DISCOVERY AND CHARACTERIZATION OF LUNG CANCER BIOMARKERS USING MONOCLONAL ANTIBODY PROTEOMICS AND LIQUID CHROMATOGRAPHY - MASS SPECTROMETRY OF GLYCOPEPTIDES

A dissertation presented

by

Dongdong Wang

To
The Department of Chemistry and Chemical Biology

in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in the field of

Chemistry

Northeastern University
Boston, MA
December, 2010
DISCOVERY AND CHARACTERIZATION OF LUNG CANCER BIOMARKERS USING MONOCLONAL ANTIBODY PROTEOMICS AND LIQUID CHROMATOGRAPHY - MASS SPECTROMETRY OF GLYCOPEPTIDES

by

Dongdong Wang

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry and Chemical Biology in the Graduate School of Arts and Sciences of Northeastern University

December 2010
ABSTRACT

Current proteomics strategies for biomarker efforts have two limitations: the neglect of protein glycosylation in discovery/qualification and the lack of antibodies for verification/validation. The main goal of this thesis is to address these two problems. For the lack of high quality antibodies, monoclonal antibody (mAb) proteomics - high throughput mAb generation and disease-specific mAb screening - was applied to lung cancer biomarker discovery. mAb proteomics is demonstrated to generate high quality mAbs in a high throughput manner, an important characteristic for biomarker verification/validation. For glycosylation characterization, mass spectrometry (MS)-based methods for sensitive, site-specific discovery and quantitation were developed and applied to characterize the glycosylation changes of the lung cancer biomarker candidate haptoglobin. These developed techniques for glycosylation characterization can be also applied to other glycoproteins and therefore will contribute to the emerging glyco biomarker field.

In Chapter 2, together with Biosystem International (BSI), we generated an mAb library using mAb proteomics against lung cancer patient plasma protein mixtures. Cancer specific mAbs were identified by high throughput screening, and three mAbs were produced in large quantity for antigen (Ag) identification purposes. An MS-based strategy combined with immunoassay was developed for Ag identification. One of the antigens was found to be plasma glycoprotein haptoglobin (Hpt). The anti-Hpt Ab-Ag interaction was characterized using various techniques, and the mAb was found to have higher affinity for native plasma Hpt than the corresponding reduced form. Furthermore, the glycan on the Hpt was found to be essential for the Ab-Ag interaction. The characterization of mAbs in this work revealed that the global mAb
proteomics process can generate high-quality lung cancer specific mAbs capable of recognizing proteins in their native state.

In Chapter 3, a robust, ultrasensitive LC-MS platform, combining the strength of 10 µm i.d. porous layer open tubular (PLOT) LC column with linear ion trap-collision induced dissociation / electron transfer dissociation - mass spectrometry (LTQ-CID/ETD-MS), was successfully developed for site-specific protein glycosylation characterization. Using Hpt as a proof-of-concept model glycoprotein, glycopeptide identification, glycosylation site elucidation, glycan quantitation and site occupancy determination, were demonstrated to be completed within 10 LC-MS runs at a total sample consumption of 100 fmol (~ 13 ng Hpt). The developed platform was shown to have sufficient sensitivity to allow characterization of site-specific protein glycosylation from trace amounts of glycosylated proteins.

In Chapter 4, a glycopeptide selected reaction monitoring (SRM) method was developed to sensitively and reproducibly quantify protein glycosylation isoforms in a site-specific manner. A feasibility study was first performed to ascertain the potential of glycopeptide SRM analysis using Chip LC-quadrupole time-of-flight mass spectrometer (Chip LC-qToF) and Chip LC-triple quadrupole mass spectrometer (Chip LC-QQQ). Glycopeptide fragmentation pattern was systematically studied, and glycan oxonium ions were determined to be preferred SRM product ions. The glycopeptide SRM method featured high sensitivity, high reproducibility and large dynamic range. Site-specific glycosylation differences of Hpt, affinity purified from individual non-small cell lung cancer, healthy controls and inflammation controls (rheumatoid arthritis) plasma samples, were revealed using the glycopeptide SRM method. Further verification/validation of these discovered site-specific Hpt glycosylation isoforms may increase the specificity of Hpt as a lung cancer biomarker.
ACKNOWLEDGEMENTS

The completion of this dissertation reflects the support I received throughout my graduate career from many people. I would like to give many thanks to the following individuals, although I know it is impossible for me to identify all.

In the first place I would like to record my gratitude to my advisor, Dr. Barry L. Karger, for your supervision, advice, and guidance in every stage of this research and so to this thesis. Above all, you provided me unflinching encouragement and support in various ways. Your scientist intuition has made you as a constant oasis of ideas and passions in science, which exceptionally inspire and enrich my growth as a student, a researcher and a scientist I want to be. I feel very privileged to have such an awesome advisor.

Many thanks go in particular to Dr. Marina Hincapie, Dr. Tomas Rejtar, Dr. Billy Wu and Dr. Andras Guttman. I would like to thank you for your assistance and support. The great intellectual conversations and your constructive comments immensely help me on my projects. I am lucky enough to have the opportunity to work together with you during my graduate career.

I gratefully thank Dr. Shujia (Daniel) Dai, Dr. Quanzhou Luo, Dr. Guihua (Eileen) Yue, Dr. Sangwon Cha, Dr. Zoltan Szabo for your valuable advice in science discussion and careful explanation of detailed experiments.
I also benefited from working with my labmates, from both Dr. Karger and Dr. Hancock’s groups: Dr. Lingyun Li, Dr. Ye Gu, Dipak Thakur, Zhenke (Jack) Liu, Siyang (Peter) Li, Chen Li, Wenqin Ni, Dr. Xiaoyang Zheng, Dr. Haitao Jiang, Qiaozhen (Cheryl) Lu, Majlinda (Linda) Kullolli, Zhi (Janet) Zeng, Rafalko Agnes, Yi Wang, Fateme Tousi, Suli Liu, Yue Zhang, Fangfei Yan, for your tremendous help, scientific discussion, and friendship. I couldn’t ask for a better lab group.

To my committee members, Dr. Zhaohui (Sunny) Zhou, Dr. John R. Engen, Dr. Graham B. Jones, for your invaluable feedbacks, helpful advice, and sacrifice of precious time.

I would also acknowledge the Department of Chemistry and Chemical Biology staffs, colleagues, and friends: Dr. William S. Hancock, Dr. David E. Budil, Dr. David A. Forsyth, Dr. Ira S. Krull, Jeffrey Kesilamn, Felicia Hopkins, Richard Pumphrey, Jean Harris and Jana Volf, for all of your help.

My special thanks to the protein analytics group, process department at Abbott: Dr. Hongcheng Liu, Dr. Czeslaw Radziejewski, Gary Welch, Chris Chumsae and Kathreen Gifford, for your support at the beginning of my professional career.

Finally and most importantly, I would like to thank my wife and my parents. Their support, encouragement, patience and love were undeniably the bedrock upon which my life has been built. It was under their watchful eyes that I gained the ability to tackle challenges ahead.
TABLE OF CONTENTS

Abstract ................................................................................................................. 2
Acknowledgements ................................................................................................. 4
Table of Contents ..................................................................................................... 7
List of Figures ......................................................................................................... 13
List of Tables .......................................................................................................... 16

Chapter 1  Overview of technologies: antibody library generation and protein
glycosylation characterization.............................................................................. 17
  1.1 Abstract........................................................................................................... 18
  1.2 Biomarker....................................................................................................... 19
    1.2.1 Biomarker discovery and qualification....................................................... 19
    1.2.2 Biomarker verification and validation......................................................... 21
    1.2.3 Limitations for the current proteomics method.......................................... 24
  1.3 Antibody library technique............................................................................. 25
    1.3.1 Polyclonal and monoclonal antibody......................................................... 25
    1.3.2 Polyclonal and monoclonal antibody production....................................... 26
    1.3.3 Antibody library technique........................................................................ 29
      1.3.3.1 Monospecific polyclonal Abs............................................................... 29
      1.3.3.2 Monoclonal Ab proteomics................................................................. 30
  1.4 Protein glycosylation....................................................................................... 33
1.4.1 Monosaccharides and N-linked glycans ................................. 33
1.4.2 Glycan synthesis ................................................................ 36
1.4.3 Glycosylation changes in cancer ........................................... 38
1.5 Techniques for protein glycosylation characterization ............. 40
1.5.1 Intact glycoprotein ............................................................... 40
1.5.1.1 Lectin based technique ...................................................... 41
1.5.1.2 Capillary electrophoresis ..................................................... 41
1.5.1.3 MS analysis ................................................................. 42
1.5.2 Glycan analysis ................................................................. 44
1.5.2.1 N-Glycan release from protein backbone ......................... 45
1.5.2.2 Glycan derivatization ......................................................... 47
1.5.2.3 High pH anion exchange chromatography - pulsed amperometric detection ................................................................. 49
1.5.2.4 Capillary electrophoresis - laser induced fluorescence ........... 49
1.5.2.5 Hydrophilic interaction chromatography ............................... 49
1.5.2.6 Reverse phase liquid chromatography ................................ 50
1.5.2.7 Porous Graphitized Carbon .............................................. 50
1.5.2.8 MS based technique ......................................................... 51
1.5.3 Glycopeptide analysis ......................................................... 54
1.5.3.1 Glycoprotein protease digestion ..................................... 54
1.5.3.2 MS/MS fragmentation techniques ..................................... 55
1.5.3.3 LC-MS analysis of glycopeptides ..................................... 60
1.6 Glycoproteomics studies ....................................................... 63
1.6.1 Proteocentric glycoproteomics................................................. 63
1.6.2 Glycocentric glycoproteomics.............................................. 63
1.6.3 Monoproteic glycoproteomics............................................... 64
1.7 References................................................................. 67

Chapter 2 Antigen identification and characterization of lung cancer specific
monoclonal antibodies produced by mAb proteomics................. 77
2.1 Abstract............................................................... 78
2.2 Introduction.......................................................... 79
2.3 Experimental section.................................................... 81
2.3.1 Materials.......................................................... 81
2.3.2 Plasma sample collection and immunogen preparation........... 81
2.3.3 Mouse immunization, hybridoma generation and mAb screening..... 82
2.3.4 Immunoaffinity purification.......................................... 83
2.3.5 SDS-PAGE and Western blotting..................................... 84
2.3.6 In-gel and in-solution tryptic digestion............................... 85
2.3.7 LC-ESI-MS and MALDI-TOF-MS analysis.......................... 86
2.3.8 Deglycosylation and glycan-specific staining......................... 87
2.3.9 Surface plasmon resonance (SPR) analysis........................... 87
2.3.10 Sandwich ELISA................................................ 88
2.3.11 Hybridoma cell culture and mAb purification from supernatant... 89
2.4 Results and discussion.................................................. 91
2.4.1 Hybridoma generation and screening................................ 93
2.4.2 Antigen identification for 3 lung cancer specific mAbs………………… 95
2.4.3 Characterization of 2 hybridomas…………………………………… 102
2.4.4 Western blotting………………………………………………………. 106
2.4.5 Surface plasmon resonance…………………………………………… 111
2.4.6 Sandwich ELISA of Hpt……………………………………………… 113
2.5 Conclusions…………………………………………………………… 117
2.6 References……………………………………………………………. 118

Chapter 3 Ultrasensitive characterization of site-specific glycosylation for
affinity purified haptoglobin from lung cancer patients plasma using
10 µm i.d. PLOT LC-LTQ-CID/ETD-MS…………………………… 123
3.1 Abstract……………………………………………………………… 124
3.2 Introduction………………………………………………………….. 125
3.3 Experimental………………………………………………………… 128
3.3.1 Samples and reagents………………………………………………… 128
3.3.2 Haptoglobin immunoaffinity purification…………………………….. 128
3.3.3 Trypsin and PNGase-F digestion……………………………………. 129
3.3.4 PLOT LC-LTQ-CID/ETD-MS……………………………………….. 130
3.3.5 Data processing………………………………………………………. 132
3.4 Results and discussion……………………………………………… 133
3.4.1 Haptoglobin affinity purification…………………………………….. 136
3.4.2 Glycopeptides T1 and T3…………………………………………….. 138
3.4.3 Core/Antenna fucosylation isomer identification…………………. 146
Chapter 4  Site-specific protein glycosylation quantification using glycopeptide selected reaction monitoring on Chip LC-QQQ platform  164
  4.1  Abstract  165
  4.2  Introduction  166
  4.3  Experimental  168
    4.3.1  Plasma samples and reagents  168
    4.3.2  Haptoglobin purification  169
    4.3.3  Haptoglobin digestion  170
    4.3.4  HPLC-Chip-MS system  170
    4.3.5  qToF-MS/MS  171
    4.3.6  QQQ-MS/MS  172
  4.4  Results and Discussion  173
    4.4.1  Glycopeptide SRM development strategy  173
    4.4.2  Glycopeptide fragmentation pattern in the quadrupole mass spectrometers  177
    4.4.3  Transitions for glycopeptide and peptide SRM  183
    4.4.3.1  Transitions for glycopeptide SRM  183
# LIST OF FIGURES

| Figure 1-1 | Pipeline for the development of novel protein biomarkers | 23 |
| Figure 1-2 | Antibody production workflow | 28 |
| Figure 1-3 | mAb proteomics workflow: hybridoma generation and mAb screening | 32 |
| Figure 1-4 | Biosynthesis of the N-linked glycan | 37 |
| Figure 1-5 | Multiple mechanisms regulating the formation of glycans | 39 |
| Figure 1-6 | Intact glycoprotein analysis | 43 |
| Figure 1-7 | Common derivatization for glycan | 48 |
| Figure 1-8 | Glycan analysis using CE, LC and MS | 53 |
| Figure 1-9 | Glycopeptide analysis using different MS/MS technique | 59 |
| Figure 1-10 | PLOT LC-ESI-MS analysis at ultralow flow rate (20 nL/min) | 62 |
| Figure 1-11 | Representative glycoproteomics studies | 66 |
| Figure 2-1 | General workflow for lung cancer biomarker discovery using mAb proteomics | 92 |
| Figure 2-2 | Antibodies and hybridomas characterized in this work | 94 |
| Figure 2-3 | Antigen identification workflow | 98 |
| Figure 2-4 | Antigen identification for mAb #1 | 99 |
| Figure 2-5 | Antigen identification for mAb #2 | 100 |
| Figure 2-6 | Antigen identification for mAb #3 | 104 |
| Figure 2-7 | Antibody purification from hybridoma supernatant and mAb reactivity test using Western blotting | 105 |
| Figure 4-4 | The effect of charge state on glycopeptide fragmentation……….. | 180 |
| Figure 4-5 | The effect of glycan structure on glycopeptide fragmentation…… | 181 |
| Figure 4-6 | The effect of peptide sequence on glycopeptide fragmentation…….. | 182 |
| Figure 4-7 | Glycopeptide and peptide SRM……………………………………… | 187 |
| Figure 4-8 | Collision energy (CE) influence on oxonium ion intensity………… | 191 |
| Figure 4-9 | The influence of fragmentor voltage on glycopeptide SRM………. | 193 |
| Figure 4-10 | Most intense charge state determination………………………… | 196 |
| Figure 4-11 | Four orders magnitude of linearity (50 amol to 500 fmol) is achieved using the developed glycopeptide SRM……………… | 199 |
| Figure 4-12 | Site-specific glycoforms quantitation by SRM on individual Hpt samples………………………………………………………….. | 200 |
LIST OF TABLES

Table 1-1  A) Common monosaccharides in N-linked glycans; B) Types of N-linked glycans……………………………………………………….. 35
Table 1-2  Enzymes used for glycan release………………………………………………………………………………………………….. 46
Table 1-3  MS/MS techniques for glycopeptide analysis……………………………………………………………………………… 58
Table 2-1  Hybridoma tissue culture medium with HAT supplement…………. 90
Table 2-2  Proteins identified from the IP pulldown mixture using mAb #1…… 96
Table 2-3  Hpt concentration determined from individual lung cancer and healthy plasma……………………………………………………….. 116
Table 3-1  Haptoglobin tryptic glycopeptides………………………………………………………………………………………………….. 137
Table 3-2  Theoretical mass and m/z matrix for the identified site-specific glycoforms……………………………………………………….. 143
Table 3-3  Site-specific glycan structures identified on Hpt………………………………. 144
Table 4-1  The 26 glycoform masses and m/z values (z=3, 4, 5) for the Hpt samples…………………………………………………………… 185
Table 4-2  Amino acid compositions with nominal masses of 204, 274 and 366……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………
Chapter 1

Overview of Technologies:
Antibody Library Generation and Protein Glycosylation Characterization
1.1 Abstract

This chapter provides an introduction to the principles and technologies involved in research conducted in this thesis. One important application of mass spectrometry-based proteomics is the discovery of protein biomarkers for the early diagnosis of human disease. An overview of the concepts and workflow in current proteomics biomarker efforts is first provided, with the focus on the 2 challenges: lack of high quality immunoaffinity agents for biomarker verification/validation and neglect of protein post translational modification (e.g. glycosylation) for discovery/qualification.

Immunoaffinity agents are discussed in detail including polyclonal antibody (pAb), monoclonal antibody (mAb), their specificity and production, with a focus on emerging high throughput Ab library generation technique such as mAb proteomics. The introduction of protein glycosylation includes the structure, in vivo synthesis of glycans and the relationship of specific structures to cancer, followed by an overview of techniques used for glycosylation characterization. The discussed techniques are categorized by their applicable targets, namely intact glycoprotein, released glycan and glycopeptides. Finally, the 3 glycoproteomics research fields (proteocentric, glycocentric and monoproteic) are discussed.
1.2 Biomarker

A biomarker is a molecule used as an indicator of a biological state. It is a characteristic that can be measured and evaluated for normal biological processes, pathogenic development, or pharmacologic responses to a therapeutic intervention. A biomarker can be any type of biological molecule, such as protein, DNA/RNA, metabolite or a pattern of molecules, and can take many forms from these molecules. Among the types of biomarkers, proteins control many biological activities and can be broadly affected in disease. Proteomics thus holds great promise for biomarker discovery. In the following, current protein biomarker pipelines (Figure 1-1) including 4 phases (discovery, qualification, verification and validation) are briefly described.

1.2.1 Biomarker discovery and qualification

Protein biomarker “discovery” is non-targeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) profiling and relative abundance comparison to identify differentially expressed proteins among various biological/pathological states (usually 5-10 samples). Many types of samples including cell lines, plasma/sera and tissues have been used for biomarker discovery. Among these, plasma/serum is the most widely used, mainly due to its easy accessibility and accepted application to diagnostics.

For biomarker discovery, the workflow usually starts with enzymatic protein digestion (mostly by trypsin) to generate constituent peptides. The generated peptides are separated by LC and the amino acid sequences are determined by mass spectrometry. In MS/MS experiment, the peptide precursors are fragmented by collision induced dissociation (CID) to produce product ions. The resulting MS/MS spectra are searched against protein databases. The peptide identification is achieved by correlating the experimentally observed MS/MS fragments with the
theoretical MS/MS spectra. Protein identities are assigned based on the identified peptides and protein abundances are compared (normal vs. disease) using methods such as “spectral counts”\(^8\) to identify differentially expressed proteins.

The results of biomarker “discovery” experiments are a list of protein biomarkers (usually > 10s but < 100) with their expression ratios, based on the relative quantitation of the MS data. Some of these discovered biomarkers are false positives. False positives are proteins which were determined to be up/down-regulated proteins but are actually not. The high false-discovery rate is not only due to individual variability across the human samples (caused by race, age, gender etc) but also simply due to analytical bias\(^1\). The LC-MS technique used in “discovery” methods is data-dependent, i.e. in each measurement it samples a fraction of the peptides based on their relative intensities. Data-dependent LC-MS technique usually biases towards high abundance peptides. Also, multiple data-dependent LC-MS analysis are required to compare protein expression across different samples: for each analysis, only a subset of the peptides will be sampled, thus make it challenging to generate complete and consistent data sets\(^1\). Also, for low abundance targets, their signals may randomly exceed the detection limits due to the limited instrument dynamic range and sensitivity and result in artificial differences. Due to these reasons, these discovered biomarkers are referred to as “potential biomarkers” or “biomarker candidates”\(^1\).

After the non-targeted discovery, these “biomarker candidates” will be transferred to the “qualification” phase. “Qualification” is the assessment of the differential expression observed in the “discovery” phase using targeted MS methods such as selected reaction monitoring (SRM)\(^9\). Selected reaction monitoring is usually performed with triple quadrupole instrument. In SRM, a pre-defined precursor is selected in quadrupole 1 (Q1), fragmented by CID in
quadrupole 2 (Q2), and one or several pre-defined fragments detected in quadrupole 3 (Q3). The signal intensity of the transition (precursor to fragment) is used to represent the abundance of peptide and further the protein thereof. Combined with LC retention time, the tandem mass selection endows better selectivity and sensitivity, compared to non-targeted MS methods.

The change from non-targeted data-dependent MS profiling to targeted SRM quantitation brings in improved reproducibility, sensitivity and enhanced linear dynamic range. For both “discovery” and “qualification” phases, the primary focus is to ascertain the consistent association between the identified “biomarker candidates” and disease. In both phases, the principal goal is to achieve the highest sensitivity: disease state will be identified using these discovered biomarkers (sensitivity). The consistent “biomarker candidates” should also be analyzed using biological pathway analysis, where applicable, to determine whether or not these candidates make biological sense. The “candidates”, which show consistent ratio changes and also make biological senses, are passed on to “verification” and further “validation” phases.

1.2.2 Biomarker verification and validation

In the “verification” phase, immunoassays such as enzyme-linked immunosorbent assay (ELISA, Figure 1-1) is used to test a larger number of human samples (approximately over hundreds, cancer v.s normal). ELISA is a gold standard quantitative assay for biomolecule measurement. Two antibodies are required for an ELISA assay: one antibody is immobilized onto solid surface (usually by passive adsorption) for capture purpose, while the second one is labeled with fluorescence tag or enzyme (for example horseradish peroxidase) for detection purpose. It is in this phase that antibody is replacing MS as the working horse for biomarker efforts. Antibody-based techniques are known to have high signal-to-noise ratio (fluorescence or
chemiluminescence detection), enhanced throughput (96 or 384 plates), multiplexing capability and is widely accepted in clinical environment. In the “verification” phase, a broader range of cases (for example, different stage of disease) and controls involving environmental, genetic and biological variations are tested. A profound change of “verification” is the sample variation, which was carefully minimized by choosing well-matched disease and control in “discovery” and “qualification”, is re-introduced. Therefore, specificity - the likelihood that a normal sample will give negative signal using biomarker candidates - began to be assessed. After being screened in the “verification” phase, only a few candidates (< 5) will suffice the requirement for disease diagnosis or prognosis monitoring purpose. These verified candidates will be subjected to formal “validation”. In “validation”, at least thousands of samples representing full variation for the targeted population are tested. The purpose is to determine whether or not these verified biomarker candidates can became a “validated” biomarker1.
Figure 1-1. Pipeline for the development of novel protein biomarkers. This figure is modified from Rifai\textsuperscript{1}. 

\begin{itemize}
  \item **Discovery**
    \begin{itemize}
      \item Non-targeted LC-MS/MS (Data-dependent profiling)
    \end{itemize}
  \item **Qualification**
    \begin{itemize}
      \item Targeted LC-MS/MS (Selected reaction monitoring)
    \end{itemize}
  \item **Verification**
    \begin{itemize}
      \item Prototype Immunoassay (Ab-based technique)
    \end{itemize}
  \item **Validation**
    \begin{itemize}
      \item High throughput Immunoassay
    \end{itemize}
\end{itemize}

Analytes:
- Discovery: 100s-1000
- Qualification: <100
- Verification: 10s
- Validation: <5

Samples:
- Discovery: 5-10
- Qualification: <100
- Verification: 100s
- Validation: 1000s
1.2.3 Limitations for the current proteomics method

Despite intense interest and investment for biomarker efforts, few novel biomarkers have been adopted in clinical practice. Many reasons are contributing to this limited success. Here, two limitations of current biomarker workflow are pointed out and discussed in detail.

