHLLFGSNVTDCSGNFCLFR	3	955.42505	-0.32
P02790 (HEMO_HUMAN)	Hemopexin	ALPQPQNVTSLLGCTH	2	1043.01135	-0.72
		NGTGHGNSTHHGPEYMR	4	601.75702	-1.84
		SWPAVGNCSALR	2	877.40692	0.21
P04004 (VTNC_HUMAN)	Vitronectin	NGSLFAFR	2	630.80817	-1.31
		NISDGFDPDNDVDAALALPAHSYSYGR	3	1041.15723	-0.85
		NNATVHEQVGGPSLTSDLQAQSK	3	910.77271	-1.09
P04196 (HRG_HUMAN)	Histidine-rich glycoprotein	HSHNNSSDLHPHK	4	493.97345	-1.23
		VENTTVYYLVLDVQESDCSVLSR	3	1013.48505	0.22
P04220 (MUCB_HUMAN)	Ig mu heavy chain disease protein	GLTFQQNASSMCPDQDTAIR	2	1331.58264	-1.05
P04275 (VWF_HUMAN)	von Willebrand factor	GLQPTLTNPGECPNFTCACR	3	933.42133	-1.75
		HCDGNVSSCGDHPSEGCFPPDK	3	990.04144	-0.94
		MEACMLNGTVIGPGK	2	971.94257	-1.48
		TTCNPCPLGYKEENNTGECCGR	3	989.40027	-1.53
P05362 (ICAM1_HUMAN)	Intercellular adhesion molecule 1	ANLTVVLLR	2	674.38959	-0.83
		LNPTVTYGNDSFSAK	2	981.96198	-0.63
P05543 (THBG_HUMAN)	Thyroxine-binding globulin	VTACHSSQPNTLYK	3	675.98578	-0.31
P07602 (SAP_HUMAN)	Proactivator polypeptide	TNSTFVQALVEHVK	3	641.32916	-1.27
P08603 (CFAH_HUMAN)	Complement factor H	ISEENETTCYMGK	2	955.89624	-1.58
		MDGASNVTCINSR	2	887.38361	-1.74
		SPDVINGSPISQK	2	845.92181	-1.1
P10909 (CLUS_HUMAN)	Clusterin	LANLTQGEDQYYLR	2	1016.98761	-1.5
		MLNTSSLLEQLNEQFNWVSR	2	1379.66248	-2.33
P12259 (FA5_HUMAN)	Coagulation factor V	EDNAVQPNSSYTYVWHATER	3	906.07233	0.54
P12821 (ACE_HUMAN)	Angiotensin-converting enzyme	VTNDTESDINYLK	2	987.47559	0.05
P20851 (C4BPB_HUMAN)	C4b-binding protein beta chain	LGHCPDPVLVNGEFSSSGPVNVSDK	3	987.4646	-1.3
P22105 (TENX_HUMAN)	Tenascin-X	GNPLTSPASITFTGLEAPR	2	1190.09937	-0.51
P22897 (MRC1_HUMAN)	Macrophage mannose receptor 1	TAHCNESFYFLCK	3	675.95636	-1.17
P25311 (ZA2G_HUMAN)	Zinc-alpha-2-glycoprotein	AREDFIMETLKDIVEYYNDSNGSHV	4	937.9458	-0.92
		LQGR			