First, for the “discovery” and “qualification” phases, protein identification and relative quantitation are based on digested peptides. These measured, non-modified peptides serve as surrogates for their corresponding proteins. However, a human protein is more than a simple combination of unmodified peptides. There are several hundred co- and post-translational modifications that will affect protein structure, function and interaction\(^\text{10}\). A prominent example is phosphorylation. The attachment/detachment of a phospho group to a specific Ser/Thr/ Tyr of a protein can serve as function switch and change the protein function significantly\(^\text{10}\). Monitoring non-modified peptides will omit the information conveyed by phosphorylation. Another example, and a far more complex one, is protein glycosylation\(^\text{11}\). Aberrant glycosylation has been recognized as a hallmark of cancer for more than 30 years\(^\text{12}\). However, protein glycosylation is often neglected in biomarker research, mainly due to its complex nature. Compared to phosphorylation, which has already been intensively characterized using mass spectrometry\(^\text{10}\), glycosylation structure analysis using MS is still under development and its incorporation into biomarker efforts at a relatively early stage\(^\text{12}\).

Second, for the “verification” and “validation” phases, well characterized high quality antibodies are essential. Also, due to the requirement of an ELISA sandwich assay, multiple antibodies targeting different epitopes of a protein are necessary. However, for most biomarker studies, there are often no available antibodies for qualified “biomarker candidates”. The
generation of high quality antibodies can be quite time-consuming, costly, and the final outcome is unpredictable. Typically, antibody is produced against one antigen at a time; thus, technologies that can generate libraries of Abs within reasonable time frame are ideal platforms for biomarker research. A number of methods for mAb library generation have been proposed\textsuperscript{13,14}, and some of the most important are described below.

1.3 Antibody library technique

1.3.1 Polyclonal and monoclonal antibody

Polyclonal antibodies (pAbs) are antibodies produced by multiple B-cells\textsuperscript{15}. Polyclonal antibodies target multiple epitopes (e.g. both linear and conformation epitopes) on the same antigen and recognize antigen from different orientations. This characteristic makes pAbs attractive reagents for detection or affinity purification of its associated antigen from complex matrix. Compared to monoclonal antibodies (mAbs), which are produced from single clone of B-cell through hybridoma technique\textsuperscript{15}, pAbs are relatively easy and cost-effective to produce in short period. However, because pAbs are produced by immunization of living animal, batch-to-batch variation is inevitable, even when the immunization is made to the same animal\textsuperscript{16}. This variation includes differences in antibody reactivity (i.e. specificity and affinity) and titre, and thus pAbs in general suffer from bad reproducibility. Due to this reason, pAbs are usually not the choice of reagents for clinical practice. Monoclonal antibodies, on the contrary, are targeting a single epitope with consistent and defined specificity. The continuous culture of hybridomas offers a reproducible and potentially inexhaustible supply of mAbs\textsuperscript{15}. Consequently, mAbs enable the development of standardized immunoassay systems with consistent and predictable quality, which is essential for a biomarker be accepted in clinical diagnostic test.
1.3.2 Polyclonal and monoclonal antibody production

Typically, antibody is produced against one antigen at a time\(^{15}\) (Figure 1-2). For pAb, the production starts with the immunization of animals (usually rabbit, goat or horse) with pure antigen using Freund's adjuvant (emulsion with mineral oil) or other adjuvants\(^ {17}\). Adjuvant is an immunological agent that enhances the effect of antigens while having little effects when given by itself. Adjuvant is often included to augment the animal’s immune response\(^ {17}\). After immunization (usually span several weeks with repeated immunization), blood samples are obtained for measurement of serum antibodies using ELISA. Usually large animals are used for pAb production, simply because large animal will enable the collection of large volume of antibody-rich blood/serum. The collected blood/serum can be used directly for detection purpose or the antibody can be purified based on the downstream application requirement.

For mAb, the production also starts with the animal immunization with the antigen, but in this case, mouse is the most widely used (Figure 1-2-B)\(^ {16}\). Mouse is repeated immunized and, when the antibody titer is high enough, as determined by ELISA, the mouse is euthanized and its spleen is removed to generate hybridoma cell\(^ {16}\). Hybridoma cell is generated by fusing antibody-producing spleen cells, which have a limited life span, with immortal myeloma cells (cancer cells) (Figure 1-2-B). The fusion of spleen and myeloma cells is achieved by co-centrifugation with polyethylene glycol. The resulting cell mixture (spleen cell, myeloma cell and hybridoma cell) are cultured in hypoxanthine-aminopterin-thymidine (HAT) selection medium, in which only the hybridoma will survive\(^ {16}\). The hybridoma cells are distributed into 96/384 well plates with the supply of growth factors (usually from fetal bovine serum) that promote their growth. A so-called “limiting dilution” at this time will ensure that only single clone exists in one well. Tissue culture of the single clone will produce supernatants containing the mAb targeting a
single epitope (Figure 1-2-B). In other case, the hybridoma can be re-injected into mouse to produce ascites fluid, which has a higher yield for the mAb production than hybridoma tissue culture\textsuperscript{16}.
Figure 1-2. Antibody production workflow. (A) Polyclonal antibody production. (B) Monoclonal antibody production. For traditional antibody (pAb and mAb) production, the antibody is produced against one antigen at a time.
1.3.3 Antibody library technique

Biomarker “verification” and “validation” needs large number of high quality immunoaffinity reagents, thus technologies that can generate libraries of Abs within reasonable time are ideal platforms for biomarker research. Several initiatives are underway to produce antibody libraries including the HUPO antibody initiative \(^{18}\) and ProteomeBinders\(^{19}\), etc. Both pAb and mAb libraries have been generated. Based on the research purpose of the generated libraries, they can be roughly divided into 2 types. For one type, the purpose is to generate as many Abs as possible and serve as a resource for others, as may be needed for biomarker “verification” and “validation”. While for the 2\(^{nd}\) type, the purpose is not only to generate but also to screen an Ab library to identify cancer-specific Abs. The 1\(^{st}\) type is best represented by monospecific polyclonal Abs\(^{14}\), while the 2\(^{nd}\) type is typified by mAb proteomics\(^{13}\), developed by our collaborators at Biosystems International. In the following, these two techniques are briefly described.

1.3.3.1 Monospecific polyclonal Abs

The monospecific polyclonal antibodies (msAbs) are essentially pAbs generated against synthetic short peptides, namely protein epitope signature tags (PrESTs)\(^{14}\). The PrESTs are usually 6-10 amino acids long synthetic peptides, which are designed / synthesized to contain unique epitopes present on protein surfaces suitable for initiating immune response. These synthetic peptides are used in immunization, screening and the purification of the pAbs. Because the immunization is performed using a complex peptides mixture (not pure antigen), therefore, a large number of pAbs targeting many peptides could be generated in a short period. The pAbs,
which are targeting the same peptide, are affinity purified using the individual PrEST as the
ligand from polyclonal sera.

The purified msAbs have been used to systematically profile the gene products of
chromosome 21 in human tissue samples\textsuperscript{14}. This approach is promising, but since peptides are
used for immunization, the antibody may exhibit reduced binding to the epitope when present in
a three dimensional structure (native protein). Another drawback of such approach is that post-
translational modifications, such as glycosylation, which are important for protein function, are
given no consideration. In addition, the production of the msAb is non-reproducible, even from
the same animal, since msAbs are essentially polyclonal antibodies\textsuperscript{15}.

1.3.3.2 Monoclonal Ab proteomics

Compared to pAbs, mAbs can be produced continuously with defined characteristics in a
reproducible manner, and are powerful reagents for biomarker research\textsuperscript{15}. One challenge for
large scale mAb library generation is to produce a large number of pure immunogens with
sufficient amount to suffice the requirement of mouse immunization, hybridoma screening and
mAb characterization\textsuperscript{20, 21}. Recently, our collaborator Biosystem International developed a novel
mAb library method combining the high throughput mAb generation and cancer-specific mAb
screening (Figure 1-3)\textsuperscript{13}. In this scheme, the mouse is immunized with fractionated plasma
protein mixture to generate a hybridoma library (mAb library thereof). The mAbs are subjected
to high throughput ELISA screening to identify the cancer-specific mAbs, although at this stage,
the identities of the antigens (protein biomarkers) causing the differentiation between the cancer
and control are unknown.
For a given antigen, the immune response is proportional to the relative concentration of the antigen and its immunogenicity (often unknown)\textsuperscript{13}. To generate antibodies against majority low abundant plasma proteins, high abundant protein needs to be depleted. The immunogen mixture is prepared from pooled cancer patients’ plasma using 3-step affinity chromatography fractionation. First, the amount of high concentration plasma proteins is decreased using a multi-affinity removal system (MARS) column. Second, the glycoprotein contents from depleted plasma sample is enriched using a multi-lectin affinity chromatography (M-LAC) column\textsuperscript{22, 23}. Last, the enriched M-LAC material is further normalized using a patented rabbit anti-human pAb column\textsuperscript{24}.

In this thesis, the 3-step plasma fractionation was applied to pooled non-small cell lung cancer patients plasma\textsuperscript{25}. The resulting (glyco)-protein mixture was used to immunize mice and generate mAb-producing hybridomas (Figure 1-3-A). Using high throughput ELISA screening against lung cancer and normal plasma sample (Figure 1-3-B), the lung cancer specific mAbs were identified before their antigen identities were determined. In chapter 2, the antigen identification and characterization for 3 lung cancer specific mAbs (indicated in Figure 1-3-C) are described.
Figure 1-3. mAb proteomics workflow: hybridoma generation and mAb screening. (A) The 3-steps fractionated lung cancer plasma protein mixture was used to immunize mice using complete (first immunization) and incomplete Freund’s adjuvant (subsequent immunizations). Hybridoma libraries were generated through fusions of SP2/0-Ag14 hybridoma partner cells (myeloma) with spleen cells from immunized mouse. (B) Indirect capture ELISA was used to screen the mAb produced by hybridomas. Each hybridoma was tested with two tracers (cancer and normal). Tracers are biotinylated plasma protein mixture. (C) Results for mAb screening using indirect capture ELISA. Each dot represents one mAb, and bold dots represent mAb specific for pooled cancer plasma (Y axis) or normal (X axis). mAb #1, 2, and 3 were produced from ascites fluids and were subject to antigen identification, mAb characterization and immunoassay development, as described in detail in Chapter 2.
1.4 Protein glycosylation

Glycosylation is one of the major co-/post-translational modifications of proteins, involving the attachment of glycans to the protein backbone through asparagine (N-linked) or serine/threonine (O-linked). Glycosylation is important for protein folding, subcellular localization, turnover and activity, and contributes significantly to the cell proliferation and interaction. Most secreted proteins such as those circulating in plasma are usually glycosylated. For protein biomarker discovery, glyco-based marker is lagging far behind than peptide based, mainly due to the complexity of glycosylation in nature. However, the rapid development of techniques for glycosylation characterization are changing the landscape.

The work presented in this thesis focuses on the characterization of N-linked protein glycosylation. In the following, N-linked protein glycosylation is introduced. The topics covered include the commonly encountered N-linked glycan structure, its synthesis, function and current understanding of glycosylation-disease relationship.

1.4.1 Monosaccharides and N-linked glycans

The commonly encountered monosaccharides in N-linked glycans are listed in Table 1-1-A. N-linked glycans contain N-acetylglucosamine (GlcNAc) linked via an amide bond to asparagine residue of a protein. There is a sequence motif for N-linked glycosylation Asn-X-Ser/Thr, where X can be any amino acid except proline. Asn-X-Cys can also result in N-linked glycosylation, but in a much lower frequency. All the N-linked glycans share a trimannosyl chitobiose core structure, as shown in Table 1-1-B. Multiple antennae (usually 2-4) extends from the core structure. Based on the compositions of antennae, the N-linked glycans can be
classified into three subgroups namely high mannose, complex and hybrid type (Table 1-1-B). The glycans that contain only mannose residues in their antenna region are classified as “high mannose type”, while those containing N-acetyllactosamine (Galα1-3/4 GlcNAc) are called “complex type”. The “hybrid type” contains both mannose residues and N-acetyllactosamine attached to the core structure. Monosaccharides such as fucose (Fuc) and N-acetyleneuraminic acid (NeuAc) could also decorate the glycans and are most often α-linked to GlcNAc residue.
### Table 1-1-A. Common monosaccharides in N-linked glycans

<table>
<thead>
<tr>
<th>Monosaccharide</th>
<th>Structure*</th>
<th>Donor#</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mannose (Man)</td>
<td><img src="image1" alt="Mannose Structure" /></td>
<td>GDP-Man</td>
<td><img src="image2" alt="Green" /></td>
</tr>
<tr>
<td>Galactose (Gal)</td>
<td><img src="image3" alt="Galactose Structure" /></td>
<td>UDP-Gal</td>
<td><img src="image4" alt="Yellow" /></td>
</tr>
<tr>
<td>N-acetylglucosamine (GlcNAc)</td>
<td><img src="image5" alt="N-acetylglucosamine Structure" /></td>
<td>UDP-GlcNAc</td>
<td><img src="image6" alt="Blue" /></td>
</tr>
<tr>
<td>N-acetyleneuraminic acid (NeuAc)</td>
<td><img src="image7" alt="N-acetyleneuraminic acid Structure" /></td>
<td>CMP-NeuAc</td>
<td><img src="image8" alt="Purple" /></td>
</tr>
<tr>
<td>Fucose (Fuc)</td>
<td><img src="image9" alt="Fucose Structure" /></td>
<td>GDP-Fuc</td>
<td><img src="image10" alt="Red" /></td>
</tr>
</tbody>
</table>

* Cyclic forms of the monosaccharides are shown here. Cyclic structure is formed by the reaction between the aldehye group at C-1 (C-2 for NeuAc) and the hydroxyl group at C-5 (C-6 for NeuAc). This reaction produces either of two stereoisomers, the α and β anomers, which differ only in the stereochemistry around the hemiacetal carbon C-1 (C-2 for NeuAc).

# Donor is used by the glycosyltransferases for glycan synthesis.

### Table 1-1-B. Types of N-linked glycans

<table>
<thead>
<tr>
<th>Type</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Mannose</td>
<td><img src="image11" alt="High Mannose Structure" /></td>
</tr>
<tr>
<td>Complex</td>
<td><img src="image12" alt="Complex Structure" /></td>
</tr>
<tr>
<td>Hybrid</td>
<td><img src="image13" alt="Hybrid Structure" /></td>
</tr>
</tbody>
</table>

* The figures for glycan structures are adopted from Varki.  

---

35
1.4.2 Glycan synthesis

Glycan synthesis involves complex and sequential enzymatic reactions to add (by glycosyltransferase) or remove (by glycosidase) specific monosaccharides from the processed polypeptide/protein backbone\textsuperscript{11}. Two cell organelles are heavily involved in protein glycosylation synthesis, namely endoplasmic reticulum (ER) and Golgi, with glycosyltransferase and glycosidase bounded to the membranes of both organelles. The glycans formed in ER are usually simple and homogeneous. The diversity of protein glycosylation mainly results from Golgi glycosylation machinery\textsuperscript{11}.

An over-simplified description of glycan synthesis is provided in this section\textsuperscript{11}. Glycan synthesis starts on the cytoplasmic side of the ER membrane (Figure 1-4). The lipid-linked precursor dolichol phosphate (Dol-P) is first glycosylated by glycosyltransferases to generate Dol-P-P-Man\textsubscript{5}GlcNAc\textsubscript{2} (step 1), which is then flipped into the ER lumen(step 2), where 7 more sugars are added to form Dol-P-P-Glucose\textsubscript{3}Man\textsubscript{9}GlcNAc\textsubscript{2}. The Glucose\textsubscript{3}Man\textsubscript{9}GlcNAc\textsubscript{2} is transferred by oligosaccharyltransferase (not shown in the Figure) to the newly synthesized linear peptide (step 3). After deglucosylation - reglucosylation cycles, more monosaccharides are trimmed away, and the polypeptide is folded into correct 3D structure with a glycan composition of Man\textsubscript{8}GlcNAc\textsubscript{2} (step 4). The folded protein is then transferred to the Golgi, where a diverse array of glycosyltransferases and glycosidases resides, for further trimming (step 5) and terminal decoration (such as addition of fucose and sialic acids, step 6) to produce a matured glycoprotein.
Figure 1-4. Biosynthesis of the N-linked glycan. Endoplasmic reticulum and Golgi are involved in glycan synthesis, with the ER generating simple and homogenous glycan and Golgi producing much more complex oligosaccharides. Detailed description can be found in Section 1.4.2. This figure is adopted from Varki11.
1.4.3 Glycosylation changes in cancer

As described in previous section, a precise collaboration of enzymes (glycosyltransferase and glycosidase) is needed to decorate the matured glycoprotein with “right” glycan structures. Cancer has profound influence on gene transcription and could affect protein glycosylation in many different ways\textsuperscript{28}. Seven possible mechanisms are schematically shown in Figure 1-5. Among them, mechanisms 3 and 5 - enzyme function control by phosphorylation and enzyme catalytic domain release into circulation - are not yet observed experimentally\textsuperscript{28}. The perturbation of the glycosylation machinery by cancer could generate glycoproteins with different (aberrant) glycan structures. Aberrantly glycosylated glycoproteins may enter blood and could therefore serve as cancer biomarkers for diagnostic purpose. There are multiple glycosylation sites per protein, as a result, the information conveyed by glycosylation is amplified compared to peptide alone. Also, the same glycoprotein produced from different tissues will have different glycoform distribution profile\textsuperscript{11}, therefore protein glycosylation may also convey tissue-specificity information. As such, protein glycosylation may potentially be a category of biomarker, with better sensitivity and specificity than protein alone.

It should be noted that many established cancer biomarkers are glycosylated proteins. Examples include carcinoembryonic antigen for colorectal cancer\textsuperscript{29}, alpha-fetoprotein (AFP) for liver cancer\textsuperscript{30}, cancer antigen 125 for ovarian cancer\textsuperscript{31} and prostate-specific antigen (PSA) for prostate cancer\textsuperscript{32}, etc. Fucosylated AFP has been reported to be able to differentiate chronic liver disease from liver cancer\textsuperscript{30}. Similarly, a special glycoform of PSA, when probed with maackia amurensis agglutinin lectin, could separate prostate cancer patients from benign patients\textsuperscript{32}. These examples support the importance of developing glycosylation characterization techniques for discovering glyco-biomarker.
Figure 1-5. Seven proposed mechanisms for the controlling of protein glycosylation. The 1\textsuperscript{st} one is the gene transcription of the enzymes (glycosyltransferases and glycosidases) responsible for glycans synthesis; The 2\textsuperscript{nd} one is the synthesis and transport of the monosaccharide donors (nucleotide sugars) to either ER or Golgi; The 3\textsuperscript{rd} one is the function switching of the glycosyltransferases and glycosidases by phosphorylation; The 4\textsuperscript{th} one is the competition for the same glycoprotein substrate between different glycosyltransferases and glycosidases; The 5\textsuperscript{th} one is the trafficking of the enzymes into different compartments of either ER or Golgi; The 6\textsuperscript{th} one is the release of the catalytic domains of the enzymes into the cytoplasmic area and further into circulation; The 7\textsuperscript{th} and the last one is the turnover of the generated glycoproteins. This figure is adopted from Ohtsubo\textsuperscript{28}. 
1.5 Techniques for protein glycosylation characterization

Protein glycosylation characterization\(^{33}\) includes the following tasks: glycan composition, glycan sequence/branching, linkage, site-specificity and quantity. Due to the complex nature of protein glycosylation, until now, no single technique could accomplish all of above mentioned tasks. Multiple analytical techniques are usually combined to achieve comprehensive characterization of protein glycosylation\(^{33}\). Intact glycoprotein, due to its relatively large molecular weight, various glycans per site (micro-heterogeneity) and many glycosylation sites per protein, can seldom be studied structurally in detail. Glycoprotein(s) are usually cleaved into smaller fragments – either glycans or glycopeptides using glycosidase or protease - for further structural determination. The released glycan or glycopeptides can be characterized by various types of separation methods such as HPLC and CE or by mass spectrometry (derivatization is often necessary for glycan analysis). Due to the sample complexity, hyphenated techniques such as LC-MS are essential for glycosylation characterization\(^{26,34}\). In the following, a detailed discussion of each individual technique is given and categorized by targets such as intact glycoprotein, released glycan, and glycopeptide digests.

1.5.1 Intact glycoprotein

Intact glycoprotein analysis has been typically performed using methods such as lectin-based technique, capillary electrophoresis and mass spectrometry, either separate or in combination. Usually, the purpose of intact glycoprotein analysis is to determine the general category of presented glycans (lectin) or glycoform distribution (CE, MS). Due to the limitation of current techniques, the information related to detailed glycan structure and site-specificity is not easily obtained.
1.5.1.1 Lectin based technique

Lectins are proteins that specifically bind glycans according to their structural epitopes. Due to its affinity and specificity, lectin is useful reagent for separating and/or detecting glycoproteins with different oligosaccharide moieties\textsuperscript{12}. Lectin affinity chromatography, for example, has been used to separate glycoproteins from complex sample matrices. Multiple-lectin affinity chromatography has been developed to capture a broad range of glycoproteins from plasma/serum\textsuperscript{35}. Another technique, antibody-lectin sandwich assay (ALSA)\textsuperscript{36, 37}, uses specific lectin for detection purposes (Figure 1-6-A). In ALSA, an antibody is immobilized onto solid phase to capture target proteins from a complex mixture. The captured protein is probed with labeled lectin for detection. By comparing disease and normal samples, the differentially expressed glycoform of the target protein can be identified. The disadvantage for lectin based technique is the lack of information on the exact nature of the glycan structures being measured, including the sequences of the glycans and their attachment sites.

1.5.1.2 Capillary electrophoresis

Capillary electrophoresis (CE) features high resolving power and fast separation speed. Capillary electrophoresis has been used for the assessment of glycoprotein heterogeneities, especially for sialic acid containing molecules. One example is shown in Figure 1-6-B, where transferrin, a marker for congenital disorder of glycosylation (CDG), was subject to CE profiling\textsuperscript{38}. Four different sialoforms of transferrin could be separated within 8 mins, with undersialylated transferrins (SA\textsubscript{0-2} in Figure 1-6-B) clearly separated from normally sialylated molecules (SA\textsubscript{4}).
1.5.1.3 MS analysis

Intact glycoproteins can be resolved to individual glycoforms by MALDI-TOF-MS and/or ESI-MS. One example is shown in Figure 1-6-C\textsuperscript{39}. Ovalbumin was directly infused into the mass spectrometer and \(\sim\) 20 scans were averaged for this spectrum. By comparing with theoretical masses, 11 glycoforms could be identified for ovalbumin. In this example, a low resolution triple quadrupole MS instrument was used. Recently, high resolution, high mass accuracy instruments such as qTOF or FT-ICR-MS have been increasingly adopted for this purpose. Also, CE has been interfaced to mass spectrometry to achieve improved separation of protein glycoforms. An example of CE-FT-ICR-MS analysis of intact recombinant therapeutic glycoprotein is shown in Figure 1-6-D (heat map)\textsuperscript{40}. Many protein glycoforms could be observed and the 20 most intense ones were labeled with their corresponding masses, and the differences in glycan composition are highlighted using color coded arrows.
Figure 1-6. Intact glycoprotein analysis. (A) Antibody-lectin sandwich assay\textsuperscript{36}; (B) CE analysis of intact transferrin\textsuperscript{38}. SA: sialic acid; (C) Intact chicken ovalbumin glycoform analysis using a low resolution triplequadrupole instrument\textsuperscript{39}; (D) CE-LTQ-FT-ICR-MS analysis of a recombinant human chorionic gonadotrophin\textsuperscript{40}. The X-axis represents migration time, while Y-axis represents m/z and the shade of the spots indicates signal intensity.
1.5.2 Glycan analysis

Glycan analysis has long been the method of choice for protein glycosylation characterization. Glycan analysis is technically easy and straightforward and there are various developed approaches available for glycan analysis. The disadvantage of glycan analysis is the site-specificity information is lost, unless the glycan is released from individually purified glycopeptide. Also, multiple-step sample preparation including glycan release, derivatization and clean up is necessary, and sample loss is inevitable. These make it challenging when the sample amount is limited, such as those derived from precious biological material.

The released glycans are hydrophilic molecules with no chromophores for UV detection. Thus, glycans are usually derivatized to increase the interaction between the glycans and the respective chromatography stationary phase and/or assist detection. Chromatographic and electrophoretic separations are widely used to “profile” glycans, as detailed in sections 1.4.2.3 to 1.4.2.8. By comparing with the known standards, the structure of the to-be-characterized glycans can be determined. Exoglycosidases have been widely used, usually multiple enzymes in a sequential format, for glycan structure determination. By detecting the shifts of glycan elution or migration (e.g. LC or CE) or m/z difference (e.g. MS measurement) after serial exoglycosidase treatments, the glycan composition and their structure could be determined. Mass spectrometry is becoming increasingly important for glycan analysis. Glycan composition can be directly determined using high resolution MS measurement such as ToF or FT-ICR-MS. Glycan sequence, branching and/or linkage information can be obtained using CID-MS/MS fragmentation. In the following, glycan release is first summarized, followed by the commonly used chemical derivatization. The analytical techniques used for glycan analysis is given at last.
1.5.2.1 N-Glycan release from protein backbone

Chemical methods such as alkaline borohydride treatment (\(\beta\)-elimination)\(^{43}\) and hydrazinolysis\(^{44}\) have been used to release N-linked oligosaccharides. Although these methods can be effective, the harsh reaction conditions are often accompanied by side-reactions; moreover, the protein may not be fully recovered. Enzyme is an ideal alternative for glycan release. N-linked glycans can be released enzymatically from glycoproteins using glycoamidases, endoglycosidases and exoglycosidases (Table 1-2)\(^{45}\). PNGase F, the most widely used enzyme for release of N-linked glycans, is glycoamidase. PNGase F cleaves intact N-linked glycans from peptides and converts previously occupied asparagine to aspartic acid. PNGase F has a wide specificity, cleaving all N-linked glycans except those having \(\alpha\) 1-3 linked fucose to the inner most GlcNAc of the N-linked core structure. This type of glycan usually exists in fungi and could be cleaved using PNGase A\(^{45}\). Endoglycosidases is usually more specific than glycoamidase, such as endoglycosidase H (Endo H), which cleaves high mannose and hybrid glycan structures. Endo H cleaves between the two GlcNAc residues of the chitobiose core, leaving one GlcNAc still attached to the protein\(^{45}\). Also, there are many exoglycosidases available with different substrate specificity (Table 1-2). Exoglycosidases are specific for both monosaccharide residue and linkage type, and therefore are invaluable reagents for detailed structural elucidations of the glycans\(^{11}\).
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>EC number</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Glycanase</td>
<td>3.5.1.52</td>
<td>cleaves between Asn of oligomannose, complex, or hybrid N-glycans; requires at least one amino acid at both the amino terminal and carboxyl terminal of Asn</td>
</tr>
<tr>
<td>(Peptide-N-glycosidase A, PNGase A, glycoamidase A)</td>
<td>3.5.1.52</td>
<td>cleaves between Asn of oligomannose, complex or hybrid N-glycans; will work if core is $\alpha$ 1-3 fucosylated, but ineffective when sialic acid is present; requires at least one amino acid at both the amino terminal and carboxyl terminal of Asn;</td>
</tr>
<tr>
<td>Endo-</td>
<td>3.2.1.96</td>
<td>cleaves between the two N-acetylglucosamine residues in the core of high mannose N-glycans</td>
</tr>
<tr>
<td>Endoglycosidase F</td>
<td>3.2.1.96</td>
<td>cleaves between the two N-acetylglucosamine residues in the core of high mannose or biantennary complex N-glycans</td>
</tr>
<tr>
<td>Sialidase (neuraminidase)</td>
<td>3.2.1.18</td>
<td>nonreducing terminal sialic acids</td>
</tr>
<tr>
<td>Fucosidase</td>
<td>3.2.1.51</td>
<td>nonreducing terminal fucose</td>
</tr>
<tr>
<td>$\beta$-galactosidase</td>
<td>3.2.1.23</td>
<td>nonreducing terminal $\beta$-galactose</td>
</tr>
<tr>
<td>$\alpha$-mannosidase</td>
<td>3.2.1.24</td>
<td>nonreducing terminal $\alpha$-mannose</td>
</tr>
<tr>
<td>$\beta$-N-Acetylhexosaminidase</td>
<td>3.2.1.52</td>
<td>nonreducing terminal HexNAc</td>
</tr>
</tbody>
</table>

*The figure is adopted from Varki\textsuperscript{11}.*
1.5.2.2 Glycan derivatization

Once the glycan are released, they can be analyzed by various techniques. To improve its detection and/or separation, glycan usually needs to be derivatized before separation or MS analysis. A summary of commonly used derivatization reactions is shown in Figure 1-7. N-linked glycans released by enzymes have reducing ends, which in solution exist as equilibrium between cyclic hemiacetals and open-ring aldehydes (Figure 1-7-A). Glycans are typically reduced to alditols to prevent peak splitting of α- and β- anomers (Figures 1-7-A and Figure 1-7-D), which otherwise can be separated by high resolution separation techniques such as LC or CE. Permethylation, the conversion of all hydrogen presented in -OH or -NH group to methyl group, is important for MS measurement(Figure 1-7-A). Permethylation increase the stability and hydrophobicity of the derivatized glycans, which further facilitates MS analysis16. Chromophore, fluorophore, or stable isotope tag can be easily attached to a glycan using reductive amination(Figure 1-7-A). Three commonly used fluorescence tags are shown in Figure 1-7-B. Reductive amination requires that the glycan have a reducing end; for glycans that do not have a reducing end, such as those attached to a peptide moiety, the cis-diol on the glycan can be oxidized to generate aldehyde groups before further derivatization (Figure 1-7-C). Although effective and widely adopted, chemical derivation causes significant sample losses during the derivatization workup, which makes glycan characterization not suitable for analysis of precious biological/clinical samples.
Figure 1-7. Common derivatization for glycan. (A) Released glycans exists as equilibrium between chemically non-reactive hemiacetal and chemically reactive aldehyde forms. Reducing agent such as sodium borohydride can reduce aldehyde to –OH group. Permethylation will convert all the N- or O-linked hydrogen into methyl group. The aldehyde group could also be conjugated to primary amine using so-called Schiff base reaction (reductive amination). Reductive amination is usually used to attach a fluorophore to the glycans. R, glycan moiety, R’, primary amine alkyl group. This figure is modified from Zaia\textsuperscript{34}. (B) Commonly used fluorophore. APTS is triply negatively charged at pH 7, which makes it especially suitable for CE analysis. (C) A glycan attached to protein/peptide can be oxidized to generate an aldehyde group for further reductive amination. (D) Glycan mutarotation: interconversion of α and β anomers, which differ only in the stereochemistry around the hemiacetal carbon.

1.5.2.3 High pH anion exchange chromatography - pulsed amperometric detection
High pH anion-exchange chromatography (HPAEC) - pulsed amperometric detection (PAD)\(^{46}\) is an established method for glycan analysis. The separation in HPAEC is carried out at highly alkaline mobile phase (pH >12), in which glycans will carry negative charge. The separation is based on ionic interactions between the negatively charged carbohydrates and the anion exchange resins. The advantage of HPAEC is the high resolution and labile free quantitation. The disadvantage is, at high pH, side reactions of the reducing end of GlcNAc can occur\(^{33}\). Also, the high salt concentration of HPAEC mobile phase prevents its interfacing to MS.

1.5.2.4 Capillary electrophoresis - laser induced fluorescence

Glycans are derivatized with a fluorescence tag for capillary electrophoresis - laser induced fluorescence (CE-LIF) analysis and separated on the basis of charge and size\(^{47}\). CE has sufficient resolution to separate linkage and/or position isomers, see Figure 1-8-A. Glycans released from mAb were labeled with aminopyrene trisulfonic acid (APTS) and subjected to standard CE-LIF analysis (polyvinyl alcohol (PVA) coated capillary to minimize EOF)\(^{47}\). G1 and G1’, which differ only at the position of one galactose on the biantennary structure, were successfully separated by CE. Four early migrating species (labeled 1-4, Figure 1-8-A), were later characterized by CE-LIF-MS analysis, and determined to be sialylated species. Sialylated species migrate faster because of additional negative charges.

1.5.2.5 Hydrophobic interaction chromatography

Hydrophilic interaction chromatography (HILIC)\(^{48}\), separating glycans based on size, hydrophilicity, and/or charge (zwitterion - HILIC), is also an effective chromatography mode. HILIC can be considered as a subtype of normal phase LC (NPLC), where the stronger elution
solvent is water in contrast to reversed phase where it is the weakest solvent. An example of HILIC separation is shown in Figure 1-8-B\textsuperscript{48}. In this case, glycan released from a standard glycoprotein was subjected to amide-HILIC-ESI-MS analysis. The peaks labeled with capital letters were on-line determined by MS. A disadvantage of HILIC is that the resolution of structural isomers is generally not observed (compare the peak shape and peak width in Figure 1-8-A and 1-8-B).

1.5.2.6 Reverse phase liquid chromatography

For reverse phase - LC (RPLC) analysis, reductive amination is utilized to increase the hydrophobicity of glycans. Aminopyridine, 2-aminobenzamide (2-AB), 2-aminobenzoic acid (2-AA) are widely used\textsuperscript{33,41}. One example is shown in Figure 1-8-C, where glycans were released from a therapeutic glycoprotein, 2-AB reductively aminated and RPLC separated\textsuperscript{49}. Three peaks can be observed, corresponding to neutral (B), mono- (BS), and disialylated (BS\textsubscript{2}) biantennary oligosaccharides. Ion pairing reagents such as dibutyl amine can be added to mobile phase to separate native acidic glycans (sialic acid containing oligosaccharides) using RPLC\textsuperscript{50}. The addition of positively charged ion-pairing agents will shield the negatively charged glycan and therefore increase the interaction between glycan and stationary phase. However, the ion-pairing agent (high concentration of salt) will make it challenging for the online coupling between RPLC and MS.

1.5.2.7 Porous graphitized carbon

Porous graphitized carbon (PGC) is much more hydrophobic than C\textsubscript{18} stationary phase and can be used directly for carbohydrate analysis\textsuperscript{51}. The separation of glycans is mainly based
on hydrophobic and/or dipole-dipole interaction. Porous graphitized carbon chromatography can be used to separate structural isomers such as anomers. To prevent peak splitting due to anomer interconversion (Figure 1-7-D), the glycans are usually reduced to alditol or reductively aminated. The disadvantage of PGC is the potential for irreversible adsorption of contaminants from the sample matrix as well as low recovery from multi-sialylated compounds.

1.5.2.8 MS based technique

Similar to chromatography / electrophoresis, MS has also been used for glycan profiling (Figure 1-8-D)\(^49\). It will be interesting to compare Figure 1-8-C and Figure 1-8-D, since they were analyzing glycans released from the sample glycoprotein\(^49\). Three more glycans could be easily identified from MS profiling, where RPLC fails to do so, mainly due to the glycan co-elution. For high resolution, high mass accuracy detector such as TOF, FT-ICR or Orbitrap-MS, the direct molecular weight measurement usually will be sufficient to determine the monosaccharide composition for individual glycan\(^34\). Glycan mixtures can be analyzed directly using MS\(^2\) or MS\(^n\) experiment. Glycan sequence, antenna branching, and even linkage information (to variable extent) can be determined through glycan fragmentation\(^34\).

Glycans are usually permethylated before MS measurement\(^34\). As already mentioned above (section 1.4.2.2), permethylation improves the glycan hydrophobicity by converting all the hydrogen atoms (-OH or –NH linked) into methyl groups. For ESI-MS analysis, this conversion facilitates glycan ionization and stabilizes labile glycans, especially sialic acids containing glycans. It has also been found that permethylated glycans produce information-rich fragments in MS/MS analysis than native glycans\(^52\). MS/MS fragmentation of glycans usually produces two types of fragments: glycosidic cleavages where the fragmentation occurs between
monosaccharides and cross-ring cleavages that involve the fragmentation of two bonds on the same monosaccharide residue. Domon and Costello (Figure 1-8-E) proposed a systematic nomenclature for describing glycan fragments generated by CID-MS/MS analysis. In this system, A, B, and C type ions refer to ions retaining charge on the non-reducing terminus, whereas X, Y, and Z are the ions having charge on the reducing terminus (Figure 1-8-E). B, C, Y and Z type ions are resulted from glycosidic cleavages, and are useful for glycan sequence and branching determination. A and X type ions are cross-ring fragments, usually have weaker intensities than that of inter-ring fragments, and can be used to obtain linkage information.
Figure 1-8. Glycan analysis using CE, LC and MS. (A) Standard CE-LIF analysis of APTS-labeled, N-linked glycans released from mAb\(^47\). The identifications (G0, G1, G1’, G1-1, G1’-1 and G2) are based on enzymatic treatment with exoglycosidases and/or co-migration with commercially available standards. (B) HILIC-MS analysis of underivatized glycan\(^48\). (C) RPLC chromatogram of the 2-aminobenzamide-derivatized N-glycans from CD4 glycoprotein\(^49\). (D) Electrospray mass spectrum of the permethylated oligosaccharides from CD4 glycoprotein\(^49\). (E) Nomenclature for glycan MS/MS fragments\(^53\).
1.5.3 Glycopeptide analysis

Glycopeptides are arguably the targets of choice for site-specific characterization of protein glycosylation, since the identified glycan can be directly correlated with the attachment site and peptide\(^\text{33}\). Traditional site-specific glycosylation characterization involves glycopeptide separation, fraction collection, glycan release and MS analysis on glycans and deglycosylated peptides. The identified peptide sequence can be correlated with glycan structures to obtain site-specific glycosylation information\(^\text{33}\). The rapid development of mass spectrometry, including hardware design, novel fragmentation and further data treatment, is changing the field of direct analysis of glycopeptides.

In this thesis, the works presented on glycosylation characterization are performed on glycopeptide level. In the following, a brief summary of glycopeptide generation using protease digestion is first presented, followed by the comparison of various MS/MS techniques, with the focus on the type of information could be obtained with respect to glycan or peptide moiety. Last, the challenges of current LC-MS strategy for glycopeptides analysis are discussed.

1.5.3.1 Glycoprotein protease digestion

Glycosylated proteins can be digested into glycopeptides using either specific enzymes such as trypsin, Glu-C, Asp-N, Lys-C or non-specific enzymes such as pronase\(^\text{54}\). Trypsin is the most widely used enzyme in proteomics studies. Trypsin specifically cuts protein sequences at the basic K/R residues; the basic amino acids facilitate peptide ionization during positive ESI. However, other enzymes can also be used including Glu-C, Asp-N or Lys-C. Pronase\(^\text{54,55}\) is a mixture of endo- and exo-peptidases. Pronase usually cuts the unmodified protein into individual amino acids. For glycosylated proteins, due to the steric hindrance of glycan, the amino acids
around the glycosylation site will be cut at much slower rate; therefore, by controlling the concentration of the enzyme and digestion time, glycans attached to short peptides (3-6 AAs long) fragments could be obtained. One disadvantage of pronase digestion is that the same glycan will be attached to multiple short peptides with different AA lengths$^{54,55}$. This variability will decrease the signal intensity and further complicate the MS data interpretation and quantitation (same glycan, different m/z, and different elution time). Also, much more stringent protocol, compared to trypsin digestion, is needed to achieve reproducible results.

1.5.3.2 MS/MS fragmentation techniques

MS/MS fragmentations have been used for glycopeptide analysis, as summarized in Table 1-3. The information on glycopeptide structure that can be obtained using various MS/MS techniques is different, as schematically represented in Figure 1-9.

Collision induced dissociation (CID) is the most widely used MS/MS technique in proteomics study and can be performed in wide range of instruments including ion trap (IT)$^7$ and quadrupole instrument. Infrared multiphoton dissociation (IRMPD) is a MS/MS technique specially developed in fourier transform - ion cyclotron resonance (FT-ICR) instrument$^{56,57}$. Both CID and IRMPD$^{56}$ are vibrational fragmentation techniques, which favor dissociation of the weakest bond in the precursor molecule. In CID, the precursor is accelerated by electric potential and collides repeatedly with neutral gas molecules$^7$. Fragment ions are formed when the internal energy of the ion exceeds the strength of the weakest bond in the molecule. In IRMPD, the precursor internal energy is increased by the absorption of multiple infrared photons, which cause energetic vibration throughout the molecule and bond fragmentation$^{56}$. For glycopeptides, the weakest bonds in the molecule are usually glycosidic linkages. Dissociation of glycopeptides
using CID or IRMPD therefore generally produces abundant ions from cleavage of glycosidic bonds, with little or no peptide backbone dissociation.

On the contrary, electron transfer dissociation (ETD)\textsuperscript{58} and electron capture dissociation (ECD)\textsuperscript{59} are radical induced fragmentation methods, which will preserve the labile post translational modifications (PTMs) such as glycosylation during fragmentation. In ECD\textsuperscript{59}, which is performed only in FT-ICR, multiply protonated peptides are “mixed” with highly excessive number of thermal electrons (i.e. electrons generated by heated filament). When the positive charged peptide captures one electron (negative charged), the peptide backbone will fragment by a so-called “nonergodic process”: a process so fast that no intramolecular vibrational energy redistribution is involved. Similar to ECD, the ETD fragmentation, which is performed in ion trap, is also initiated by the capture of one electron, which is transferred from electron-carrying molecule fluoranthene\textsuperscript{58}. Dissociation of glycopeptides using ETD and ECD typically generates peptide backbone fragments with the intact glycan still attached to the peptide moiety, allowing the site of glycosylation to be determined.

High energy C-trap dissociation (HCD)\textsuperscript{60}, a recently introduced MS/MS fragmentation approach in the Orbitrap, actually occurs in an octopole located behind the C-trap (a curved linear ion trap), which works as an ion pusher to inject ions into Orbitrap for high resolution mass measurement. The name of HCD comes from its original instrument design\textsuperscript{60}, where the fragmentation occurs at C-trap in elevated energy. The glycopeptide fragmentation using HCD is dominated by glycosidic linkage breakage, since it essentially is CID-type fragmentation.

MALDI-TOF/TOF is another type of MS/MS fragmentation, which features much higher precursor energy (> keV)\textsuperscript{34}. As a comparison, for all other techniques, the kinetic energy of the precursor is usually less than ~50 eV. The commonly observed glycopeptide fragments using
MALDI-TOF/TOF includes both glycan linkage and peptide sequence breakage\textsuperscript{26}, which provide information on both peptide and glycan moiety of the glycopeptide.

For glycopeptide analysis, a combination of CID/ETD or IRMPD/ECD-MS/MS can be a preferred choice, since both glycan structure and peptide sequence could be characterized using the same MS instrument (IT-CID/ETD or FT-ICR-IRMPD/ECD). Currently, the application of FTICR-IRMPD/ECD-MS/MS is mainly confined to academic field\textsuperscript{56}, while IT-CID/ETD-MS/MS has more utilization in both academic research and industry application. We choose linear ion trap (LTQ)-CID/ETD-MS/MS to develop our strategy for sensitive site-specific characterization of protein glycosylation, as described in detail in chapter 3.
<table>
<thead>
<tr>
<th>Fragmentation</th>
<th>Description</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CID</strong>&lt;sup&gt;7&lt;/sup&gt;</td>
<td>Precursor ion is accelerated by electrical potential (energy level &lt; 20 eV); collision with neutral gas molecules (helium, nitrogen or argon, depending on instrument) to generate fragments</td>
<td>IT, quadrupole&lt;sup&gt;#&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>IRMPD</strong>&lt;sup&gt;56, 57&lt;/sup&gt;</td>
<td>Precursor ion absorb infrared light, which cause vibrational excitation, collision with neutral gas molecule to generate fragments</td>
<td>FT-ICR</td>
</tr>
<tr>
<td><strong>ETD</strong>&lt;sup&gt;58&lt;/sup&gt;</td>
<td>ETD occurs in ion trap instrument. An electron is transferred from singly charged anion (usually fluoranthene) to multiply charge precursor ion. An activated electron species (4-5.5 eV) is generated which undergoes dissociation and charge reduction</td>
<td>IT</td>
</tr>
<tr>
<td><strong>ECD</strong>&lt;sup&gt;59&lt;/sup&gt;</td>
<td>Irradiation of multiply charged precursor ions with low energy electrons (~ 6 eV), generating fragments and charge reduced radical species</td>
<td>FT-ICR</td>
</tr>
<tr>
<td><strong>HCD</strong>&lt;sup&gt;60&lt;/sup&gt;</td>
<td>Similar to quadrupole CID</td>
<td>Orbitrap&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>TOF/TOF</strong>&lt;sup&gt;26&lt;/sup&gt;</td>
<td>High energy CID; Precursor is accelerated by high electric potential to achieve &gt; keV kinetic energy.</td>
<td>MALDI-TOF/TOF</td>
</tr>
</tbody>
</table>

<sup>#</sup> IT has the ability for MS<sup>n</sup>, while quadrupole only good for MS<sup>2</sup>. <sup>*</sup>HCD actually occurs in an octopole located behind C-trap, the fragments generated in the octopole were transferred back to C-trap and injected into Orbitrap for high mass measurement.
Figure 1-9. Glycopeptide analysis using different MS/MS technique. The length of the arrow approximately indicates the information richness could be obtained using the specified MS/MS technique. This figure is inspired by Dr. Manfred Wuhrer’s talk.
1.5.3.3 LC-MS analysis of glycopeptides

Due to microheterogeneity (many glycoforms on single site), a specific glycopeptide usually constitutes only a small proportion of the peptides generated by enzymatic digestion of glycoprotein/protein mixture. Compared to peptides, glycopeptides has lower ionization efficiency (due to the attached hydrophilic glycan), and its ionization will be suppressed by non-glycosylated peptides. For LC-MS analysis, liquid chromatography can separate glycopeptides from interfering peptides (to variable extent), therefore decreasing ion suppression. A recent multi-institution study indicates that LC-MS technique is sufficiently mature and can provide results comparable to standard method such as glycan profiling using HILIC-LC. Furthermore, LC-MS analysis of glycopeptides features easy sample preparation, and can provide direct information on both glycan composition (MS level) and structure (MS/MS level).

For current LC-MS, the typical sample consumption level ranges from 0.5-5 pmol/injection for 75 μm i.d. column to 50 pmol/injection for 300 μm i.d. column and ~200 pmol/injection for 1mm i.d. column. Due to the large dynamic range, complex glycan structure and requirement for attachment site determination, multiple LC-MS/MS runs are necessary, which demands much larger amount of sample consumption (10s to 100s pmol). More sensitive LC-MS techniques, which could handle trace amounts of sample (total consumption level at 100s fmol), can be important for glycoproteomics and glyco-biomarker discovery.