		DIVEYYNDSNGSHVLQGR	2	1208.05029	-0.58
P29622 (KAIN_HUMAN)	Kallistatin	FLNDTMAVYEAK	2	875.9071	-1.31
P33151 (CADH5_HUMAN)	Cadherin-5	EVYPWYNLTVEAK	2	980.97449	-0.49
		LDRENISEYHLTAVIVDK	4	616.81628	-0.86
P36980 (FHR2_HUMAN)	Complement factor H-related protein 2	LQNNENNISCVSR	2	969.93781	-1.37
P43121 (MUC18_HUMAN)	Cell surface glycoprotein MUC18	CGLSQSQGNLSHVDWFSVHK	4	659.55792	0.18
P43652 (AFAM_HUMAN)	Afamin	YAEDKFNETTEK	2	912.40637	-0.82
P51884 (LUM_HUMAN)	Lumican	AFENVTDLQWLILDHNLLENSK	3	987.82745	-1.71
		KLHINHNNLTESVGPLPK	2	1180.625	-1.26
		LGSFEGLVNLTFIHLQHNR	4	636.83252	-0.44
		LHINHNNLTESVGPLPK	3	744.72113	-0.84
P55058 (PLTP_HUMAN)	Phospholipid transfer protein	EGHFYYNISEVK	3	612.28394	-0.51
		VSNVSCQASVSR	2	821.88141	-2.05
Q06033 (ITIH3_HUMAN)	Inter-alpha-trypsin inhibitor heavy chain H3	NAHGEEKENLTAR	3	606.62115	-1.28
Q07954 (LRP1_HUMAN)	Prolow-density lipoprotein receptor-related protein 1	LTSCATNASICGDEAR	2	1037.94812	-0.7
		TCVSNCTASQFVCK	2	1005.92548	-0.84
		WTGHNVTVVQR	3	549.27698	-1.03
Q08380 (LG3BP_HUMAN)	Galectin-3-binding protein	ALGFENATQALGR	2	848.92206	-1.21
Q12860 (CNTN1_HUMAN)	Contactin-1	ANSTGTLVITDPTR	2	897.95062	-1.57
Q13201 (MMRN1_HUMAN)	Multimerin-1	FNPGAESVVLNSSTLK	2	1006.50555	-1.62
Q16610 (ECM1_HUMAN)	Extracellular matrix protein 1	QGNNHCTWK	3	532.23114	-0.86
Q92859 (NEO1_HUMAN)	Neogenin	TPASDPHGDNLTYSVFYTK	3	821.37988	0.2
P01033 (TIMP1_HUMAN)	Metalloproteinase inhibitor 1	FVGTPEVNQTTLYQR	2	1051.51794	-0.12
P09172 (DOPO_HUMAN)	Dopamine beta-hydroxylase	SLEAINGSGLQMGLQR	2	1020.00049	-1.23
O14786 (NRP1_HUMAN)	Neuropilin-1	GPECSQNYTTPSGVIK	2	1043.97766	-0.25
		RGPECSQNYTTPSGVIK	3	748.35339	-1.81
Q6YHK3 (CD109_HUMAN)	CD109 antigen	INYTVPQSGTFK	2	852.42139	-1.15
Q76LX8 (ATS13_HUMAN)	A disintegrin and metalloproteinase with thrombospondin motifs 13	IAIHALATNMGAGTEGANASYILIR	3	959.82507	-1.98
Q7Z7M0 (MEGF8_HUMAN)	Multiple epidermal growth factor-like domains protein 8	ALLTNVSSVALGSR	2	868.96649	-1.15
Q8TER0(SNED1_HUMAN)	Sushi, nidogen and EGF-like domain-containing protein 1	GYCLASNGSHSCTCK	3	684.27612	-0.63
Q9BXJ4 (C1QT3_HUMAN)	Complement C1q tumor necrosis factor-related protein 3	TGTVDNNTSTDLEK	2	857.89722	0.35
Q9H8L6 (MMRN2_HUMAN)	Multimerin-2	FNTTYINIGSSYFPEHGYFR	3	921.42493	-0.68
Q9HDC9 (APMAP_HUMAN)	Adipocyte plasma membrane-associated protein	AGPNGTLFVADAYK	2	886.93152	-1.8
Q9ULI3 (HEG1_HUMAN)	Protein HEG homolog 1	LNNSTGLQSSSVSQTKE	2	1000.48065	-0.31
		NSSGPDLSWLHFYR	3	676.64954	-0.25
		SHAASDAPENLTLAETADAR	3	834.73163	-1.59
Q9Y6R7 (FCGBP_HUMAN)	IgGFc-binding protein	YLPVNSSLLTSDCSER	2	1095.50769	-1.73