Recently, our group developed a 10 μm i.d. porous layer open tubular (PLOT) column for ultrasensitive proteomics analysis (Figure 1-10-A). Operating at 20nL/min, the PLOT column generates ultrasmall ESI droplets, resulting in improved ionization and decreased ion suppression. The ESI droplet diameter generated at 20 nL/min is ~ 70 nm, calculated according the equation described by Wilm and Mann. As a comparison, the diameter of droplet formed at
200 nL/min is ~ 470 nm. Additionally, the electrospray emitter can be placed nearer to the MS inlet than in comparable configurations, which improves the sampling efficiency at low flow rates (Figure 1-10-B)\textsuperscript{39}. We postulate PLOT column, when coupled to LTQ-CID/ETD-MS/MS, could provide sensitive glycopeptide analysis and should be able to contribute to the emerging glycoproteomics field. Currently, the practice in glycoproteomics filed can be divided into 3 sub-fields\textsuperscript{54}, namely proteocentric, glycocentric and monoproteic, which are described below.
Figure 1-10. PLOT LC-ESI-MS analysis at ultralow flow rate (20 nL/min). (A) scan electron micrograph of PLOT crosssection. (B) Schematic representation of the configuration of conventional electrospray and nanoelectrospray LC-ESI-MS. This figure is adopted from Smith69.
1.6 Glycoproteomics studies

Glycoproteomics can be viewed as a branch of proteomics that identifies, catalogs, and characterizes proteins containing carbohydrates as a post-translational modification. Based on the difference of research focus, there are 3 major type glycoproteomics studies, namely proteocentric, glycocentric and monoproteic\textsuperscript{54}.

1.6.1 Proteocentric glycoproteomics

Proteocentric glycoproteomics is the large-scale identification of glycosylated proteins and mapping of glycosylation sites; all information regarding the glycans is lost. Proteocentric glycoproteomics generally involves the enrichment of glycan-bearing peptides/proteins using hydrazide coupling\textsuperscript{70} or lectin\textsuperscript{71}, followed by standard proteomic analysis. Two typical proteocentric glycoproteomics are the hydrazide capture method\textsuperscript{70} and filter assisted sample preparation (FASP) plus lectin method\textsuperscript{71}. PNGase F treatment converts glycosylated asparagine into aspartic acid. Standard LC-MS/MS experiments will be able to identify previously glycosylated glycosylation sites, based on the observation of newly generated sequence motif of Asp-X-Ser/Thr. The glycan information is neglected in this type of experiment. As stated in Section 3, the loss of this information limits the biological conclusions that may be drawn from such experiments.

1.6.2 Glycocentric glycoproteomics

In contrast to proteocentric, in glycocentric study, glycoprotein identities and sites information are usually not characterized. The focus of glycocentric glycoproteomics is the composition, branching, structure and relative quantitation of oligosaccharides released from
proteins\textsuperscript{72, 73}. The common work flow involves glycoprotein enrichment, glycan release, derivatization, and profiling by chromatography, electrophoresis and/or mass spectrometry methods. The individual glycan structure can be deduced from the glucose unit\textsuperscript{42}, by comparison with standards, exoglycosidase treatment and/or MS/MS analysis. Again, all information related to specific sites/proteins is lost. Glycocentric glycoproteomics has been used in comparative studies, with the goal of discovery of potential glycan-markers for early diagnosis of cancer (Figure 1-11-A)\textsuperscript{74}.

\textbf{1.6.3 Monoproteic glycoproteomics}

The third major category of glycoproteomics is monoproteic\textsuperscript{54}. Compared to proteocentric and glycocentric, monoproteic features detailed and site-specific characterization of protein glycosylation status. The work presented in this thesis belongs to monoproteic glycoproteomics (Figure 1-11-B). The workflow of monoproteic glycoproteomics is similar to that of standard shotgun proteomics, and generally involves protein purification, protein digestion and LC-MS/MS analysis. Pure or simple protein mixture is the preferred target; since simple sample matrix will decrease the glycopeptide suppression and also facilitate data-interpretation.

For monoproteic glycoproteomics, the target glycoprotein can be deduced from multiple sources such as fundamental biological research, biomarker studies and/or existing biomarkers (Figure 1-11-B). Site-specific glycoforms of the target glycoprotein can be identified using sensitive LC-MS analysis, such as PLOT LC-MS described in Chapter 3. Site-specific differential quantification of protein glycoforms using newly developed glycopeptide SRM
(Chapter 4) can identify cancer-specific glycoforms, which could become potential biomarkers, possibly more specific than either the protein or glycan alone.
Figure 1-11. Representative glycoproteomics studies. (A) Glycocentric glycoproteomics. This figure is adopted from Bosques. (B) monoproteic glycoproteomics.
1.7 References


44. Patel, T.; Bruce, J.; Merry, A.; Bigge, C.; Wormald, M.; Jaques, A.; Parekh, R., Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. *Biochemistry* 1993, 32, (2), 679-93.


Chapter 2

Antigen Identification and Characterization of Lung Cancer Specific Monoclonal Antibodies Produced by mAb Proteomics*

* The main part of this chapter has been published in *Journal of Proteome Research* 2010, 9(4), 1834-1842.
2.1 Abstract

A mass spectrometric (MS)-based strategy for antigen (Ag) identification and characterization of globally produced monoclonal antibodies (mAbs) is described. Mice were immunized with a mixture of native glycoproteins, isolated from the pooled plasma of patients with non-small cell lung cancer (NSCLC), to generate a library of mAb-secreting hybridomas. Prior to immunization, the pooled NSCLC plasma was subjected to 3-sequential steps of affinity fractionation, including high abundant plasma protein depletion, glycoprotein enrichment and polyclonal antibody affinity chromatography normalization. In this work, in order to demonstrate the high quality of the globally produced mAbs, we selected 3 mAbs of high differentiating power against a matched, pooled normal plasma sample. After production of large quantities of the mAbs from ascites fluids, Ag identification was achieved by immunoaffinity purification, SDS-PAGE, Western blotting and MS analysis of in-gel digest products. One antigen was found to be complement factor H, and the other two were mapped to different subunits of haptoglobin (Hpt). The 2 Hpt mAbs were characterized in detail in order to assess the quality of the mAbs produced by the global strategy. The affinity of one of the mAbs to the Hpt native tetrameric form was found to have a \( K_D \) of roughly \( 10^{-9} \) M and to be 2 orders of magnitude lower than the reduced form, demonstrating the power of the mAb proteomics technology in generating mAbs to the natural form of the proteins in blood. The binding of this mAb to the \( \beta \)-chain of haptoglobin was also dependent on glycosylation on this chain. The characterization of mAbs in this work reveals that the global mAb proteomics process can generate high-quality lung cancer specific mAbs. The mAb proteomics strategy can be applied to other diseases to produce mAbs in a high throughput manner and will contribute to the future biomarker research efforts, especially for biomarker verification/validation.
2.2 Introduction

The search for clinical protein biomarkers in body fluids such as plasma continues to be an area of active research\textsuperscript{1-4}. There have been significant efforts to develop various platforms using liquid chromatography-mass spectrometry (LC-MS) -based shotgun proteomics to discover markers associated with disease. The dynamic range of plasma proteins, the heterogeneity due to post-translational modifications, and the overall sample complexity continue to be major challenges for shotgun proteomics-based platforms. While various strategies have been advanced, there is general agreement that the results to date have not realized the potential of the LC-MS approach, and verification/validation of candidate markers for clinical application remains a major hurdle\textsuperscript{3}. The verification/validation of putative markers requires large (thousands) and controlled population studies, and LC-MS methods up to now have lacked sufficient throughput to make these studies routine; however, progress is being made in the development of technologies for qualification (prior to verification/validation) of biomarkers using selected reaction monitoring (SRM) MS method.

One strategy for biomarker verification/validation is to raise antibodies (Abs) to the discovered candidate markers for immunoassay analysis, e.g. ELISA. The ELISA method can be highly sensitive, can be a high throughput method, and is widely accepted as the gold standard in clinical medicine and biological research. On the other hand, ELISA development is time consuming, and the generation of Abs with high affinity and specificity cannot be certain. Recently, the use of well characterized monospecific polyclonal Abs (msAbs) as capture reagents has been proposed as a technology for clinical proteomics and biomarker validation\textsuperscript{5} and several initiatives are underway to generate Ab libraries\textsuperscript{6,7}. These technologies include the production of scFv antibody subunits using phage display\textsuperscript{6} and the generation of polyclonal Abs
(pAbs) against recombinant peptides. In addition, the use of monoclonal Ab (mAb) libraries generated against mixtures of immunogens from tissue organelles or plasma has been suggested as an alternative ("reverse proteomic") approach for the discovery of biomarkers.

Another global approach, namely mAb proteomics, has recently been introduced using a platform for generating large libraries of mAbs in a high throughput manner. Mice are immunized with complex biological mixtures of proteins, e.g. from enriched plasma. A large hybridoma library is generated, and the mAbs are subjected to high throughput ELISA screening to identify disease specific mAbs. The antigen(s) to which the mAb(s) are generated are initially unknown. It is the purpose of this work to demonstrate a method to identify the antigens associated with specific mAbs. Further, the work examines the quality of the mAbs produced in order to provide insight into the effectiveness of the global mAb library approach.

The technology cited above has been used to discover lung cancer specific mAbs. The immunogen in this study was prepared by 3-sequential steps of affinity fractionation of pooled plasma from patients with non-small cell lung cancer (NSCLC), as described in the Experimental Section. Mice were immunized with the resulting fractionated mixtures to generate a library of mAb-secreting hybridomas. A high throughput ELISA method was developed to screen for lung cancer specific hybridomas, comparing the reactivity of supernatants against biotinylated plasma samples from lung cancer patients to matched normal plasma samples. Three mAbs were selected for antigen identification and characterization because of their high differentiating powers against normal matched samples, and these mAbs were produced in large quantities from ascites fluids.

A mass spectrometric-based method combined with a simple immunoassay was developed for antigen identification. The target of one mAb was found to be to complement
factor H and the other two were mapped to different subunits of the same molecule, haptoglobin (Hpt). The Hpt mAbs were characterized by surface plasmon resonance analysis. The affinity of one of the mAbs was found to be approximately 2 orders magnitude higher for the Hpt native tetrameric form than the reduced form. Unlike other methods based on recombinant proteins or peptide conjugates, this result demonstrates that the mAb proteomics approach to produce large antibody libraries is able to generate mAbs to the actual form of proteins that exists in plasma, i.e. the native form. The complementary pair of mAbs was further used to develop an ELISA test for Hpt in plasma.

2.3 Experimental Section

2.3.1 Materials

Sodium cyanoborohydride, dimethyl pimelimidate•2HCl (DMP), EZ-link™ plus activated peroxidase were purchased from Thermo Fisher Scientific (San Jose, CA). POROS® Affinity Protein G beads and horseradish peroxidase (HRP)-conjugated goat anti-mouse secondary antibody were from Life Technologies (Carlsbad, CA). Sequencing grade trypsin was obtained from Promega (Madison, WI). Chemiluminescence HRP substrate was obtained from GE Healthcare (Buckinghamshire, UK). Colorimetric HRP substrate, tetramethylbenzidine (TMB), was from BioFX (Owings Mills, MD). All other reagents were obtained from Sigma-Aldrich (St. Louis, MO).

2.3.2 Plasma sample collection and immunogen preparation

Plasma from 20 patients diagnosed with non-small cell lung cancer (NSCLC) and 20 healthy control patients, matched by age, gender and race, were obtained from Proteogenex
(Culver City, CA). According to the histological diagnosis, 80% of the cancers were squamous cell carcinoma, 15% were adenocarcinoma, and 5% were undifferentiated NSCLC. All of the patients except one were free of distant metastasis, with primary tumors larger than 2 cm.

Equal volumes of plasma from each patient or matched control were pooled. The pooled plasma was subjected to 3-sequential steps of affinity fractionation. First, the concentration of the twelve most abundant plasma proteins was reduced using the Seppro LC10 IgY depletion immunoaffinity column (SepproLC, Beckman, Fullerton, CA), according to the protocol of the manufacturer. Then, the glycoproteins in the flow-through from the depletion column were enriched using a multi-lectin affinity column (M-LAC)\(^1\)\(^2\). Finally, the eluate from M-LAC column was subjected to a polyclonal affinity column for normalization\(^1\)\(^3\). The normalized M-LAC material of native glycoproteins from the NSCLC plasma was used for immunization to produce the mAb-secreting hybridoma library.

2.3.3 Mouse immunization, hybridoma generation and mAb screening

The generation and screening of the hybridoma library was carried out by our collaborators at Biosystem International (France). Briefly, female mice (8–12 week old) were immunized using complete (first immunization) and incomplete Freund’s adjuvant (subsequent immunizations) or aluminum hydroxide gel. Hybridoma libraries were generated through fusions of SP2/0-Ag14 hybridoma partner cells (ATCC, Teddington, Middlesex, UK) with spleen cells from immunized mouse. Hybridoma supernatants were generated via standard procedures in 96- and 24-well tissue culture plates. Tissue culture wells positive in high throughput screening procedures were selected for cloning. Hybridomas were cloned in selective media (hypoxantin, aminopterin and thymidine, HAT) and adapted to regular media subsequently by gradual
medium exchange. Lung cancer specific mAb-secreting hybridomas were selected by their reaction in a high throughput ELISA screening.

For hybridoma supernatants screening, 96-well plates were coated overnight at 4°C with goat anti-mouse IgG (20 μg/mL in 50 mM sodium carbonate buffer, pH 9.6). The plates were washed with PBS/ 0.05% v/v Tween (PBST), and blocked with 0.5% BSA in PBS for 60 min at room temperature. Following washing with PBST, 25 μL of hybridoma supernatant was added to the well and the plates were incubated for 90 min at room temperature. The plates were washed again with PBST and 25 μL tracers consisting of a predetermined optimal dilution of the biotinylated plasma protein mix, generally 10 μg/mL in PBST / 10 % fetal bovine serum was added to the wells. The tracer was biotinylated depleted plasma protein mix, which was prepared using an EZ-Link™ Sulfo-NHS-LC- biotinylation kit. Following incubation for 60 min at 37 °C, the plates were washed with PBST, and 25 μL streptavidin-HRP conjugate was added to the wells. The plates were then incubated for 60 min at 37 °C. Finally, the plates were washed with PBST, 25 μL ophenylenediamine dihydrochloride (OPD) was added, followed by incubation for 10–30 min. The reaction was stopped by the addition of 25 μL 4M H₂SO₄, and the absorbance was read at 490 nm using a Bio-Tek ELS 800 reader (Bio-Tek Instruments, Winooski, VT, USA).

2.3.4 Immunoaffinity purification

Three mAbs (mAb #1, #2 and #3), which shown significant differential power between lung cancer and normal, were produced in large quantities from ascites fluids and purified by means of HiTrap Protein G (GE Healthcare) and Protein L (Thermo Fisher Scientific) affinity chromatography. The purified mAbs were crosslinked to POROS® Affinity Protein G beads
using DMP and packed into PEEK columns (2 × 20 mm, 63 µL) for HPLC affinity purification. For affinity purification, a Shimadzu HPLC system (LC-10AD, Columbia, MD) was used. Plasma was mixed with an equal volume of binding buffer (100 mM phosphate, 150 mM NaCl, pH 7.5), and 50 µL of sample was injected into the HPLC system. The unbound proteins were washed away with 5 column volumes of binding buffer, and then the retained proteins were eluted with 100 mM glycine (pH 2.5). The eluted proteins were collected directly into an Amicon spin column (MWCO: 5 kDa, Ultra-15, Bedford, MA) and buffer exchanged to PBS by means of centrifugation.

2.3.5 SDS-PAGE and Western blotting

Using a NuPAGE® Novex® 4 - 12% Bis-Tris mini-gel system (Life Technologies), SDS-PAGE was conducted under either reduced or non-reduced conditions. For reduced SDS-PAGE, the affinity purified antigen was mixed with lithium dodecyl sulfate (LDS) sample buffer and reducing agent (dithiothreitol (DTT)) and then incubated at 70 °C for 10 mins before loading on to the gel. For non-reduced SDS-PAGE, the sample was mixed with LDS sample buffer without addition of DTT. Gels were stained with Coomassie blue for protein detection.

For Western blotting, the proteins on the SDS-PAGE gel were electroblotted onto a nitrocellulose membrane using XCell II Blotting Module (Life Technologies). The membrane was treated with blocking solution (0.5% BSA in PBS) for 1 hour with gentle shaking, followed by incubation with 1/2000 dilution of the BSI mAb (mAb stock concentration 1 mg/mL) in blocking solution for 1 hour. The membrane was quickly washed with 18 MΩ water, followed by 3 consecutive washes with TBS/ 0.05% v/v Tween (TBST) (10 min × 20 mL/each). The HRP-conjugated goat anti-mouse secondary antibody was 1/40,000 diluted in blocking solution, and
the membrane was incubated for 1 hour, followed by washing as above. Chemiluminescence HRP substrate was incubated with the membrane for 2 min before detection using FluorChem™ Imaging System (Alfa Innotech, San Leandro, CA).

### 2.3.6 In-gel and in-solution tryptic digestion

For in-gel digestion, the protein bands of interest were excised and minced into small pieces (~ 0.5 mm²). Coomassie dye was removed by 2-3 cycles wash with acetonitrile and rehydration with 100 mM ammonium bicarbonate. After destaining, the proteins were reduced by incubation with 10 mM DTT for 30 min at 56 °C and then alkylated with 55 mM iodoacetamide (IAA) for 60 min at room temperature in the dark. Trypsin was added to the gel pieces and incubated for 40 mins on ice. The unabsorbed trypsin solution was replaced with 50 mM ammonium bicarbonate, and sufficient volume was added to cover the gel pieces. Samples were incubated overnight at 37 °C, and then the supernatant was removed and stored. Gel pieces were further extracted with 5% formic acid at 37 °C for 15 min with constant shaking. The extracted solution was combined with the stored supernatant and completely dried using a Speed Vac. Peptides were redissolved in 10 µL 0.1% formic acid for LC-ESI-MS or MALDI-TOF-MS analysis.

For in-solution digestion, the protein solution (IP pull-down mixture, M-LAC material or normalized M-LAC material) was denatured by 6M guanidine·HCl, reduced by 10 mM DTT (30 min at 56 °C) and alkylated by 50mM IAA (60 min at room temperature in the dark). The reduced and alkylated protein solution was buffer exchanged to 50 mM ammonium bicarbonate using Microcon ultracentrifuge device (5 kDa MWCO, Millipore, Billerica, MA) for 5 cycles (10 000× g for 5 min per cycle). Trypsin was added into the protein solution at ~ 1:25 ratio (enzyme
to protein) and incubated for 6 hrs. The enzymatic reaction was stopped by 5% formic acid and the digest was stored at -80 °C until analysis.

**2.3.7 LC-ESI-MS and MALDI-TOF-MS analysis**

For LC-ESI-MS, an LTQ-FT mass spectrometer (Thermo Fisher Scientific) interfaced with an Ultimate 2000 nanoLC (Dionex, Sunnyvale, CA) was used. A 75 μm i.d. capillary column (New Objective, Woburn, MA) was packed in-house with Magic C₁₈ (3 μm, 200 Å pore size) stationary phase (Michrom Bioresources, Auburn, CA). Peptides were eluted at a flow rate of 200 nL/min, and MS/MS spectra were obtained in a data-dependent mode in which the 8 most intense peaks in each MS scan were chosen for fragmentation. An m/z width of +/- 1 Da was employed to isolate the peptide precursor ions, and a 35% normalized collision energy was used to fragment the isolated peptides. Dynamic exclusion was utilized with exclusion duration of 30 s and no repeat counts. Protein identification was achieved by searching the Swiss-Prot human protein database using SEQUEST.

MALDI-TOF mass spectra were acquired using an Applied Biosystems 4700 TOF/TOF Proteomics Analyzer equipped with delayed extraction and a 200-Hz repetition rate UV laser (355 nm). The instrument was externally calibrated using angiotensin I. The matrix, α-cyano-4-hydroxycinnamic acid (CHCA, Mass PREP, Waters, Milford, MA), was prepared in acetonitrile/water, containing 0.1% trifluoroacetic acid (TFA) (50:50, v/v), at a concentration of 7 mg/mL. 0.5 µL of the in-gel digest were spotted onto the MALDI plate with 0.4 µL matrix using a thin-layer spotting method. Protein identification was accomplished by searching the Swiss-Prot human protein database using MASCOT in the peptide mass fingerprinting (PMF) mode.
2.3.8 Deglycosylation and glycan-specific staining

Enzymatic deglycosylation using PNGase F was achieved in either native or reduced conditions. For deglycosylation of the native protein, affinity purified antigen was mixed with 250 mM phosphate buffer (pH 7.5) at a 1:4 ratio, followed by the addition of PNGase F (1 Sigma unit enzyme for 20 µg Hpt) and incubation at 37 °C for up to 4 days. Two additional aliquots of PNGase F were transferred into the reaction mixture at the end of the 2\textsuperscript{nd} and 3\textsuperscript{rd} days, respectively. For deglycosylation under reduced conditions, the sample was mixed with reaction buffer (250 mM phosphate buffer, pH 7.5) and denaturation buffer (1% SDS and 1% mercaptoethanol), and then heated at 100 °C for 5 mins. After cooling to room temperature, PNGase F was added (1 Sigma unit enzyme for 20 µg Hpt), and the mixture was allowed to incubate at 37 °C for 3 - 4 hours with constant shaking.

Glycan-specific staining was accomplished using periodic acid and a Schiff\textsuperscript{’}s base reagent\textsuperscript{14}. Briefly, the gel was first fixed overnight with 40% ethanol and 7% acetic acid in order to minimize protein diffusion. Glycans attached to the glycoprotein were oxidized by 1% periodic acid and 3% acetic acid for 1 hour to generate aldehyde groups. The gel was then washed thoroughly with deionized water to remove periodic acid. After washing, the gel was incubated with Schiff\textsuperscript{’}s base reagent for 1 hour in the dark. A solution of 0.6% potassium metabisulfite and 3% acetic acid was used to wash the gel in order to prevent color fading.

2.3.9 Surface plasmon resonance (SPR) analysis

Surface plasmon resonance was performed on a T100 instrument (Biacore, Uppsala, Sweden). The sensor chip (type CM5, series S) surface was first activated with a 1:1 mixture of
0.2 M N-ethyl-N’-(3-dimethylaminopropyl) carbodiimide (EDC) and 0.05 M N-hydroxysuccinimide (NHS) in water, followed by a goat anti-mouse antibody solution (30 μg/mL in 10mM sodium acetate, pH 5.0). The goat anti-mouse antibody (mouse antibody capture kit, Biacore) was immobilized on the sensor chip surface through amine coupling, and the remaining unreacted NHS binding sites were blocked with 1 M ethanolamine-HCl, pH 8.5. For SPR analysis, mAb was first passed through the chip channel, and, after washing with running buffer, the serially diluted sample was introduced and the interaction recorded in real-time. Regeneration of the sensor surface between analysis cycles was achieved using 10 mM glycine-HCl, pH 1.7. The running buffer was 10 mM HEPES, 150 mM NaCl, 3.0 mM CaCl₂, 1.0 mM EGTA, and 0.005% Tween-20, pH 7.4. All binding data were analyzed using the Biomolecular Interaction Analysis evaluation program version 3.1 (Biacore).

2.3.10 Sandwich ELISA

mAb #1 was conjugated to HRP as follows: a total of 100 μg of EZ-link™ plus activated peroxidase was reconstituted in 10 μL ultrapure water and added to the 100 μL antibody solution (1 mg/mL mAb #1 in PBS, pH 7.8). Immediately, 1.1 μL of freshly-prepared sodium cyanoborohydride (0.5 M in PBS) was added to the antibody solution, and the solution was then incubated for 3 hours at R.T. (final concentration of sodium cyanoborohydride, 5 mM). Four L of 3 M ethanolamine was added to the antibody solution to quench the reaction. The reaction solution was buffer exchanged to PBS and incubated overnight with Sepharose Protein G beads (Thermo Fisher Scientific) at 4 °C. The HRP-mAb #1 conjugate was captured and eluted from the Protein G column using 0.1 M glycine, pH 2.5, and the eluate was immediately neutralized with 1 M Tris (1/10 volume), followed by buffer exchange to PBS. The activity of the HRP-mAb
#1 conjugate was confirmed by its ability to turn the TMB substrate from colorless to blue. For the sandwich ELISA, mAb #2 (20 μg/mL in PBS) was coated onto a 96-well plate. Serially diluted standard Hpt solution (0.005 - 10 μg/mL) was added into each well, followed by incubation with HRP-mAb #1 conjugate for standard curve generation. A negative control (blank) was used to monitor the non-specific signal. For Hpt concentration determination in crude plasma, the plasma was diluted 600 times with PBS. For Hpt concentration determination in M-LAC material, the resulting M-LAC material was directly used with no further treatment.

2.3.11 Hybridoma cell culture and mAb purification from supernatant

Two mAb-secreting hybridoma cell lines were cultured using hypoxanthine-aminopterin-thymidine (HAT)-medium shown in Table 2-1. A total of 250 mL of supernatant was collected for each hybridoma. The supernatant was filtered using 0.22 μm filtration unit (Millipore, Billerica, MA). The mAb was purified from the supernatant by means of protein G and protein L affinity chromatography (Life Technologies, Carlsbad, CA), according to the protocol provided by the manufacturer. SDS-PAGE was used to determine the recovery of purified Ab. The reactivity of the mAb (in the supernatant) was tested by probing separated plasma protein using the supernatant in a Western blotting format, as described in the Results and Discussion.
<table>
<thead>
<tr>
<th>Component</th>
<th>Supplier</th>
<th>Catalog No.</th>
<th>Initial Concentration</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPMI 1640 w/o glutamine</td>
<td>Invitrogen</td>
<td>21870-076</td>
<td>1X</td>
<td>415ml</td>
</tr>
<tr>
<td>Antibiotic-antimycotic solution</td>
<td>sigma</td>
<td>A5955-100mL</td>
<td>10000 ug /ml (100X)</td>
<td>5ml</td>
</tr>
<tr>
<td>GlutaMAX I supplement</td>
<td>Invitrogen</td>
<td>35050-061</td>
<td>200 mM (100X)</td>
<td>5ml</td>
</tr>
<tr>
<td>Sodium Bicarbonate 7.5%</td>
<td>Invitrogen</td>
<td>25080-094</td>
<td>7.5%</td>
<td>5ml</td>
</tr>
<tr>
<td>FBS (low IgG)#</td>
<td>Invitrogen</td>
<td>16250-086</td>
<td>(5 ug/mL)</td>
<td>50ml</td>
</tr>
<tr>
<td>Sodium Pyruvate MEM 100MM</td>
<td>invitrogen</td>
<td>11360-070</td>
<td>100 mM (100X)</td>
<td>5ml</td>
</tr>
<tr>
<td>Non ess amino acids</td>
<td>invitrogen</td>
<td>11140-050</td>
<td>100mM (100X)</td>
<td>5ml</td>
</tr>
<tr>
<td>HAT supplement</td>
<td>invitrogen</td>
<td>21060-017</td>
<td>50X</td>
<td>10mL</td>
</tr>
</tbody>
</table>

*HAT : sodium hypoxanthine, aminopterin and thymidine. Eliminate unfused myeloma cells

# FBS, fetal bovine serum
2.4 Results and Discussion

The general workflow of mAb proteomics for lung cancer biomarker discovery is shown in Figure 2-1. Pooled lung cancer patients plasma was subjected to 3-sequential steps of affinity fractionation, including high abundant plasma protein depletion, glycoprotein enrichment and polyclonal antibody affinity chromatography normalization. The resulting protein mixtures were used for mouse immunization and hybridoma generation. High throughput ELISA screening identifies hybridomas which can differentiate between lung cancer and normal. Lung cancer specific mAb was produced in large quantity by either hybridoma tissue culture or from ascites fluid. The mAb is purified in large quantity for antigen identification, antibody-antigen interaction characterization and immunoassay development, which are the focus of this work. A mass spectrometric (MS)-based strategy combined with simple immunoassay for antigen identification was developed and characterization of globally produced monoclonal antibodies was performed. Surface plasmon resonance proved high quality mAb (affinity constant at nM level) has been produced using mAb proteomics. This global approach for production and screening of mAb library can potentially be developed as diagnostic immunoassay kits after the use of the mAbs for the discovery and ultimate validation of biomarkers.
Figure 2-1. General workflow for lung cancer biomarker discovery using mAb proteomics.
2.4.1 Hybridoma generation and screening

We collaborated with Biosystems International to discover lung cancer specific mAbs. Pooled lung cancer patients plasma was used to prepare the immunogen, as described above (section 2.2.2). A total of 3848 hybridomas were generated, among which, 1051 high/medium mAb-producing hybridomas were underwent ELISA screening and 184 mAbs were found to be able to differentiate pooled lung cancer from normal. Among them, the majority (160) was found to be specific for the cancer and 24 are specific for the normal. Three lung cancer specific mAbs (#1, #2, and #3) were selected to produce in large quantity from ascite fluids and subject to antigen identification, mAb characterization and immunoassay development. Also, 2 hybridoma cell lines were cultured in lab and their reactivities were characterized, too. A summary of experiments accomplished is shown in Figure 2-2.
Figure 2-2. Antibodies and hybridomas characterized in this work. (A) mAbs #1, 2 and 3 were from BSI. An antigen identification method was developed combining mass spectrometric (MS)-based strategy with simple immunoassay. The antigens were identified to be Hpt-β, Hpt-α and CFH for mAb #1, 2 and 3, respectively. The antibody-antigen interaction between anti-Hpt and its antigen was characterized using Western blotting, SPR and sandwich ELISA. (B) Two hybridomas were cultivated and each mAb purified using protein G and protein L chromatography. These two hybridomas were also characterized using Western blotting (screening).
2.4.2 Antigen identification for 3 lung cancer specific mAbs

Immunoprecipitation followed by MS analysis is typically used for antigen identification\(^9\),\(^10\), but during affinity purification, not only the antigen but also its natural interacting partners and non-specific “sticky” proteins can be co-precipitated with the antigen. As a result, MS analysis will not directly yield a single protein identification, but rather a list of proteins. For example, when mAb #1 was used to IP pull-down the antigen from plasma, after trypsin digestion and LC-ESI-MS analysis, more than 30 proteins were identified. The top 9 proteins (Table 2-2) were all identified with more than 10 peptide digest fragments. Such a result makes it challenging to identify the antigen unambiguously and requires additional steps. In one case\(^9\), a cDNA expression library was used to screen generated Abs, and the results from the screening were correlated with MS data to determine the antigen identity.
## Table 2-2, Proteins identified from the IP pulldown mixture using mAb #1*

<table>
<thead>
<tr>
<th></th>
<th>Protein Name</th>
<th>Peptide Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Haptoglobin</td>
<td>103</td>
</tr>
<tr>
<td>2</td>
<td>Serum albumin</td>
<td>33</td>
</tr>
<tr>
<td>3</td>
<td>C4b-binding protein α chain</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>Apolipoprotein A-I</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>Hemoglobin subunit β</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>Hemoglobin subunit α</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Ig alpha-1 chain C region</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>Ig kappa chain C region</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>Ig gamma-3 chain C region</td>
<td>10</td>
</tr>
</tbody>
</table>

*The criteria used are as following: Xcorr >= 1.9 when charge is +1, >= 2.2 when charge is +2, >= 3.8 when charge is +3; ΔCn >= 0.1 and Peptide Prophet >= 0.95.
Here we used Western blotting to identify the protein band from the IP pull-down, followed by MS analysis for the antigen identification. As shown in Figure 2-3, the antigen protein was first isolated from plasma using an immunoaffinity column with the specific mAb; the eluate was then separated by SDS-PAGE. Two separate gels were run: one was used for Coomassie blue staining and the other for Western blotting. The Western blotting membrane was probed with the same mAb used in the immunoaffinity purification step, and the protein band detected by the antibody was targeted for MS analysis. The theoretical molecular weight of the identified protein was also compared with that of the SDS-PAGE / Western blotting results. mAb #1 was found to target the Hpt-β chain. The identification of the antigen using MALDI-TOF-MS is shown in Figure 2-4, with a total of 11 tryptic peptides (56% sequence coverage of the β-chain) identified with a mass accuracy of better than 20 ppm. The major peaks from Hpt-β are annotated in Figure 2-4-C. The mAb #2 was determined to target Hpt-α (Figure 2-5), with both the α1 and α2 chains of Hpt15 being recognized by this mAb. The identification was accomplished using LC-ESI-MS. The sequence coverage for Hpt-α1 and Hpt-α2 are 40% and 52%, respectively. Two representative MS/MS spectra resulting from Hpt-α1 (Figure 2-5-A, band 4,) and Hpt-α2 (Figure 2-5-A, band 3) are shown in Figure 2-5-C. The peptide fragments (b- and y-ion series, Figure 2-5-C) unambiguously determined the identity of the antigen.
Figure 2-3. Antigen identification workflow. The affinity purified antigen mixture was separated by SDS-PAGE, and then probed with a specific mAb to locate the binding antigen. The MW region, which was reactive in Western blotting, was excised from the gel, digested with trypsin and sequenced by mass spectrometry.
Figure 2-4. Antigen identification for mAb #1. Reduced SDS-PAGE of the mAb #1 affinity purified mixture from pooled lung cancer patient plasma and stained with Coomassie blue (A). Bands 1 and 3 are the heavy (~ 50 kDa) and light chains (~ 25 kDa) of IgG, respectively. The protein (band 2) migrating at ~ 44 kDa was recognized by the mAb #1 (B) and was subsequently identified as Hpt-β chain by MALDI-TOF-MS (C).
Figure 2-5. Antigen identification for mAb #2. Reduced SDS-PAGE of an affinity purified mixture (5 µg) from IgG- and albumin-depleted pooled lung cancer plasma, stained with Coomassie blue (A). Bands 1 and 2 are albumin and Hpt-β chain, respectively. Bands 3 (~18 kDa) and 4 (~10 kDa) were both recognized by mAb #2 (B) and were subsequently identified as the Hpt-α2 and Hpt-α1 chain by LC-ESI-MS (C).
In plasma, native Hpt circulates as an \((\alpha\beta)_2\) tetramer with N-linked glycosylation sites on its \(\beta\)-chain\(^{15}\). The \(\alpha\)-chain is polymorphic, and Hpt-\(\alpha_2\) is the product of gene duplication with approximately twice the length of Hpt-\(\alpha_1\)^{16,17}. Haptoglobin (either Hpt-\(\alpha\)^{18}, Hpt-\(\beta\)^{19-21} or whole Hpt protein\(^{22}\)) has previously been suggested to be candidate marker for various cancers, including lung cancer. However, in our study, we expected that the majority of Hpt from the lung cancer plasma would be removed following immunodepletion with the IgY-antibodies. We performed a shotgun proteomic analysis (see Experimental Section) on the M-LAC and normalized M-LAC fractions to estimate the amount of Hpt in the two samples. Haptoglobin was identified in both samples with total peptide counts of 14 and 3 for the M-LAC and normalized M-LAC fractions, respectively. This result indicates the presence of meaningful amounts of Hpt in the samples used for immunization. To accurately determine the levels of Hpt, we measured the Hpt concentration in the M-LAC sample (the protein mixture after M-LAC glycoprotein enrichment) using a sandwich ELISA assay developed with mAbs #1 and #2 (see below). The concentration of Hpt in M-LAC sample was determined to be \(\sim120\) ng/mL. Unfortunately, there was not sufficient amount of normalized M-LAC sample for direct Hpt concentration measurement by ELISA. However, based on the peptide counts for the Hpt before and after normalization, the levels of Hpt were estimated to have been further reduced by at least 5 fold (14 peptides vs 3 peptides). Therefore, it can be estimated that the level of Hpt injected into the mouse was in the low ng/mL level. This estimation suggests that this global method of immunization has the potential of eliciting an immune response to low abundance proteins, such as those originating from tumor cells but secreted into the bloodstream, if proper sample pretreatment is achieved. In this study, we have only focused on 3 of the discriminating mAbs; however, at least 184 hybridomas have been shown to discriminate between lung cancer and
matched controls. Future antigen identification of this set of mAbs could reveal potentially novel markers for NSCLC.

Turning to mAb #3, the antigen was identified in a similar manner as above and found to be complement factor H (CFH, Figure 2-6). Complement factor H is a plasma glycoprotein mainly produced in the liver in healthy people; however, CFH has also been observed to be expressed at high levels in several tumor cell lines including cervix\textsuperscript{23}, bladder\textsuperscript{24} and ovary tissue\textsuperscript{25}. As a key inhibitor of the alternative complement pathway\textsuperscript{26}, CFH has been suggested to be utilized by tumor cells to inhibit the complement-mediated immune response\textsuperscript{27}. Recently, lung cancer cells (NSCLC) were found to specifically express high levels of complement factor H to protect tumor cells from being destroyed by complement-dependent lysis\textsuperscript{28, 29}.

### 2.4.3 Characterization of 2 hybridomas

Two hybridoma cell lines were cultured and the supernatant collected. The antibody was first purified using a Protein G column. Protein G which targets the Fc domain of the Ab will capture both mouse mAb and bovine antibody (from cell culture medium). A large amount of Ab was affinity purified using the Protein G column, as shown in Figure 2-7-A. The mouse mAb was further purified using a Protein L column. Protein L specifically binds antibody containing kappa light chains, which are usually presented in a mouse antibody but not in a bovine antibody. Therefore, Protein L will separate the mouse Ab away from the bovine antibody. However, after protein L purification, no antibody was obtained, as evidenced by SDS-PAGE (Figure 2-7-A). The possible reasons are the produced mAb does not contain a kappa light chain or very little amount was produced.
The mAb reactivity against depleted plasma was tested using Western blotting. Albumin and IgG depleted plasma was separated by SDS-PAGE, transferred to a membrane and probed with hybridoma supernatant or fetal bovine serum (FBS). The Western blotting image of the membrane, probed with FBS and hybridoma #1, is shown in Figure 2-7-B. A similar pattern is observed for membranes probed with either FBS or the hybridoma #1 supernatant. Hybridoma #2 gave similar results as hybridoma #1. The similar pattern between FBS and the two hybridomas indicates these two hybridomas are low antibody-producers and further characterization of these hybridomas was stopped.
Figure 2-6. Antigen identification for mAb #3. Reduced SDS-PAGE of an affinity purified mixture from IgG- and albumin-depleted lung cancer plasma, stained with silver nitrate (A). The protein with a MW of ~ 140 kDa was recognized by the mAb (B) and was identified as complement factor H by LC-ESI-MS (C).
Figure 2-7. Antibody purification from hybridoma supernatant and mAb reactivity test using Western blotting. (A) SDS-PAGE of affinity purified antibody. Left, antibody purified using protein G; Right, antibody purified using protein L. lane 1, molecular weight marker; lane 2, hybridoma #1; lane 3, hybridoma #2. (B) Western blotting analysis against albumin and IgG depleted human plasma. Left gel was probed with fetal bovine serum (bovine antibody) and right gel was probed with supernatant from hybridoma #1 (mouse antibody). Lane 1, 1:10 dilution; lane 2, 1:100 dilution. The bands corresponding to antibody heavy and light chains are present, mainly due to the cross-reactivity of the secondary detection antibody.
2.4.4 Western blotting

It is well known that glycans attached to an antigen could affect Ab-Ag interaction\textsuperscript{23, 30}, either directly through hydrogen bonding and/or electrostatic interaction, or indirectly by glycosylation-induced conformational changes of the protein. To test whether the affinity of anti-Hpt-\(\beta\) was influenced by Hpt glycosylation, the binding of mAb #1 was examined for both the glycosylated and deglycosylated form of Hpt. PNGase F was used to remove the N-linked sugars from Hpt, and the deglycosylated protein was subsequently examined by reduced SDS-PAGE and Western blotting (Figure 2-8). The identified Hpt-\(\beta\) chain has a MW ~ 44 kDa (Figure 2-4-A, band 2), which is 13 kDa larger than its theoretical MW (31 kDa), due to the glycosylation attached to the chain. As shown in Figure 2-8, the Hpt-\(\beta\) chain is seen to migrate with a MW of 44 kDa (Figure 2-8-A, lane 1); after PNGase F treatment, Hpt-\(\beta\) shifted to an apparent MW of 31 kDa (Figure 2-8-A, lane 2). The identities of Hpt-\(\beta\) and deglycosylated Hpt-\(\beta\) were determined using MALDI-TOF-MS (Figure 2-9). Glycan-specific staining confirmed that the shift in MW was due to the removal of the glycans (Figure 2-8-B). A separate Hpt preparation, which underwent all the deglycosylation steps except for the addition of PNGase F, was run in parallel with the deglycosylated Hpt as a positive control (Figure 2-8-C, lane 1). Western blotting indicated that N-linked glycosylation on Hpt-\(\beta\) is essential for the Ab-Ag interaction (Figure 2-8-C, lane 2). In a separate experiment, Hpt was subjected to PNGase F treatment under native conditions, followed by \textit{non-reduced} SDS-PAGE and Western blotting (Figure 2-10). As expected, the complete removal of all glycans from the native Hpt was not possible, even after a long incubation time (4 days) and repetitive additions of PNGase F. Nonetheless, the signal for Western blotting of the Ab-Ag interaction was weaker for the partially deglycosylated protein,
further suggesting that the glycan structures on Hpt-β chain are necessary for the Ab-Ag interaction.
Figure 2-8. N-linked glycosylation is essential for the mAb-Ag interaction. Glycosylated and deglycosylated haptoglobin were separated by SDS-PAGE and stained with either Coomassie blue for protein detection (A), Schiff’s base reagent to determine the deglycosylation efficiency (B) or transferred to the nitrocellulose membrane for Western blotting (C, probed with anti-Hpt-β mAb #1).
Figure 2-9. In-gel digestion and MALDI-TOF-MS analysis confirmed the band identities. Protein bands corresponding to molecular weight 44 kDa and 31 kDa were excised, in-gel digested and analyzed by MALDI-TOF-MS. Mascot database search result indicate both bands are Hpt molecule. After deglycosylation, the signal corresponding to glycopeptides, which present at high molecular weight region, disappeared, while deglycosylated glycopeptides appeared at low molecular weight region.
Figure 2-10. PNGase F Treatment of Hpt under native conditions and Western blotting Using mAb #1. (A) Deglycosylation of Hpt under native conditions. Lane 1, control; Lane 2, 2-day incubation; Lane 3, 3-day incubation; Lane 4, 4-day incubation. (B) Western blotting of deglycosylated Hpt with mAb #1. Lane 1, control; Lane 2, Hpt after 4-day incubation.
2.4.5 Surface plasmon resonance

Surface plasmon resonance (SPR) was employed to measure the affinity constant (K_D) of mAb #1 (anti-Hpt-β) against native Hpt purified from the pooled lung cancer and matched control plasma samples. Multiple rounds of affinity purification were conducted to obtain sufficiently pure Hpt. Anti-Hpt-β was immobilized onto two flow channels of the same SPR sensor chip, one for the sample and the other to serve as a reference channel or control. A serially diluted affinity purified native Hpt sample was introduced, and the interaction was recorded in real time. A similar K_D (10^-9 mol) of anti-Hpt-β against native Hpt (glycosylated tetrameric form) was observed for both the lung cancer and matched control samples (Figures 2-11-A, B).

To further characterize the reactivity of anti-Hpt-β, we measured the K_D of anti-Hpt-β with a) reduced Hpt-β chain (glycosylated monomeric form, Figure 2-11-C) and b) reduced and deglycosylated Hpt-β chain (deglycosylated monomeric form, Figure 2-11-D). Due to the limited availability of lung cancer patient plasma samples, these SPR experiments were conducted using Hpt isolated from pooled, untreated normal plasma. The results indicated that mAb #1 binds with tight affinity to the native Hpt tetramer, with a K_D of ~ 10^-9 M (Figure 2-11-B), versus 2 orders of magnitude lower for the reduced Hpt (K_D of ~ 10^-7 M, Figure 2-11-C). As expected, no interaction was observed for reduced and deglycosylated Hpt (Figure 2-11-D). The affinity constant of mAb #2 (anti-Hpt-α) against native Hpt was also measured by SPR and the K_D was found to be in the 10^-8 M range.
Figure 2-11. Surface plasmon resonance analysis of anti-Hpt-β with Hpt. Panels A-D are sensorgrams of the interaction of anti-Hpt-β with (A) native Hpt (lung cancer); (B) native Hpt (matched control); (C) reduced Hpt-β chain (matched control); and (D) reduced and deglycosylated Hpt-β chain (matched control).
The favorable interaction of mAb #1 with the tetrameric form in comparison to the monomeric form suggests that anti-Hpt-β was generated to the protein that actually exists in plasma, i.e., the native form. This is an important result since mAbs are typically produced against synthetic peptides or recombinant protein fragments which generally are not the native form. No interaction between anti-Hpt-β and the reduced and deglycosylated Hpt was observed, further suggesting that the mAb proteomics process generates mAbs to the protein as it exists in plasma. Ignoring serum albumin, over 80% of the plasma proteome is glycosylated\textsuperscript{31,32}, and many of these glycoproteins require their glycans to be functionally active. In summary, we have demonstrated that the mAb proteomics process can produce large numbers of monoclonal antibodies, some of which are of high affinity (K\textsubscript{D} in nmol range) that target the native structure and associated PTMs.

It is interesting to note that similar affinity constants of anti-Hpt-β were observed for both the lung cancer and matched control samples. Aberrant glycosylation changes on Hpt-β chain have been found for various cancers\textsuperscript{20-22}, including lung cancer\textsuperscript{22}. This result seems to suggest that, although the interaction of anti-Hpt-β mAb and the Hpt protein is affected by the presence of glycan structures, the high affinity of anti-Hpt-β may not be specific for cancer-induced glycosylation changes. Thus, the high differentiating power of this mAb for lung cancer vs. normal plasma seems to mainly result from the Hpt’s abundance difference between the lung cancer and matched control sample.