References

- (1) Budnik, B. A.; Lee, R. S.; Steen, J. A. J., Global methods for protein glycosylation analysis by mass spectrometry. *Biochim Biophys Acta, Proteins Proteomics* **2006**, 1764, (12), 1870-1880.
- (2) Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006**, 126, (5), 855-867.
- (3) Dwek, R. A.; Butters, T. D.; Platt, F. M.; Zitzmann, N., Targeting glycosylation as a therapeutic approach. *Nat Rev Drug Discov* **2002**, 1, (1), 65-75.
- (4) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A., Glycosylation and the immune system. *Science* **2001**, 291, (5512), 2370-2376.
- (5) An, H. J.; Miyamoto, S.; Lancaster, K. S.; Kirmiz, C.; Li, B.; Lam, K. S.; Leiserowitz, G. S.; Lebrilla, C. B., Profiling of glycans in serum for the discovery of potential biomarkers for ovarian cancer. *J Proteome Res* **2006**, 5, (7), 1626-1635.
- (6) Ruhaak, L. R.; Miyamoto, S.; Lebrilla, C. B., Developments in the Identification of Glycan Biomarkers for the Detection of Cancer. *Molecular & Cellular Proteomics* **2013**, 12, (4), 846-855.
- (7) Saldova, R.; Fan, Y.; Fitzpatrick, J. M.; Watson, R. W. G.; Rudd, P. M., Core fucosylation and alpha 2-3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia. *Glycobiology* **2011**, 21, (2), 195-205.
- (8) de Leoz, M. L. A.; Young, L. J. T.; An, H. J.; Kronewitter, S. R.; Kim, J. H.; Miyamoto, S.; Borowsky, A. D.; Chew, H. K.; Lebrilla, C. B., High-Mannose Glycans are Elevated during Breast Cancer Progression. *Molecular & Cellular Proteomics* **2011**, 10, (1).
- (9) Kyselova, Z.; Mechref, Y.; Al Bataineh, M. M.; Dobrolecki, L. E.; Hickey, R. J.; Vinson, J.; Sweeney, C. J.; Novotny, M. V., Alterations in the serum glycome due to metastatic prostate cancer. *J Proteome Res* **2007**, 6, (5), 1822-1832.
- (10) Okuyama, N.; Ide, Y.; Nakano, M.; Nakagawa, T.; Yamanaka, K.; Moriwaki, K.; Murata, K.; Ohigashi, H.; Yokoyama, S.; Eguchi, H.; Ishikawa, O.; Ito, T.; Kato, M.; Kasahara, A.; Kawano, S.; Gu, J. G.; Miyoshi, E., Fucosylated haptoglobin is a novel marker for pancreatic cancer: A detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. *Int J Cancer* **2006**, 118, (11), 2803-2808.
- (11) Shetty, V.; Nickens, Z.; Shah, P.; Sinnathamby, G.; Semmes, O. J.; Philip, R., Investigation of Sialylation Aberration in N-linked Glycopeptides By Lectin and Tandem Labeling (LTL) Quantitative Proteomics. *Anal Chem* **2010**, 82, (22), 9201-9210.
- (12) de Leoz, M. L. A.; Young, L. J. T.; An, H. J.; Kronewitter, S. R.; Kim, J.; Miyamoto, S.; Borowsky, A. D.; Chew, H. K.; Lebrilla, C. B., High-Mannose Glycans are Elevated during Breast Cancer Progression. *Molecular & Cellular Proteomics* **2011**, 10, (1).
- (13) Mann, B.; Madera, M.; Klouckova, I.; Mechref, Y.; Dobrolecki, L. E.; Hickey, R. J.; Hammoud, Z. T.; Novotny, M. V., A quantitative investigation of fucosylated serum glycoproteins with application to esophageal adenocarcinoma. *Electrophoresis* **2010**, 31, (11), 1833-1841.
- (14) Chen, R.; Wang, F.; Tan, Y.; Sun, Z.; Song, C.; Ye, M.; Wang, H.; Zou, H., Development of a combined chemical and enzymatic approach for the mass spectrometric identification and quantification of aberrant N-glycosylation. *J Proteomics*

2012, 75, (5), 1666-74.

(15)Geng, F.; Shi, B. Z.; Yuan, Y. F.; Wu, X. Z., The expression of core fucosylated E-cadherin in cancer cells and lung cancer patients: prognostic implications. *Cell Res.* **2004**, 14, (5), 423-433.

(16)Aoyagi, Y.; Suzuki, Y.; Isemura, M.; Nomoto, M.; Sekine, C.; Igarashi, K.; Ichida, F., THE FUCOSYLATION INDEX OF ALPHA-FETOPROTEIN AND ITS USEFULNESS IN THE EARLY DIAGNOSIS OF HEPATOCELLULAR-CARCINOMA. *Cancer* **1988**, 61, (4), 769-774.

(17)Li, D.; Mallory, T.; Satomura, S., AFP-L3: a new generation of tumor marker for hepatocellular carcinoma. *Clin. Chim. Acta* **2001**, 313, (1-2), 15-19.

(18)Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **2003**, 21, (6), 660-666.

(19)Lee, A.; Nakano, M.; Hincapie, M.; Kolarich, D.; Baker, M. S.; Hancock, W. S.; Packer, N. H., The Lectin Riddle: Glycoproteins Fractionated from Complex Mixtures Have Similar Glycomic Profiles. *OmicS-a Journal of Integrative Biology* **2010**, 14, (4), 487-499.