\textbf{2.4.6 Sandwich ELISA of Hpt}

The sandwich ELISA assay is taken to be the gold standard for measuring the concentration of proteins in crude and complex mixtures; however, the development of an
ELISA assay requires highly specific antibody pairs targeting separate epitopes on the antigen, often a difficult goal to achieve. In this study, since mAb #1 and mAb #2 were found to target different subunits on the same protein, the potential to develop an ELISA assay was examined. mAb #1 was labeled in-house with HRP and used as the detection antibody, as shown in Figure 2-12-A. Activated HRP was chosen to label the antibody, and the resulting Schiff base was reduced to form a stable secondary amine with sodium cyanoborohydride. A standard sandwich ELISA operation procedure was followed, and the calibration curve was generated using Hpt ranging from 0.005 - 10 μg/mL (Figure 2-12-B). The sandwich ELISA assay was used to measure a) the Hpt concentration in the M-LAC sample from pooled lung cancer patients plasma and b) the Hpt levels in crude plasma of a subset of individual lung cancer patients and controls. For the M-LAC sample, as noted earlier, the Hpt concentration was found to be ~ 120 ng/mL. For the individual samples, the Hpt concentration in the lung cancer patients was found to be ~ 3 mg/mL and in the normal samples to be 1 mg/mL (Table 2-3). This result confirmed the expected up-regulation of Hpt in the plasma of the lung cancer patients and served as an independent check of the cancer-specific mAb screening. The high number of mAbs generated by the mAb proteomics clearly increases the chances of producing mAb pairs suitable for the development of sandwich ELISA assays. This can be an important potential advantage of this method.
Figure 2-12. Sandwich ELISA using mAb #1 and mAb #2. (A) mAb #1 was labeled with HRP for detection and mAb #2 was used as the capture Ab. (B) The standard curve was generated using Hpt standard ranging from 0.005 - 10 µg/mL. The haptoglobin concentration in M-LAC material, lung cancer and matched control plasma sample were measured using this assay.
<table>
<thead>
<tr>
<th>Gender</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Healthy</td>
<td>Healthy</td>
<td>Lung cancer*</td>
<td>Lung cancer*</td>
<td>Lung cancer*</td>
</tr>
<tr>
<td>Smoking</td>
<td>No</td>
<td>30 years, 1 pack/day</td>
<td>No</td>
<td>45 years, 1.5 pack/day</td>
<td>35 years, 1 pack/day</td>
</tr>
<tr>
<td>Hpt Conc. (mg/mL)</td>
<td>0.82</td>
<td>0.76</td>
<td>3.4</td>
<td>2.86</td>
<td>3.2</td>
</tr>
</tbody>
</table>

* All the lung cancer samples are squamous cell carcinoma without metastasis.
2.5 Conclusions

In this work, we used MS-based methods combined with immunoassay to identify antigens for globally produced and screened mAbs. Ab-Ag interaction were characterized using Western blotting, SPR and sandwich ELISA. The Western blotting results indicated that the reactivity of mAb to the beta chain of Hpt was affected by the glycosylation on the chain. Surface plasmon resonance revealed that this mAb was generated to the native form of the protein in plasma with tight binding (K_D in the low nM range). Also, using a sandwich ELISA from two mAbs of Hpt and LC-MS based proteomics analysis, the Hpt concentration in the immunogen mix was determined to be in the low ng/mL range, demonstrating that the mAb proteomics used in this work can generate mAbs against low abundance plasma proteins, e.g. potentially proteins originating from tumor cells. The results of this study indicate that mAb proteomics is able to generate high quality mAbs in a high throughput manner and should be important for future biomarker research efforts.
2.6 References


Chapter 3

Ultrasensitive Characterization of Site-Specific Glycosylation for Affinity Purified Haptoglobin from Lung Cancer Patients Plasma Using 10 µm i.d. PLOT LC-LTQ-CID/ETD-MS*

* The main part of this chapter has been submitted to *Analytical Chemistry.*
3.1 Abstract

Site-specific analysis of protein glycosylation is important for biochemical and clinical research efforts. Glycopeptide analysis using liquid chromatography - collision induced dissociation / electron transfer dissociation - mass spectrometry (LC-CID/ETD-MS) allows simultaneous characterization of glycan structure and attached peptide site. However, due to the low ionization efficiency of glycopeptides during electrospray ionization (ESI), 200-500 fmol of sample per injection is needed for a single LC-MS run, which makes it challenging for the analysis of limited amounts of glycoprotein purified from biological matrices. To improve the sensitivity of LC-MS analysis for glycopeptides, an ultranarrow porous layer open tubular (PLOT) LC column (2.5 m x 10 µm i.d.) was coupled to a linear ion trap mass spectrometer (LTQ-CID/ETD-MS) to provide low level analysis of N-linked protein glycosylation heterogeneity. The potential of the developed strategy is demonstrated by the site-specific glycosylation characterization of haptoglobin (Hpt). A total of 26 glycoforms / glycan compositions on three Hpt tryptic glycopeptides were identified and quantified from 10 LC-MS runs with a consumption of 100 fmol Hpt digest (10 fmol per injection). Included in this analysis was the determination of the glycan occupancy level. At this sample consumption level, the high sensitivity of the PLOT LC-LTQ-CID/ETD-MS allowed glycopeptide identification and structure determination, along with relative quantitation, even for low abundant glycopeptides at ~ 100 attomole level. The PLOT LC-MS is shown to have sufficient sensitivity to allow characterization of site-specific protein glycosylation from trace levels of glycosylated proteins.
3.2 Introduction

Protein glycosylation, a biologically significant and complex co- and post-translational modification, involves the attachment of glycans to the protein backbone through asparagine (N-linked) and/or serine/threonine (O-linked)\(^1\). Glycosylation plays a key role in, among others, protein folding, subcellular localization, turnover and activity, cell proliferation and cell-cell interaction\(^2-4\). Cancer has a profound effect on the glycosylation machinery of a cell, leading to glycans that are markedly altered (either in abundance or structure) from that produced by a normal cell\(^2\). Glycosylation changes have been reported for many cancers\(^5-8\), as well as other diseases\(^9\). One of the observed alterations in cancer is site specific glycosylation, and various studies have reported on this modification. For example, N\(^{251}\) monosialylated alpha-fetoprotein (AFP) for hepatocellular carcinoma\(^5\), N\(^{45}\) fucosylated prostate specific antigen (PSA) for prostate cancer\(^6\), N\(^{134}\) sialylated kallikrein 6 for ovarian cancer\(^7\), N\(^{211}\) difucosylated tetraantennary glycan of haptoglobin for pancreatic cancer\(^8\) and N\(^{630}\) under-glycosylated transferrin for congenital disorder of glycosylations (CDGs)\(^9\).

Site-specific glycan characterization can be important not only for diagnostic purposes, but also for fundamental structural biology. Site-specific glycosylation on epidermal growth factor receptor (EGFR), for example, has been shown to modulate its function\(^10\). A point mutation of EGFR (N\(^{579}\) to Q\(^{579}\)), while maintaining glycosylation at other extra cellular sites of the receptor, can produce a functionally distinct receptor, which features decreased auto-inhibitory tether interactions and increased level of preformed dimers in the absence of EGF\(^{10}\). Such findings, as well as others\(^5-8,11\), support the importance of site-specific glycosylation characterization.
The LC-MS analysis of glycopeptides can be more difficult than analysis of non-modified peptides or oligosaccharides, as the chemical properties can be quite different between the glycan (hydrophilic) and peptide (hydrophobic) components. Since the glycosidic linkage in the glycopeptide is weaker than the peptide backbone, collision induced dissociation (CID)-MS will mainly produce glycosidic fragmentation, thus limiting information on the peptide sequence and linkage site. The addition of electron transfer dissociation (ETD) has allowed the direct LC-MS analysis of glycopeptides to characterize, in the same run, both glycan structure (CID-MS/MS) and peptide sequence/site attachment (ETD-MS/MS).

Comprehensive characterization of specific glycoproteins from biological matrices (enriched, for example, by immunoprecipitation) present significant challenges from both the limited amount of material generally available and the overall complexity of the protein. A single glycosylation site can present a number of glycans (micro-heterogeneity) and a specific glycopeptide can be a minor constituent in a peptide/glycopeptide digest mixture. Furthermore, glycopeptides will generally have poorer electrospray ionization efficiency than peptides; thus, co-eluting peptides may suppress glycopeptide signal. On top of these factors, due to the complexity of glycosylation, multiple LC-MS runs are necessary (fragmentation of different ions) to achieve comprehensive glycosylation characterization. A standard 75 µm i.d. capillary LC column routinely requires 200 to 500 fmol (or higher) sample for single LC-MS run, and multiple runs could easily result in up to 10 pmol of protein sample consumption. This level may simply not be available, e.g. from tissue or low level plasma circulating glycoproteins. For characterization of protein glycosylation, where only limited amounts of a glycosylated protein may be available, ultrasensitive LC-MS methods are required.
It is well-known that the ionization efficiency and thus the MS response will increase if the LC flow rate is decreased\textsuperscript{17-19}. Up to 100 times increase in response factor has been observed for glycan analysis, when the flow rate for ESI was decreased from 1 µL/min to ~ 30 nL/min\textsuperscript{20}. Such a low flow rate is best utilized with ultra-narrow i.d. LC columns for optimal separation, which is challenging. We introduced 10 µm i.d. porous layer open tubular (PLOT) LC columns for highly sensitive LC-MS analysis featuring high resolving power\textsuperscript{21-24}. The low level sample consumption of PLOT LC column should be especially important for multiple runs, necessary for comprehensive glycosylation analysis.

In this work, the performance of combining the 10 µm i.d. PLOT LC with LTQ-CID/ETD-MS is demonstrated in the comprehensive site specific glycan characterization of immunoprecipitated Hpt from grade 4 non-small cell lung cancer patient plasma at the sample consumption level of 100 fmol. Haptoglobin is a well-known acute phase plasma glycoprotein having an (αβ)\textsubscript{2} tetramer structure with 4 N-linked glycosylation sites on its β chain\textsuperscript{8}. The concentration of the glycoprotein in plasma increases during inflammation-associated conditions, including cancer. Recently, several studies have reported on cancer-induced glycosylation changes on Hpt\textsuperscript{8,25,26}, including prostate\textsuperscript{26}, pancreatic\textsuperscript{8} and lung cancer\textsuperscript{25}. Thus, Hpt is a relevant glycoprotein to select to demonstrate the power of the approach. Using affinity purified Hpt from pooled lung cancer patient plasma, the site-specific glycan identification, structure determination including core/antenna fucosylation isomers, relative quantitation and site occupancy, was accomplished from a total of 10 LC-MS runs, each with 10 fmol injection (total consumption of 100 fmol), allowing identification and quantitation down to 100 attomole of specific glycopeptide isoforms. The method is demonstrated to have sufficient sensitivity to allow characterization of site-specific protein glycosylation from trace amounts of glycoproteins.
3.3 Experimental

3.3.1 Samples and reagents

Plasma from 20 patients diagnosed with advanced non-small cell lung cancer (NSCLC, all grade 4 squamous cell carcinoma) was obtained from Proteogenex (Culver City, CA). The individual plasma samples were pooled on an equal volume basis. Anti-haptoglobin (Hpt) polyclonal antibody (pAb) and human Hpt ELISA kit were acquired from Genway (San Diego, CA). Protein G coated magnetic beads were purchased from Invitrogen (Carlsbad, CA). Bis(sulfosuccinimidyl) suberate (BS$_3$) was from Thermo Fisher Scientific (Waltham, MA). Sequencing grade trypsin and Glu-C were obtained from Promega (Madison, WI) and Roche Diagnostics GmbH (Penzberg, Germany), respectively. Phosphate buffered saline (PBS, 10 times concentrated) was from MP Biomedicals (Solon, OH). All other reagents were from Sigma-Aldrich (St. Louis, MO).

3.3.2 Haptoglobin immunoaffinity purification

Anti-Hpt pAb was crosslinked to the magnetic protein G beads using BS$_3$, according to the protocol supplied by the manufacturer. Briefly, 10 µg anti-Hpt pAbs in 40 µL PBS with 0.05% Tween-20 (PBST) was incubated with 0.3 mg protein G coated magnetic beads for 30 min at room temperature (RT). After a PBST wash, 250 µL of freshly-prepared 5 mM BS$_3$ (in PBST) was added to the pAb-beads, and the crosslinking reaction was carried out at RT for 30 min. The reaction was quenched by addition of 12.5 µL of 1 M tris(hydroxymethyl)aminomethane buffer (Tris, pH 7.5). The beads were washed with PBST 3 times and were then ready for use.
For Hpt purification, 1 µL of the pooled lung cancer plasma was diluted to 50 µL with PBST and then incubated with protein G beads to deplete human immunoglobulin. The resulting IgG-depleted plasma was incubated with pAb-beads for 1 hr at RT. After extensive washing (6 times with PBST, followed by 3 times with PBS), the beads were resuspended in 100 µL PBS and transferred to a clean tube. The bound Hpt was eluted using 50 mM glycine (pH 2.8) and neutralized with 1M Tris. The Hpt concentration in the pooled plasma and in the affinity purified sample was determined by the Hpt ELISA kit, according to the protocol provided by the manufacturer.

3.3.3 Trypsin and PNGase-F digestion

For trypsin digestion, immunoaffinity purified Hpt (1.4 µg, ~ 10 pmol) was buffer exchanged with 6 M guanidine hydrochloride in 50 mM ammonium bicarbonate by ultracentrifugation (10,000 × g, 10 min per cycle, 5 cycles) using a Microcon filter device with a 10 kDa MWCO membrane. The exchanged protein solution was reduced with dithiothreitol (DTT, final concentration 2 mM) for 30 min at 37 °C and alkylated with iodoacetamide (IAA, final concentration 10 mM) in the dark at RT for 30 min. The reduced and alkylated Hpt solution was then transferred to a 10 kDa MWCO filter device to remove guanidine hydrochloride and excess DTT and IAA and buffer exchanged to digestion buffer (50 mM ammonium bicarbonate, pH 7.5). Trypsin (enzyme : substrate = 1 : 50) was added to digest the protein for 4 hr at 37 °C.

For trypsin plus PNGase F digestion, the tryptic digest (0.3 µg, 2 pmol) was incubated with PNGase F(enzyme : substrate = 1 : 50 ) at 37 °C for 4 hr. The digestion was stopped by addition of 1% formic acid. The resulting Hpt digest was aliquoted at 1 µL per vial (~ 100 fmol/vial) and
stored at -80 °C. The Hpt digest was diluted 10 times with 0.1% formic acid to ~ 10 fmol/µL immediately before LC-MS analysis, and 1 µL sample (10 fmol Hpt, 1.4 ng Hpt) was injected.

### 3.3.4 PLOT LC-LTQ-CID/ETD-MS

The polystyrene-divinylbenzene (PS-DVB) PLOT LC column (2.5 m x 10 µm i.d.) was prepared according to procedures previously described23. Briefly, a degassed solution containing 200 µL styrene, 200 µL divinylbenzene, 600 µL ethanol and 5 mg azobisisobutyronitrile was filled into a 10 µm i.d. capillary pretreated with 3-(trimethoxysilyl)propyl methacrylate. Both ends of the capillary were sealed with septa, and the capillary was heated at 74 °C for ~16 h in a water bath. The column was washed with acetonitrile (> 50 column volumes) and evaluated using digested bovine serum albumin standard peptides (Michrom Bioresources, Auburn, CA), according to published protocols23, 27. A solid phase extraction (SPE) PS-DVB monolithic column (5 cm x 50 µm) was used for sample loading27. A diagram of the overall PLOT LC system can be found in Figure 3-1. The mobile phase flow was split immediately before the SPE-PLOT assembly to minimize the gradient delay, and the flow rate was maintained at 20 nL/min after the split. Gradient elution was performed using an Ultimate 3000 pump system (Dionex, Sunnyvale, CA) with mobile phase A as 0.1% (v/v) formic acid in water and mobile phase B as 0.1% (v/v) formic acid in acetonitrile. The gradient consisted of the following steps: (i) 10 min at 0% B for sample loading/washing; (ii) linear from 0 to 35% B over 35 min; (iii) linear from 35 to 80% B over 10 min; and finally (iv) isocratic at 80% B for 10 min.
Figure 3-1. PLOT LC-LTQ-CID/ETD-MS. (A) Scanning electron micrograph (SEM) of a monolith SPE column; (B) SEM of PLOT column; (C) Schematic representation of electrospray using the 10 μm i.d. PLOT column. (D) PicoClear tee allowed quick sample loading while minimizing band broadening. The 10-port valve was first set in the loading position, with tee a serving as splitter and tee b as connector. Sample was loaded onto the SPE column while salts and other ESI deleterious components were directed to waste. After loading, both the sample injection valve and 10-port valve were switched to the separation position (dashed line), and the SPE column was now positioned to be in-line with the PLOT separation column. The tees were reversed with tee a now serving as connector, while tee b acted as splitter.
An LTQ-CID/ETD-MS (Thermo Fisher Scientific, San Jose, CA) was used in all experiments. The mass spectrometer was operated in the data-dependent mode to switch automatically between full-scan MS, CID-MS$^2$ and ETD-MS$^2$. Briefly, after a survey full-scan MS spectrum from m/z 400 to 2000 (at a target value of 30,000 ions), a subsequent CID-MS$^2$ (at a target value of 20,000 ions and 35% normalized collision energy) and ETD-MS$^2$ (at a target value of 20,000 ions) were performed on the same precursor ion. The precursor ion was isolated with a ± 2 m/z width, starting from the most intense ion in the survey scan. CID-MS$^2$ and ETD-MS$^2$ were repeated four additional times to fragment the second to the fifth highest precursor ion. For ETD-MS$^2$, the ion-ion reaction time between the isolated precursor and reagent anion (fluoranthene) was 100 ms, with the reagent ion (at a target value of 200,000) 10 fold higher than the precursor ion$^{14,28}$. The total cycle time (11 scans) of 1.6 s was continuously repeated for the entire LC-MS run under data-dependent conditions with dynamic exclusion (30 seconds for 2 repeated spectra).

### 3.3.5 Data processing

For peptide identification, the mass spectrometric raw data was database searched against the Hpt sequence (SwissProt P00738) using SEQUEST (Thermo Fisher Scientific). Peptide identification was based on the following criteria: $\Delta$Cn $\geq$ 0.1, peptide probability < 0.001, Xcorr $\geq$ 1.9, 2.5 and 3.8 for singly, doubly and triply charged ions, respectively. For glycopeptide characterization including glycan structure, peptide sequence and site of attachment, CID-MS/MS and ETD-MS/MS were manually annotated, as described below.
3.4 Results and Discussion

In this work, we describe a 10 µm i.d. PLOT LC-LTQ-CID/ETD-MS for sensitive characterization and relative quantitation of site-specific N-linked protein glycosylation. A 2.5 m PLOT LC column was interfaced online with an LTQ-CID/ETD mass spectrometer. We demonstrate that site-specific characterization of protein glycosylation, including glycopeptide identification, glycan site elucidation and quantitation and site occupancy determination, can be completed within 10 LC-MS runs at a total sample consumption of 100 fmol (~13 ng Hpt). For a glycoprotein that contains only fully occupied glycosylation sites, even less LC-MS runs and sample consumption would be required.

The general strategy for protein glycosylation analysis using 10 µm i.d. PLOT LC-LTQ-CID/ETD-MS is shown in Figure 3-2. The target glycoprotein is first affinity purified from a biological mixture. A polyclonal antibody, rather than monoclonal antibody, is used to pull down as many target molecule isoforms as possible. The purified glycoprotein (Hpt) is digested with a suitable enzyme (trypsin) to generate a mixture of peptides and glycopeptides, which are then analyzed by the PLOT LC-LTQ-CID/ETD-MS. As noted, a total of 10 LC-MS runs (Figure 3-2, Panels A-C) were performed for comprehensive characterization of Hpt glycosylation. For each individual LC-MS run, only 10 fmol of protein digest was injected. For this sample amount, the PLOT LC-MS had sufficient sensitivity to characterize and quantitate lower level glycans present down to the 100 attomole level (dynamic range of 100). Compared to a C18 column, the PS-DVB based column has lower hydrophobicity\textsuperscript{21}, and its open tubular structure improves the recovery of large glycopeptides\textsuperscript{23}. Operating at the flow rate of 20 nL/min, the high-resolution PLOT LC column results in improved analyze ionization and decreased ion suppression.
compared to conventional size (75 µm i.d.) columns\textsuperscript{29}. All of these features together contribute to improved sensitivity for glycopeptide analysis.
Figure 3-2. General strategy for site-specific characterization of protein glycosylation using the 10 µm i.d. PLOT LC-LTQ-CID/ETD-MS.
3.4.1 Haptoglobin affinity purification

Haptoglobin is a high abundant plasma glycoprotein having an $(\alpha\beta)_2$ tetramer structure (M.W. \(\sim\) 130 kDa). Even though its concentration is relatively high, we chose Hpt as a model glycoprotein to demonstrate the performance of the ultrasensitive strategy, because aberrant glycosylation changes have been reported for Hpt in many cancers\textsuperscript{8,26} including lung cancer\textsuperscript{25}. A pool of 20 individual advanced non-small cell lung cancer patient plasmas (stage 4) was selected. One microliter of the pooled lung cancer plasma (2.7 μg/μL Hpt, \(\sim\) 20 pmol Hpt) was diluted 50 fold using PBS, then IgG depleted and finally Hpt affinity purified using magnetic beads, previously crosslinked with anti-Hpt polyclonal antibodies. Approximately 1.4 μg Hpt (\(\sim\) 10 pmol) could be obtained with a recovery of \(\sim\) 50 % and a purity of greater than 80%, as determined by SDS-PAGE. The Hpt concentration in the plasma and affinity purified sample was measured by ELISA.

Haptoglobin contains 4 N-linked glycosylation sites on its $\beta$-chain (N\textsuperscript{184}, N\textsuperscript{207}, N\textsuperscript{211} and N\textsuperscript{241}). Since Hpt is known to be easy to digest\textsuperscript{30}, trypsin was used as the proteolytic enzyme. For other glycoproteins, a strategy of multiple enzymes may be necessary\textsuperscript{31}. Trypsin digestion will generate 3 glycopeptides namely T1, T2 and T3, with T2 having two glycosylation sites (Table 3-1). The two glycans on T2 are very close and thus difficult to digest into single site glycopeptides. Although it has been reported that these two sites can be separated into single glycosylated peptides\textsuperscript{8}, in this work, for simplicity, we treat the glycopeptide T2 as a single entity since our goal is to demonstrate the ability of the strategy to achieve low level glycosylation characterization.
<table>
<thead>
<tr>
<th>Glycopeptide</th>
<th>Amino acid (AA) sequence</th>
<th>AA number (glycosylation site)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>MVSHHN$^{184}$LTTGATLINEQWLLTTAK</td>
<td>24 (1)</td>
</tr>
<tr>
<td>T2</td>
<td>NLFLN$^{207}$HSEN$^{211}$ATAK</td>
<td>13 (2)</td>
</tr>
<tr>
<td>T3</td>
<td>VVLHPN$^{241}$YSQVDIGLIK</td>
<td>16 (1)</td>
</tr>
</tbody>
</table>
3.4.2 Glycopeptides T1 and T3

Glycopeptides, sharing the same peptide sequence but differing in glycan structure, will closely elute in reversed phase LC separation\textsuperscript{15}. When subjected to CID fragmentation, glycopeptides will produce low molecular weight glycan oxonium ions such as m/z 366 (Hex-HexNAc\textsuperscript{+}), 528 (Hex-HexNAc-Hex\textsuperscript{+}) and 657 (NeuAc-Hex-HexNAc\textsuperscript{+}), etc\textsuperscript{9,32}. These oxonium ions can be used to identify the chromatographic peaks of glycopeptides in a complex LC-MS chromatogram. Three data-dependent PLOT LC-LTQ-CID/ETD-MS analyses were initially performed consuming a total of 30 fmol of the Hpt tryptic digest. Figure 3-3-A shows the base peak chromatogram for one of the LC-MS runs. A signal intensity of 2E6 was obtained at this low sample level; this intensity is comparable to that obtained for ~ 200 fmol Hpt, or 20 fold higher, using a 75 μm packed C\textsubscript{18} column for LC/MS (data not shown). To locate glycopeptides from the base peak chromatogram, the above three glycan oxonium ions at m/z of 366, 528 and 657 were extracted from the MS/MS spectra (Figure 3-3-B). Together with ETD-MS/MS (see below), the elution times for T1 and T3 were identified to be approximately 39 and 34.5 min, respectively (Figure 3-3-B). In addition, the retention time of T2 (to be discussed later) was estimated to be 23.5 min; the shorter time is expected given the two glycosylation sites on T2.
Figure 3-3. Oxonium ions indicate glycopeptide elution time. (A) Base peak chromatogram of PLOT LC-MS analysis for 10 fmol Hpt tryptic digest. (B) Oxonium ions were extracted to locate the glycopeptides in the complex LC-MS chromatogram. NL: normalized level. Key to symbols: Empty circle = Hexose, filled square = HexNAc, filled diamond = NeuAc.
ETD-MS/MS was used to identify the peptide sequence of the moderate to high intensity glycopeptides for T1 and T3. An example is shown in Figure 3-4-A, where a T1 glycopeptide (m/z 1222.3), eluting at 39.2 min, was fragmented by ETD. Nine c and 15 z’ ions could be identified, with the intact glycan structure attached to c6-9, c11-12 and c14-15. The complete z’ ion series and 4 c ions are annotated in Figure 3-4-A. This ETD-MS/MS spectrum identified the peptide sequence to be MVSHHN184LTTGATLINEQWLLTAK and confirmed the glycosylation site (N184). Although LC-MS analysis on PNGase-F treated tryptic digests can also correlate the peptide moiety to its corresponding glycopeptide, the ETD-MS/MS used here can be conducted during the same run as CID-MS/MS, thus consuming less sample. Moreover, ETD-MS/MS provides direct evidence of the site of glycan attachment to the peptides, which is not available from the PNGase-F method. While low signal intensity glycopeptides did not produce ETD fragmentation spectra (ETD is less sensitive than CID), since all glycopeptides from the same peptide eluted within a 2 minute window (data not shown) and the separation of T1, T2 and T3 was much greater than 2 minutes, it was reasonable to associate the low intensity glycopeptides with T1, T2 or T3 based on retention time. Returning to Figure 3-4, the glycan mass was determined to be 2206.1 Da, based on the mass difference between the observed precursor glycopeptide (m/z 1222.3, z = 4) and the theoretical peptide mass (2683.1 Da). A glycan composition of NeuAc2 HexNAc4Hex5 (theoretical mass 2205.9 Da) was assigned, and the glycan structure was constructed based on the known Man3GlcNAc2 core and the CID-MS/MS spectra. CID fragmentation in Figure 3-4-B generates glycan oxonium ions (366 and 657) and glycopeptide fragments, the latter allowing identification of the glycan structure as a disialylated biantennary complex (see inset in Figure 3-4-B).
Figure 3-4. Glycopeptide analysis using ETD and CID-MS/MS. (A) ETD-MS/MS cleaves the peptide backbone, generating c and z' ions. (B) CID-MS/MS cleaves the glycan linkage, generating oxonium ions and glycopeptide fragments. (C) Averaged mass spectra contain all the glycans on this glycopeptide. NL: normalized level. Key to symbols: empty circle = galactose, filled circle = mannose, filled square = HexNAc, filled diamond = NeuAc, filled triangle = fucose.
In order to determine the various glycoforms on glycopeptide T1, the MS spectra collected during the elution time window of 39 ± 1 min were averaged (Figure 3-4-C). GlycoMod\textsuperscript{34} was used to assign the possible glycoforms by adding the specific glycan masses to the corresponding peptide. The theoretical m/z values were then calculated and compared with the observed m/z values in the averaged spectra to identify the glycans. Table 3-2 lists all the identified site-specific glycans with their corresponding molecule mass and m/z. The glycan structures on glycopeptide T1, supported by CID-MS/MS in either the data-dependent or targeted MS/MS mode (see below), are presented in Figure 3-4-C. In a similar manner, 9 glycans were assigned on glycopeptide T3, ranging from biantennary to tetraantennary structures, with or without fucosylation (Figure 3-5-C). It should be noted that, for the initial 3 data-dependent PLOT LC-LTQ-CID/ETD-MS runs, high quality CID-MS/MS spectra were obtained for 5 and 6 glycans on glycopeptides T1 and T3, respectively. Four additional targeted LC-MS runs (CID-MS/MS only, Figure 3-2, panel B) were necessary, focusing on missed low abundant glycopeptide precursors (T1-02 and T1-04 on glycopeptide T1, Figure 3-4-C; T3-06, T3-07 and T3-09 on glycopeptide T3, Figure 3-5-C).
<table>
<thead>
<tr>
<th>Glycopeptide</th>
<th>Peptide mass</th>
<th>Glycoform&lt;sup&gt;2&lt;/sup&gt;</th>
<th>Glycan Structure&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Glycoform mass&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Glycoform m/z&lt;sup&gt;4&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>2681.1</td>
<td>T1-01</td>
<td>2-1S-0F</td>
<td>4595.8</td>
<td>1532.9 1149.9 920.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T1-02</td>
<td>2-1S-1F</td>
<td>4741.9</td>
<td>1581.6 1186.5 949.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T1-03</td>
<td>2-2S-0F</td>
<td>4887.0</td>
<td>1630.0 1222.8 978.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T1-04</td>
<td>2-2S-1F</td>
<td>5033.2</td>
<td>1678.7 1259.3 1007.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T1-05</td>
<td>3-2S-0F</td>
<td>5252.4</td>
<td>1751.8 1314.1 1061.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T1-06</td>
<td>3-3S-0F</td>
<td>5543.6</td>
<td>1848.9 1386.9 1199.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T1-07</td>
<td>3-3S-1F</td>
<td>5889.8</td>
<td>1897.6 1423.4 1139.0</td>
</tr>
<tr>
<td>T2</td>
<td>1459.6</td>
<td>T2-01</td>
<td>(2-2S-0F)+(2-1S-0F)</td>
<td>5580.3</td>
<td>1861.1 1396.1 1117.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2-02</td>
<td>(2-2S-0F)+(2-2S-0F)</td>
<td>5871.6</td>
<td>1958.2 1469.9 1175.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2-03</td>
<td>(2-2S-0F)+(3-2S-0F)</td>
<td>6236.9</td>
<td>2080.0 1580.2 1248.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2-04</td>
<td>(2-2S-0F)+(3-2S-1F)</td>
<td>6383.0</td>
<td>2128.7 1596.8 1277.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2-05</td>
<td>(2-2S-0F)+(3-3S-0F)</td>
<td>6528.1</td>
<td>2177.0 1633.0 1306.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2-06</td>
<td>(2-2S-0F)+(3-3S-1F)</td>
<td>6674.3</td>
<td>2225.8 1669.6 1335.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2-07</td>
<td>(2-2S-0F)+(4-3S-0F)</td>
<td>6893.5</td>
<td>2298.8 1724.4 1379.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2-08</td>
<td>(2-2S-0F)+(4-3S-1F)</td>
<td>7039.6</td>
<td>2347.5 1760.9 1408.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2-09</td>
<td>(2-2S-0F)+(4-4S-0F)</td>
<td>7184.7</td>
<td>2395.9 1797.2 1437.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2-10</td>
<td>(2-2S-0F)+(4-4S-1F)</td>
<td>7330.9</td>
<td>2444.6 1833.7 1467.2</td>
</tr>
<tr>
<td>T3</td>
<td>1796.1</td>
<td>T3-01</td>
<td>2-1S-0F</td>
<td>3710.8</td>
<td>1237.9 928.7 743.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3-02</td>
<td>2-2S-0F</td>
<td>4002.1</td>
<td>1335.9 1001.5 801.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3-03</td>
<td>2-2S-1F</td>
<td>4148.2</td>
<td>1383.7 1038.3 830.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3-04</td>
<td>3-2S-0F</td>
<td>4367.4</td>
<td>1456.8 1082.9 874.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3-05</td>
<td>3-2S-1F</td>
<td>4513.5</td>
<td>1505.5 1129.4 903.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3-06</td>
<td>3-3S-0F</td>
<td>4658.7</td>
<td>1553.9 1165.7 932.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3-07</td>
<td>3-3S-1F</td>
<td>4804.8</td>
<td>1602.8 1202.2 962.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3-08</td>
<td>4-2S-0F</td>
<td>4732.7</td>
<td>1578.6 1184.2 947.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3-09</td>
<td>4-3S-0F</td>
<td>5024.0</td>
<td>1675.7 1257.0 1005.8</td>
</tr>
</tbody>
</table>

# Glycoforms identified on Hpt. The structures of corresponding glycoforms can be found in Table S2.

<sup>1</sup> Glycan structure (text format) of corresponding glycoform. The first arabic number represents branclhing; second arabic number represents sialic acid number. "S" means sialic acid; the third arabic number represents fucose number. "F" means fucose.
<table>
<thead>
<tr>
<th>T1-01</th>
<th>T1-02</th>
<th>T1-03</th>
<th>T1-04</th>
<th>T1-05</th>
<th>T1-06</th>
<th>T1-07</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Structure" /></td>
<td><img src="image2" alt="Structure" /></td>
<td><img src="image3" alt="Structure" /></td>
<td><img src="image4" alt="Structure" /></td>
<td><img src="image5" alt="Structure" /></td>
<td><img src="image6" alt="Structure" /></td>
<td><img src="image7" alt="Structure" /></td>
</tr>
<tr>
<td>T2-01</td>
<td>T2-02</td>
<td>T2-03</td>
<td>T2-04</td>
<td>T2-05</td>
<td>T2-06</td>
<td>T2-07</td>
</tr>
<tr>
<td><img src="image8" alt="Structure" /></td>
<td><img src="image9" alt="Structure" /></td>
<td><img src="image10" alt="Structure" /></td>
<td><img src="image11" alt="Structure" /></td>
<td><img src="image12" alt="Structure" /></td>
<td><img src="image13" alt="Structure" /></td>
<td><img src="image14" alt="Structure" /></td>
</tr>
<tr>
<td>T2-08</td>
<td>T2-09</td>
<td>T2-10</td>
<td>T3-01</td>
<td>T3-02</td>
<td>T3-03</td>
<td>T3-04</td>
</tr>
<tr>
<td><img src="image15" alt="Structure" /></td>
<td><img src="image16" alt="Structure" /></td>
<td><img src="image17" alt="Structure" /></td>
<td><img src="image18" alt="Structure" /></td>
<td><img src="image19" alt="Structure" /></td>
<td><img src="image20" alt="Structure" /></td>
<td><img src="image21" alt="Structure" /></td>
</tr>
<tr>
<td>T3-05</td>
<td>T3-06</td>
<td>T3-07</td>
<td>T3-08</td>
<td>T3-09</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image22" alt="Structure" /></td>
<td><img src="image23" alt="Structure" /></td>
<td><img src="image24" alt="Structure" /></td>
<td><img src="image25" alt="Structure" /></td>
<td><img src="image26" alt="Structure" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* key to symbols can be found in Figure 3.4.
Figure 3-5. Glycoform identification on glycopeptide T3. (A) ETD-MS/MS (with supplemental activation) generated peptide fragments c and z’ ions. (B) CID-MS/MS generated oxonium ions and glycopeptide fragments. (C) Averaged base peak chromatogram contained all the glycans on glycopeptide T3. m/z 1527.2 is a co-eluting nonglycosylated peptide. NL: normalized level. Key to symbols can be found in Figure 3-4.
3.4.3 Core/Antenna fucosylation isomer identification

Three glycoforms on glycopeptide T1 were found to be fucosylated (Figure 3-4-C, glycoforms 2, 4 and 7). In humans, the major sites of fucosylation of N-glycans are either antenna or core GlcNAc\(^1\). To differentiate the core from antenna isomers, a theoretical fragmentation pattern was first determined to obtain the position-specific fragments. Five Y ions (for core-fucosylation) and 2 B ions (for antenna-fucosylation) (Figure 3-6-A, nomenclature according to Domon and Costello, 1988)\(^{35}\) were calculated to be sufficient to differentiate the position of the fucose attachment\(^{36}\). The CID-MS/MS spectra of glycoforms 2, 4 and 7 of T1 were manually surveyed for these theoretical fragments. The data suggested that glycoforms 2 and 4 are core-fucosylated, while glycoform 7 is antenna fucosylated (Figure 3-6-B). The fucosylated glycoforms on T3 were characterized in a similar manner, with all the 3 fucosylated glycoforms (glycoforms 3, 5 and 7) being found to be antenna-fucosylated (Figure 3-5-C). These antenna fucosylated glycans are likely to be Lewis x type structures\(^{37}\). Exoglycosidase digestion could potentially be performed to determine the fucose linkage\(^{37}\), but with more sample consumption; however, this was not pursued in this study.
Figure 3-6. Fucosylation position (core/antenna) differentiation based on CID-MS/MS. (A) Theoretical position-specific fragments, which can differentiate core/antenna isomers. (B) CID-MS/MS spectra for 3 fucosylated glycoforms. The glycan structures are shown in the inset. Key to symbols can be found in Figure 3-4.
3.4.4 Glycopeptide T2

Glycopeptide T2 with two glycosylation sites eluted earlier than the other two glycopeptides. The observed glycan masses on T2 were also higher than those on T1 and T3, ranging from 4.1 to 5.8 kDa (Figure 3-7). We hypothesized that both sites are fully occupied, because, for the commonly encountered N-linked Hpt glycans, if a single glycosylation site were occupied, the attached glycan will only contribute up to a maximum of 3.8 kDa additional mass to the peptide (assuming the largest glycan is sialylated tetraantennary complex glycan with core fucosylation and bisecting HexNAc). Ten major glycan compositions could be identified on T2, based on the observed m/z and MS/MS fragments, and their potential structures are shown in Figure 3-7. It was not possible to identify the exact glycan isoform structure, i.e. which site contained which specific glycan when the two glycans differed from each other. Moreover, electron transfer dissociation could not be used as it was unable to cleave the peptide backbone between the closely spaced glycosylation sites at N207 and N211.
Figure 3-7. Glycan compositions identified on glycopeptide T2. The empty circle represents charge state +5, while the filled circle represents charge state +4. NL: normalized level. Key to symbols can be found in Figure 3-4.
3.4.5 Glycopeptide isoform quantitation

During positive ESI-MS analysis of glycopeptides, protonation occurs mainly on the peptide moiety; thus, the signal intensity of the glycopeptide is largely dependent on the peptide structure of the molecule. It is therefore reasonable to assume that the response factors of all glycoforms of a specific peptide are similar\(^{15}\). However, the ESI process typically generates multiply charged species from a single precursor\(^{38}\), which will complicate ESI-MS based quantitation. This is schematically shown in Figure 3-8-A. Therefore, peak areas (PAs) summed across all charge state ions were used to determine the relative abundance of a specific glycoform (Figure 3-8-B). The peak area of each individual glycoform was normalized to the total peak area of all glycoforms of the specific glycopeptide (relative quantitation)\(^{39}\).

The quantitation results, based on the 3 initial LC-MS runs (Figure 3-2,panel A), are shown in Figure 3-8-B. The X axis represents the specific glycoform, and the Y axis is the percentage abundance of different isoforms for a given glycopeptide in the glycosylated protein. For glycopeptide T1, the signal intensity ratio between the most abundant (T1-03, Figure 3-4-C) and least abundant glycoform (T1-02, Figure 3-4-C) was ~ 80 (Figure 3-8-B). This result suggests ~ 2 orders of magnitude difference in abundance for these 2 glycoforms or only ~ 100 attomoles of glycoform T1-02 with a 10 fmol sample injection. Importantly, the high sensitivity of the PLOT LC-MS allowed reproducible quantitation at this low level of sample amount. Similar dynamic ranges were observed for glycans on glycopeptide T3 (Figure 3-5-C, T3-02 and T3-08). For glycopeptide T2, the glycoforms appear to be more evenly distributed compared to T1 and T3. The PA results provide a basis for future comparison of glycosylation patterns across different samples, e.g. individual patient plasmas.
For non-small cell lung cancer, sialic acid and fucose expression on the Hpt-β chain has been reported to be elevated, albeit the glycosylation site and glycan structure contributing to the elevation was not obtained\textsuperscript{25}. We have identified partially or fully sialylated glycans with attached fucose, e.g. T1-07, T3-05 and T3-07. Highly branched glycans were also observed in our study (tetra-antennary, T3-08 and T3-09, Table 3-3), and increased branching is typical for cancer\textsuperscript{1}. In addition, core-fucosylated glycan was also observed (T1-02 and T1-04, Table 3-3). Core fucosylated glycoproteins have been reported to be involved in the development of many cancers including lung cancer\textsuperscript{40}. 
Figure 3-8. Glycopeptide quantitation. (A) Charge states result in different peak areas (PAs) for the same precursor; (B) Quantitation results using summed PA method. The glycan structures on glycopeptide T1, T2 and T3 can be found in Table 3-3.
3.4.6 Glycosylation site occupancy

It is important to determine the degree of glycosylation occupancy at a specific site. Because of the limited amount of sample, we used a simpler approach to determine the site occupancy. Automatic database searching was employed to identify nonglycosylated glycopeptides from the Hpt tryptic digest. The three initial data-dependent LC-MS runs, performed for glycopeptide identification (Figure 3-2, panel A), were database searched against the known Hpt sequence using SEQUEST. The nonglycosylated form of T3 could be identified in the three runs, while the nonglycosylated forms of T1 and T2 were not found. Extracted ion chromatograms (corresponding to nonglycosylated T1 and T2) were also examined with no peak observed. This data suggests that T1 and T2 are 100% or close to 100% occupied while T3 is partially occupied. Next, we used PNGase F to remove the N-linked glycans from the tryptic glycopeptides and convert Asn into Asp. In order to confirm complete deglycosylation, the extracted ion chromatogram of the most intense glycoform on glycopeptide T3 (glycoform T3-03, m/z 1334.5, Figure 3-5) was examined: the peak area of glycopeptide T3-03 after PNase F treatment was found to be ~3% of that without PNase F treatment, indicating that glycan removal was close to complete. Since the deglycosylated and nonglycosylated peptides have a small difference in hydrophobicity (Asp vs. Asn), we were able to separate the two peptide forms on the high resolution PLOT LC column, as shown in Figure 3-9. This separation allowed the calculation of the peak area and an estimate of the glycosylation occupancy. The occupancy was calculated to be 88 ± 2% (n = 3), assuming similar ESI-MS responses for the nonglycosylated and deglycosylated peptides. To our knowledge, this is the first report that glycopeptide T3 was partially occupied. Since we used a pooled sample in this work, 88% represents an average value, and it is thus possible that some patients will have even lower occupancy, possibly
dependent on the extent of their disease. In a separate study, the occupancy of glycopeptide T3 of Hpt from pooled normal plasma sample was determined to be 93 ± 5 %.
Figure 3-9. Glycosylation site occupancy determination. Deglycosylated and nonglycosylated peptides T3 were baseline separated using the high-resolution PLOT column. The site occupancy was calculated from the equation shown in the figure.
3.5 Conclusions

A 10 µm i.d. PLOT LC-LTQ-CID/ETD-MS has been successfully implemented for the sensitive site-specific characterization of N-linked protein glycosylation. The open tubular structure of the PLOT LC column led to good glycopeptide recovery and, the long column provided high resolving power. Furthermore and importantly, the ultralow flow rate of 20 nL/min significantly enhanced glycopeptide ionization efficiency. As demonstrated in this work, the coupling of the PLOT column to an LTQ-CID/ETD-MS (CID for glycan, ETD for peptide) should have wide application for trace level glycoprotein characterization. Twenty-six site-specific glycoforms/glycan compositions on 3 glycopeptides of Hpt were characterized and quantitated from 10 LC-MS runs using a sample consumption level of 100 fmol protein digest. Among the 10 LC-MS runs, 3 data-dependent runs were initially employed for glycopeptide identification (ETD) and glycan structure evaluation and quantitation (CID), followed by 4 additional targeted LC-MS/MS analyses (CID-MS/MS) on additional glycoforms for further in-depth glycan structure determination. Finally, three additional runs were conducted for glycosylation occupancy determination.

The PLOT LC-LTQ-CID/ETD-MS method is straightforward, robust, highly resolving and sensitive, allowing the identification of minor glycans from trace amounts of biological sample. For plasma biomarker discovery, the tissue leakage protein concentration ranges from 1ng/mL (or below) to 1 µg/mL. Assuming a M.W. of 50 KDa and a recovery of 50%, from 1 mL plasma sample, 10 fmol to 10 pmol protein will be available for LC-MS analysis. Theoretically, our strategy should be applicable for characterization of middle to high level tissue leakage glycoproteins in 1 mL of plasma. Also, the method should be useful for elucidation of
the structure-function relationship of biologically important molecules such as receptors and the role of glycosylation in their biological function.

Finally, it should be noted that, in this study, the MS spectra were interpreted manually; however, reliable bioinformatics software for glycopeptide identification/annotation will undoubtedly become available in the future. Indeed, significant progress has already been made in this direction. Thus, together with bioinformatics software, the 10 µm i.d. PLOT LC-LTQ-CID/ETD-MS will ultimately allow comprehensive characterization and quantitation of site-specific protein glycosylation from trace amounts of sample in an automatic manner. Candidate glycoforms such as those highly branched in disease can then be selected for further validation using targeted MS analysis, such as selected reaction monitoring.
3.6 References


Chapter 4

Site-Specific Protein Glycosylation Quantification

using Glycopeptide Selected Reaction Monitoring on a

Chip LC-QQQ MS Platform
4.1 Abstract

Selected reaction monitoring (SRM) is a well-established quantitation method; however, its application for site-specific protein glycosylation (glycopeptide) analysis is limited, mainly due to the complexity of the glycan structure and site micro-heterogeneity. In this study, using haptoglobin (Hpt) as a model glycoprotein, we have, to our knowledge for the first time, developed a glycopeptide SRM strategy for sensitive, site-specific quantitation of protein glycosylation isoforms. The glycopeptide fragmentation pattern was first systematically studied using a qToF mass spectrometer to which a microfluidic LC chip was attached. Glycan oxonium ions 204, 274 and 366, were found to be generated consistently for all glycopeptide structures and at higher intensities than glycopeptide fragments (Y ions). These three oxonium ions were selected as MS/MS product ions for glycopeptide SRM measurements using a QQQ mass spectrometer, to which the microfluidic LC chip was again attached. The glycan oxonium ion at m/z 366 was selected for quantitation purposes (most intense), while 204 and 274 were used to check precursor purity. The QQQ MS parameters, including fragmentor voltage (declustering and desolvation), collision energy (CE) and dwell time were systematically evaluated to maximize sensitivity and specificity. A linearity of four orders of magnitude was achieved for the dominant glycosylation isoforms. The method was applied to analyze the site-specific glycosylation differences of Hpt between individual lung cancer, healthy control and inflammation control (rheumatoid arthritis) plasma samples, each consisting of 5 patient plasmas. The results demonstrate the feasibility of using SRM to quantitate specific glycopeptide glycoforms.
4.2 Introduction

Selected reaction monitoring (SRM) has long been a widely used tool for the quantitation of small molecules for pharmacokinetic and clinical analysis and is increasingly becoming an important method for peptide / protein quantitation. Selected reaction monitoring is usually performed using triple quadrupole (QQQ) mass spectrometer. In SRM measurement, a pre-defined precursor is selected in Q1, dissociated in Q2 and fragments monitored in Q3. The signal intensity of precursor to fragment (transition) represents the peptide precursor abundance. Due to the tandem mass selection/filtering, SRM features enhanced signal-to-noise ratio and larger linear dynamic range, compared to non-targeted data-dependent MS measurements. Absolute quantitation is made possible by spiking a known amount of an isotopically-labeled peptide standard in the sample. Relative quantitation is achieved by comparison of the signal intensity of the same transition, usually after normalization, across different samples. Selected reaction monitoring has been applied to study cellular signaling networks, to quantify the full dynamic range of the proteome expressed in *S. Cerevisiae* and to verify human disease biomarker(s) discovered by non-targeted proteomics studies. However, the application of SRM for site-specific protein glycosylation (glycopeptide) quantitation has as yet not been reported.