(20)Fanayan, S.; Hincapie, M.; Hancock, W. S., Using lectins to harvest the plasma/serum glycoproteome. *Electrophoresis* **2012**, 33, (12), 1746-1754.

(21)Dai, Z.; Zhou, J.; Qiu, S.-J.; Liu, Y.-K.; Fan, J., Lectin-based glycoproteomics to explore and analyze hepatocellular carcinoma-related glycoprotein markers. *Electrophoresis* **2009**, 30, (17), 2957-2966.

(22)Segu, Z. M.; Mechref, Y., Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun Mass Spectrom* **2010**, 24, (9), 1217-1225.

(23)Saba, J.; Dutta, S.; Hemenway, E.; Viner, R., Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *International journal of proteomics* **2012**, 2012, 560391-560391.

(24)Singh, C.; Zampronio, C. G.; Creese, A. J.; Cooper, H. J., Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of N-Linked Glycoproteins. *J Proteome Res* **2012**, 11, (9), 4517-4525.

(25)Scott, N. E.; Parker, B. L.; Connolly, A. M.; Paulech, J.; Edwards, A. V. G.; Crossett, B.; Falconer, L.; Kolarich, D.; Djordjevic, S. P.; Hojrup, P.; Packer, N. H.; Larsen, M. R.; Cordwell, S. J., Simultaneous Glycan-Peptide Characterization Using Hydrophilic Interaction Chromatography and Parallel Fragmentation by CID, Higher Energy Collisional Dissociation, and Electron Transfer Dissociation MS Applied to the N-Linked Glycoproteome of *Campylobacter jejuni*. *Molecular & Cellular Proteomics* **2011**, 10, (2).

(26)Jia, W.; Lu, Z.; Fu, Y.; Wang, H.-P.; Wang, L.-H.; Chi, H.; Yuan, Z.-F.; Zheng, Z.-B.; Song, L.-N.; Han, H.-H.; Liang, Y.-M.; Wang, J.-L.; Cai, Y.; Zhang, Y.-K.; Deng, Y.-L.; Ying, W.-T.; He, S.-M.; Qian, X.-H., A Strategy for Precise and Large Scale Identification of Core Fucosylated Glycoproteins. *Molecular & Cellular Proteomics* **2009**, 8, (5), 913-923.

(27)Alvarez-Manilla, G.; Atwood, J.; Guo, Y.; Warren, N. L.; Orlando, R.; Pierce, M., Tools for glycoproteomic analysis: Size exclusion chromatography facilitates

- identification of tryptic glycopeptides with N-linked glycosylation sites. *J Proteome Res* **2006**, 5, (3), 701-708.
- (28)Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T., Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* **2003**, 21, (6), 667-672.
- (29)Lin, Z.; Lo, A.; Simeone, D. M.; Ruffin, M. T.; Lubman, D. M., An N-glycosylation Analysis of Human Alpha-2-Macroglobulin Using an Integrated Approach. *J Proteomics Bioinform* **2012**, 5, 127-134.
- (30)Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, 4, (9), 709-712.
- (31)Hart-Smith, G.; Raftery, M. J., Detection and Characterization of Low Abundance Glycopeptides Via Higher-Energy C-Trap Dissociation and Orbitrap Mass Analysis. *J. Am. Soc. Mass Spectrom.* **2012**, 23, (1), 124-140.
- (32)Malerod, H.; Graham, R. L. J.; Sweredoski, M. J.; Hess, S., Comprehensive Profiling of N-Linked Glycosylation Sites in HeLa Cells Using Hydrazide Enrichment. *J Proteome Res* **2013**, 12, (1), 337-348.
- (33)Zielinska, D. F.; Gnad, F.; Wisniewski, J. R.; Mann, M., Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **2010**, 141, (5), 897-907.
- (34)Fukuda, Y.; Aguilar-Bryan, L.; Vaxillaire, M.; Dechaume, A.; Wang, Y.; Dean, M.; Moitra, K.; Bryan, J.; Schuetz, J. D., Conserved Intramolecular Disulfide Bond Is Critical to Trafficking and Fate of ATP-binding Cassette (ABC) Transporters ABCB6 and Sulfonylurea Receptor 1 (SUR1)/ABCC8. *J. Biol. Chem.* **2011**, 286, (10), 8481-8492.
- (35)Caragea, C.; Sinapov, J.; Silvescu, A.; Dobbs, D.; Honavar, V., Glycosylation site prediction using ensembles of Support Vector Machine classifiers. *BMC Bioinformatics* **2007**, 8.

Chapter 6

Conclusion

Due to the non-template nature of protein N-glycosylation, changes of the physiological and pathological environment of cells usually induce aberrations in N-glycosylation, making N-glycosylation a promising disease biomarker. In particular, alterations of N-glycosylation have been acknowledged to correlate with various types of cancers. There have been a plethora of studies on development and application of new mass spectrometric methods for improved understanding of protein N-glycosylation. Protein N-glycosylation is highly complex in that one protein may have several N-glycosylation sites, and multiple glycoforms at each site. Efforts at characterizing protein N-glycosylation involve identification of N-glycan structures, illustration of N-glycosylation sites and the associated glycoproteins, and elucidation of glycans expressed at specific sites. Quantitative analysis of protein N-glycosylation includes quantification of specific glycoforms both globally and site-specifically, measurement of glycoprotein expression level, and determination of glycosylation occupancy at each site.

Particularly, serum glycoproteomics has attracted tremendous interest due to the promise of minimally invasive methods for monitoring cancer biomarkers. This dissertation aims at development and application of multiple mass spectrometric assays for identification of pancreatic cancer-related N-glycosylation aberrations, which encompasses qualitative and quantitative analysis of N-glycosylation at the glycan, glycopeptide and partially deglycosylated peptide levels on both individual proteins and

complex protein mixtures namely human serum, with emphasis on fucosylation, a subtype of N-glycosylation.

The global glycosylation analysis normally utilizes enzymes or chemical reactions to release N-glycans from the peptides followed by separate profiling of N-glycans and N-glycosylation sites. A MALDI-MS/MS method was utilized for identification and relative quantification of haptoglobin N-glycans in pancreatic cancer as described in Chapter 2. This workflow incorporates N-glycan purification, desialylation, and permethylation, which significantly improves the ionization efficiency of N-glycans and sensitivity of the assay, leading to the new discovery of a bifucosylated triantennary glycan in haptoglobin. Moreover, a fucosylation index based on label-free quantification of glycans was constructed for a straightforward representation and comparison of fucosylation level of haptoglobin N-glycans in various samples. The change of haptoglobin fucose contents indicates a possible means for discriminating pancreatic cancer from other noncancer conditions, and further validation is needed to evaluate its usefulness for diagnostic purposes.

For global mapping of N-glycosylation sites, more specifically core-fucosylation sites, a novel endoglycosidase-assisted strategy was implemented in both individual proteins (Chapter 3) and complex protein mixtures (Chapter 5). In Chapter 3, human alpha-2-macroglobulin was chosen as the model protein, and all eight potential N-glycosylation sites were unambiguously assigned by CID-MS² (non-core-fucosylated sites) or CID-MS³ on neutral loss product (core-fucosylated sites). The extent of core-fucosylation of alpha-2-macroglobulin was examined at individual sites using a label-free strategy as detailed in Chapter 4, where core-fucosylation of Site N396, N 410

and N1424 were found to decrease in both pancreatic cancer and chronic pancreatitis relative to normal conditions. Aberrations of alpha-2-macroglobulin core-fucosylation may be useful in improving the accuracy of the pancreatic cancer diagnostics. Furthermore, this assay can be utilized to study site-specific core-fucosylation changes in other proteins for new biomarker discovery.

Chapter 5 is an extension of the study in Chapter 3 to a large-scale core-fucosylation profiling in human serum which optimizes lectin affinity strategy for glycopeptide enrichment followed by endoglycosidase digestion. An HCD-MS/MS method was utilized for profiling of core-fucosylation for the first time and provides significantly improved performance compared to the prior site-mapping work. This study is the most comprehensive profiling of the serum core-fucome to date. Furthermore, the quantitative aspect of this workflow was evaluated by incorporating a chemical isobaric labeling method iTRAQ, which is believed to be useful in high-throughput screening of core-fucosylation aberrations as potential cancer biomarkers.

Overall, this dissertation focuses on development and application of mass spectrometric assays for characterization and quantification of protein N-glycosylation, and the assays were applied in preliminary studies of pancreatic cancer biomarkers. The results are envisioned as potentially valuable for development of pancreatic cancer diagnostics. Future work to improve the assays include: (1) improvement of separation of proteins/glycopeptides/glycans prior to LC-MS/MS analysis; (2) validation of the glycosylation sites identified in Chapter 5; (3) implementing the assay developed in Chapter 5 for pancreatic cancer biomarker discovery; (4) increasing the sample cohort for biomarker discovery purposes; and (5) increase the throughput of the studies on

individual proteins using facilitated enzyme digestion and multiplexing solid phase extraction.