Two challenges need to be overcome before glycopeptide SRM can become a widely-accepted analytical method for site-specific protein glycosylation quantitation. First, site-specific glycoforms (SRM precursors) for targeted glycoprotein analysis need to be determined at high sensitivity in order to account for low level glycoforms. As detailed in Chapter 3, this can be achieved using our newly developed PLOT LC-LTQ-CID/ETD-MS platform. Second, the fragments (SRM products) of the targeted glycoforms need to be determined in a relatively straightforward manner. Ion trap (for example LTQ) and quadrupole (for example QQQ)
instruments usually produce different fragmentation patterns for the same precursor\(^1\); therefore, the glycopeptide fragments deduced from PLOT-LC-LTQ-CID/ETD-MS cannot be directly used for SRM measurements in the QQQ instrument. In contrast, for peptide SRM method development, the fragments of a specific peptide precursor can generally be retrieved if the peptide has previously been sequenced and deposited in an available database\(^4\), predicted using various software\(^4\) or experimentally determined by infusing the synthetic peptide standards into the mass spectrometer\(^1\). For glycopeptide SRM, none of the above methods is applicable as few glycopeptides have been sequenced, and the fragmentation pattern of glycopeptides has not been studied in detail. Therefore, there is at present no database or software available to retrieve the glycopeptide fragments or to confidently predict glycopeptide fragments. Also, glycopeptide standards are not readily available, preventing the routine practice of direct infusion analysis. Due to these factors, LC-MS analyses need to be performed on the “real” sample using, in our case the qToF instrument to determine suitable fragments (transitions) before actual SRM measurement using the QQQ instrument.

The goal of this work was first to develop a sensitive and rapid glycopeptide SRM method for site-specific quantitation of glycoforms and second to demonstrate the feasibility of the method to quantitate differentially expressed site-specific glycoforms of Hpt between individual lung cancer, matched healthy and inflammation control (rheumatoid arthritis) for a limited number of patients. The ultimate goal would be to establish a sensitive and robust platform for the quantitative analysis of glycosylation isoforms from small amounts of protein sample in clinical specimens. Haptoglobin was again chosen as a model glycoprotein to develop our strategy. We first demonstrate the feasibility of glycopeptide SRM using Chip LC-qToF and Chip LC-QQQ (see Section 4.3.1). We then systematically studied glycopeptide fragmentation
using the Chip LC-qToF with the aim to develop glycopeptide fragmentation rules in quadrupole instrument for easy and rapid glycopeptide SRM method development. Experimental variables, including peptide sequence, glycan structure, charge state and collision energy were sequentially studied in individual experiment to allow detailed evaluation of the influence of the variables on glycopeptide fragmentation. Glycan oxonium ions at m/z 204, 274 and 366 were found, to be generated in a consistent manner and with higher intensities than glycopeptide fragments (Y ions) for all 26 sialylated glycoforms of Hpt under all the examined experimental conditions. These 3 oxonium ions were used as product ions for glycopeptide SRM measurements using the QQQ. The glycopeptide SRM method allowed simultaneous quantitation of all 26 glycoforms plus nonglycosylated glycopeptide T3 and 2 additional Hpt-β peptides (for normalization purpose) in a single 30 min LC-SRM run. The glycopeptide SRM approach features rapid analysis, highly reproducible and sensitive quantitation (10 fmol/injection) and large dynamic range (over 4 orders of magnitude). The method was employed to quantify site-specific glycosylation isoforms of Hpt, affinity purified from individual non-small cell lung cancer patient plasmas and matched healthy and inflammation controls (rheumatoid arthritis, RA). Rheumatoid arthritis is chosen as the inflammation control because it is a long-term autoimmune disease that leads to inflammation of the joints and surrounding tissues. Site-specific glycosylation differences were revealed using the developed glycopeptide SRM. Furthermore, the occupancy on site N\textsuperscript{241} was found to be decreased in lung cancer and arthritis, relative to healthy individuals.

4.3 Experimental

4.3.1 Plasma samples and reagents
Individual plasma samples from 5 non-small cell (squamous) lung cancer patients, 5 rheumatoid arthritis patients and 5 healthy controls (age, gender, ethnicity matched) were obtained from Proteogenex (Culver City, CA). Anti-haptoglobin (Hpt) polyclonal antibody was obtained from Sigma-Aldrich (St. Louis, MO). POROS® Affinity Protein A beads were from Life Technologies (Carlsbad, CA), and bis(sulfosuccinimidyl) suberate (BS3) from Thermo Fisher Scientific (Waltham, MA). Sequencing grade trypsin was obtained from Promega (Madison, WI), phosphate buffered saline (PBS, 10 times concentrated) from MP Biomedicals (Solon, OH), and all other reagents from Sigma-Aldrich.

4.3.2 Haptoglobin purification

Prior to Hpt purification from plasma, albumin and IgG were depleted using anti-albumin particles and POROS protein A beads. Albumin depletion beads (BAC. B.V., The Netherlands) were prepared as a 50% slurry with binding buffer (PBS) and gravity packed into Omnifit glass columns (10 mm × 20 mm) (Biochem Fluidics, Boonton, New Jersey). To prepare the IgG depletion column, a 50% slurry of POROS protein A beads (Life Technologies, Framingham, MA) in PBS was packed into a PEEK column (30 mm × 4.6). Haptoglobin depletion beads were made through crosslinking anti-haptoglobin (Sigma, St. Louis, MO) to POROS protein A beads according to the manufacturer’s protocol. The crosslinked beads were prepared as a 50% slurry with binding buffer (PBS) and packed into a PEEK column (30 mm × 4.6 mm).

A 100 µL volume of plasma was diluted with 100 µL binding buffer (1x PBS), and then the solution was added to the depletion system consisting of the anti-IgG and the anti-albumin columns. The flow through fraction was collected and concentrated down to 200 µL using Amicon 10 kDa molecular weight cut off filters (Millipore, MA). The depleted plasma was then
added to the anti-haptoglobin column (0.5 mL). The bound fraction was eluted using 0.1 M glycine buffer (pH=2.5). The collected fraction was immediately neutralized by adding 200 μL Tris buffer (pH=8). The haptoglobin fraction was desalted using Amicon 10 kDa molecular weight cut off filters (Millipore, MA). The salt concentration was decreased by a factor of 10⁴, and the final volume was 50 μL. To assure the purity of haptoglobin, 9.75 μL of each fraction was subjected to SDS-PAGE. The protein concentration of each haptoglobin fraction was measured using a spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE). The purified Hpt was subjected to standard in-solution tryptic digestion.

4.3.3 Haptoglobin digestion

The affinity purified haptoglobin was denatured by 6 M guanidine-HCl, reduced by 10 mM dithiothreitol (DTT) (30 min at 56 °C) and alkylated by 25 mM iodoacetamide (IAA) (60 min at room temperature in the dark). The reduced and alkylated protein solution was buffer exchanged to 50 mM ammonium bicarbonate using Microcon ultracentrifuge filter (5 kDa MWCO, Millipore, Billerica, MA) for 5 cycles (10 000× g for 5 min per cycle). Trypsin was added to the protein solution at ~ 1:25 ratio (enzyme to protein) and incubated for 6 hrs. The enzymatic reaction was ended with 5% formic acid, and the digest was stored at -80 °C until analysis.

4.3.4 HPLC-Chip-MS system

Liquid chromatography was performed using an Agilent series 1200 instrument (Agilent Technologies, Santa Clara, CA) consisting nanopump, capillary pump, microwell-plate autosampler, and HPLC-Chip-MS interface. An Agilent HPLC-Chip was used for both the LC-
qToF and LC-QQQ experiments. The HPLC Chip (Agilent Technologies, Santa Clara, CA) includes a 40-nL enrichment column (5 μm ZORBAX 300 SB-C18), a 50 μm (d)×75 μm (w)×43 mm (l) RPLC separation column (5 μm ZORBAX 300 SB-C18) and a nanospray emitter (50 μm orifice). The HPLC-Chip-MS interface (chip cube) mounts the HPLC-chip online to qToF or QQQ mass spectrometers. A miniature camera is included in the chip cube to visualize the spray during LC-MS analysis to ensure high and stable MS signal.

Mobile phase A was 0.1% (v/v) formic acid in water and mobile phase B 0.1% (v/v) formic acid in acetonitrile. The mobile phase gradient consisted of the following steps: (1) linear from 2-10% B over 2 mins; (2) linear from 10 - 26% B over 16 mins; (3) linear from 26-80% B in 2 mins; and then held constant for 4 min to rinse the column. Finally, the starting condition was restored over 0.1 min, and the column re-equilibrated for 6 min. The flow rate was set at 300 nL/ min. Enrichment of the analytes was performed by pumping 0% B (capillary pump) at 4 μL/min from the autosampler to the enrichment column of the HPLC Chip. An 0.5-μL aliquot of digested sample was loaded. Agilent MassHunter software (version B.01.03) was used for data acquisition and processing.

4.3.5 qToF-MS/MS

An Agilent 6530 qToF-MS was used to characterize glycopeptide fragmentation. All 26 site-specific Hpt glycoforms were subjected to targeted qToF-MS/MS analysis. The time for MS spectra collection was 200 ms (2, 000 transients), and the time for MS/MS spectrum collection ranged from 200 to 1000 ms (2, 000 to 10, 000 transients), depending on the signal intensity of the glycopeptide precursor. The analyses were performed in the positive ionization mode with a capillary voltage set at 1,800 V and a delta electron multiplier voltage of 200 V. The drying gas
flow was 5 L/min of nitrogen, at a temperature of 325°C. The collision energy ranged from 5 to 45 V, as detailed below (Section 4.3.2).

**4.3.6 QQQ-MS/MS**

An Agilent 6460 QQQ was used for the glycopeptide SRM experiment. Nonglycosylated T3 (m/z = 599.2, z = 3) and two Hpt-β-chain peptides SCAVAEYGVYVK (m/z = 673.8, z = 2) and VTSIQDWVQK (m/z = 602.7, z = 2) were also monitored using peptide SRM. Their transitions were determined using targeted qToF-MS/MS analysis (599.2→658.4, 599.2→757.5, 599.2→1346.7 for nonglycosylated T3; 673.8→319.1, 673.8→928.5, 673.8→1099.3 for SCAVAEYGVYVK; 602.7→675.4, 602.7→803.4, 602.7→1003.5 for VTSIQDWVQK) and used without further optimization during SRM measurement. The key instrument parameters for glycopeptide SRM measurement, including fragmentor voltage, collision energy and dwell time, were systematically optimized for each precursor glycopeptides ion in the QQQ and are described in detail in the Results and Discussion.
4.4 Results and Discussion

The goal of this study was to develop a glycopeptide SRM method for site-specific protein glycosylation quantitation. A total of 26 site-specific glycoforms were identified from three Hpt glycopeptides (7 for T1, 10 for T2 and 9 for T3, Table 4-1) using an ion trap mass spectrometer, as described in Chapter 3. In this work, we employed two quadrupole instruments (qToF and QQQ) to develop our glycopeptide SRM strategy. The T1, T2 and T3 glycopeptides differ in peptide sequence, glycan structure (bi- to tetra-antennary, partially/fully sialylated, core/antenna fucosylated) and number of glycosylation sites per glycopeptide (1 site for glycopeptide T1 and T3, 2 sites for T2). All 26 glycoforms were subjected to glycopeptide SRM measurements in one LC-MS run. A total of 15 plasma samples (non-small cell lung cancer, healthy control and inflammation control, 5 individuals each) were subjected to Hpt affinity purification. We combined the individual affinity purified Hpt on an equal amount basis to form a pooled sample. This sample was then used as the standard to develop the glycopeptide SRM method.

4.4.1 Glycopeptide SRM development strategy

To our knowledge, SRM analysis of glycopeptides has not been reported before. Due to this fact, we performed an initial study to ascertain the potential of glycopeptide SRM analysis. An Agilent 6530 qToF-MS instrument, which shares the “same” hardware design with an Agilent 6460 QQQ instrument (up to the collision cell), was used to identify the glycopeptide fragments (SRM products). We chose the qToF MS for this purpose is because the qToF is more sensitive than the QQQ for determining the full product ion spectrum. As an example, glycoform T3-02 (m/z = 1001.5, z = 4) was subjected to MS/MS activation with both the qToF and QQQ at
the same collision energy of 15 V, Figure 4-1. Similar MS/MS spectral patterns can be observed for both the qToF and QQQ, with the qToF having a 40 fold increase in signal compared to that of the QQQ. The resulting MS/MS fragments with the qToF include glycan oxonium ions and glycopeptide fragments (Y ions), and these results can be directly transferred to the QQQ for SRM measurements. One example is shown in Figure 4-2, where the six most intense qToF-MS/MS fragments (3 oxonium ions and 3 Y ions) from glycoform T3-04 (m/z = 1092.8, z=4) were monitored in the QQQ under the SRM mode. The glycopeptide SRM signal is obtained with a similar intensity ratio as that from the qToF. It is important to note that the same HPLC-chip was used for both the LC-qToF and LC-QQQ analyses, which facilitated method transfer between the two platforms. Glycopeptide fragmentation was then systematically evaluated using qToF-MS/MS, targeting the 26 Hpt site-specific glycoforms. This systematic evaluation defines glycopeptide fragmentation in a quadrupole instrument, an important step for rapid glycopeptide SRM method development (see below).
Figure 4-1. qToF and QQQ result in similar MS/MS fragmentation patterns for glycopeptides. Glycoform T3-02 (m/z = 1001.5, z = 4, structure in the inset) is shown as an example. The qToF is operated in the targeted MS/MS mode, while QQQ in the product ion scan mode. Glycan oxonium ions (274, 366, etc.) are more intense than glycopeptide fragments in both experiments. Note also that, in the QQQ, the 366 glycan oxonium ion is more intense than 274, while in the qToF, 274 is greater than 366. Key to symbols: empty circle = galactose, filled circle = mannose, filled square = HexNAc, filled diamond = NeuAc, filled triangle = fucose.
Figure 4-2. Glycopeptide SRM. Glycoform T3-04 (m/z = 1092.8, z = 4, structure in the inset) was subjected to targeted MS/MS analysis on the qToF. Six intense fragment ions (3 glycan oxonium ions and 3 glycopeptide ions, labeled with yellow squares) from the resulting qToF-MS/MS spectrum were monitored under the SRM mode using the QQQ. Note in this example, glycan oxonium ion 366 is more intense than 274 in both the qToF and QQQ instruments. PA: peak area. Key to symbols can be found in Figure 4-1.
4.4.2 Glycopeptide fragmentation pattern in quadrupole mass spectrometers

The influences of 1) precursor charge state, 2) glycan structure and 3) peptide sequence on glycopeptide fragmentation were systematically studied. The collision energy (CE) influence on glycopeptide fragmentation was also studied by increasing the CE for a given glycopeptide. The resulting qToF-MS/MS spectra are shown in Figures 4-3 to 4-6. Several general principles on glycopeptide fragmentation in the quadrupole instrument can be drawn from these experiments, with some of them already noted by others5.

1) Due to the weak glycosidic linkage, the default CE on the qToF instrument was too high for glycopeptide activation. The default CE value is calculated based on the following equation: \( CE (V) = 3.6 /100 \text{ Da} + 2.5 \). For glycoform T1-02 (m/z = 1186, z = 4), the instrument default collision energy is calculated to be 32V. Under this default value, the observable MS/MS fragments for glycopeptide are glycan oxonium ions with Y ions absent (Figure 4-3, CE ≥ 30 V). The commonly observed oxonium ions include 204 (HexNAc), 274/292 (NeuAc-H2O/NeuAc), 366 (Hex-HexNAc) and 657 (NeuAc-Hex-HexNAc) (Figures 4-3 to 4-6).

2) For all the examined glycoforms, the CE value, for which reasonable intensity Y ions are obtained, ranges between 10 and 15 V (Figures 4-3 to 4-6). The data (Figure 4-5) also indicate, for the same charge state, the precursor m/z of the glycopeptide does not have a significant influence on this CE value. For example, glycoforms T1-01 and T1-06 (both charge state 5), which have an m/z difference over 200, share the same optimum CE value of 15 V for Y ion generation (Figure 4-5). This result, again, reflects the fact that the glycosidic linkages are generally weaker than peptide bonds, and a CE of 30 V was sufficient to break the glycosidic linkage.
3) For quadrupole CID-MS/MS, glycan oxonium ion intensities are found to be stronger than the glycopeptide fragments (Y ions) (Figures 4-3 to 4-6). Changes in peptide sequence or glycan structure do not alter this trend (Figures 4-5 to 4-6). Furthermore, higher CE values are preferred for glycan oxonium ion generation. For example, at a CE value larger than 30 V, Y ions totally disappear (Figure 4-3); however, oxonium ions intensities are higher than that at CE 15 V (Figure 4-3). This result can be contrasted to ion trap CID-MS/MS, where Y ions are typically more intense than oxonium ions (see Chapter 3). The different mechanisms between ion trap-CID and quadrupole-CID cause this result. In ion trap-CID, once the glycopeptide is fragmented, the resulting fragments will be off the resonance frequency, and no further fragmentation will occur, while for quadrupole-CID, the resulting fragments can be still accelerated by the electric field and collide with the neutral gas to induce further fragmentation.

4) Charge state of the precursor ion has an impact on glycopeptide fragmentation. At a given CE value, the glycopeptide with the higher charge state undergoes more extensive fragmentation relative to a lower charge state ion (Figure 4-4). Similar phenomena have been observed for peptide fragmentation.1
Figure 4-3. The effect of collision energy on glycopeptide fragmentation. Glycoform T1-03 (m/z = 978, z = 5), shown as an example, was subjected to targeted qToF-MS/MS analysis with increasing CE at 5, 10, 15, 30 and 45 V. Note that with increasing CE, the intact precursor intensity decreases. Above CE 15 V, few glycopeptide fragments remain, and at CE 30 and 45 V, as expected, only oxonium ions remain detectable. Key to symbols can be found in Figure 4-1.
Figure 4-4. The effect of charge state on glycopeptide fragmentation. Glycoform T1-03 is shown as an example. Two charge state ions (m/z = 1222, z = 4; and m/z = 978, z = 5) were subjected to targeted qToF-MS/MS with CE at 15 V. Note that under the same CE, at a charge state 4, a large amount of the precursor remains intact, while for charge state 5, most of the precursor undergoes fragmentation. Key to symbols can be found in Figure 4-1.
Figure 4-5. The effect of glycan structure on glycopeptide fragmentation. Glycoform T1-01, T1-03, T1-05 and T1-06 (all charge state 5, structures shown in the inset) were subjected to targeted qToF-MS/MS analysis. For these glycoforms, the optimal CE is 15 V. Key to symbols can be found in Figure 4-1.
Figure 4-6. The effect of peptide sequence on glycopeptide fragmentation. Glycoforms T1-03 (m/z = 978, z = 5), T3-02 (m/z = 800, z = 5) and T2-02 (m/z = 1174, z = 5) were subjected to targeted qToF-MS/MS with CE at 15 V. The three glycoforms share the same glycan structures (biantennary complex type) but differ in peptide sequence. Key to symbols can be found in Figure 4-1.
4.4.3 Transitions for glycopeptide and peptide SRM

4.4.3.1 Transitions for glycopeptide SRM

Based on the qToF analysis, the mass and m/z values of targeted glycopeptide precursors are listed in Table 4-1. The elution time windows for the glycopeptides T1, T2 and T3 are 15.7-16.7, 5.6-7.2 and 13-14 min, respectively. Within an individual glycopeptide elution time window, all the glycoform m/z values can be differentiated by the QQQ (Q1 at unit resolution), even for the closely related T2-02 (tetra-sialylated, m/z = 1469, z = 4) and T2-10 (hex-sialylated, m/z = 1467, z = 5). A separate study indicated that T2-10 can be separated from T2-02 by ~ 15 seconds using the LC gradient described in the Experimental Section; therefore, these two glycoforms will not interfere with each other. Under current separation conditions (0.1% formic acid in both mobile phases), sialic acids are substantially protonated. Glycopeptide T2-10 which has six sialic acids will have a higher hydrophobicity, compared to T2-02 with four sialic acids. Thus, it would be expected to elute at a later time. Similar results have been observed by others.

On the one hand, relative to peptide SRM, glycopeptides add complexity to SRM assay development (difficult to determine the precursor); however, on the other hand, there are common preferred fragmentation tags (oxonium ions), which in fact can simplify glycopeptide SRM studies. Among the commonly observed glycan oxonium ions, 204, 274 and 366 were found in this work to be present at higher intensities than other oxonium ions and were therefore examined further. Glycan oxonium ion 204 is derived from HexNAc, 274 from sialic acid with the loss of one water molecule and 366 from Hex-HexNAc (see Figures 4-2). Table 4-2 lists the theoretical amino acid compositions of peptide fragments, which will have the same nominal mass as these 3 oxonium ions. There are only 4, 5 and 8 amino acid compositions with nominal masses at 204, 274 and 366, respectively. As a result of the low frequency of interfering ions
from peptide fragments, these oxonium ions have been successfully used by others to identify glycopeptides from unseparated peptides/glycopeptides mixtures.

In this work, the targeted glycoprotein Hpt was affinity purified from plasma. Due to the high level of Hpt in plasma and the subsequent IP pull down, we were able to target the above 3 oxonium ions, along with the retention time and precursor masses, to specifically identify and quantitate the targeted glycoforms. The advantage of targeting only oxonium ions and not Y ions is that labor-intensive, sample-consuming MS/MS experiments need not to be performed for candidate glycopeptides SRM measurements. On the other hand, for more complex mixtures, such as whole cell lysate digests, glycopeptide fragments (Y ions) will likely need to be included to assess the relative purity of the precursor ion.

Using the 3 glycan oxonium ions as SRM products, for all the 78 glycopeptide transitions (26 glycopeptide precursors, 3 transitions for each precursor), only transition 1383.7 → 366 (T3-03) has an interfering peak (Figure 4-7-A, red arrow). This peak was identified as interference because there are no peaks present at the other two SRM transitions (204 and 274). For all other glycoforms, there is no observed interfering peak. Glycopeptide SRM measurement of glycoform T3-04 is shown in Figure 4-7-A as an example.

4.4.3.2 Transitions for peptide SRM

Nonglycosylated T3 and two Hpt-β peptides were also monitored using peptide SRM. The product ions for peptide SRM (3 product ions for each precursor) were determined using targeted qToF-MS/MS. The SRM signal of nonglycosylated T3 is shown in Figure 4-7-B as an example. The 2 Hpt-β peptides SRM signals were used for normalization purpose (see below).
Table 4-1. The 26 glycoform masses and m/z values (z=3, 4, 5) for the Hpt samples

<table>
<thead>
<tr>
<th>Glycoform</th>
<th>Glycan structure</th>
<th>average</th>
<th>m/z (average)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>z=3</td>
</tr>
<tr>
<td>T1-01</td>
<td>2-1S-0F</td>
<td>4595.8</td>
<td>1533</td>
</tr>
<tr>
<td>T1-02</td>
<td>2-1S-1F</td>
<td>4741.9</td>
<td>1582</td>
</tr>
<tr>
<td>T1-03</td>
<td>2-2S-0F</td>
<td>4887.0</td>
<td>1630</td>
</tr>
<tr>
<td>T1-04</td>
<td>2-2S-1F</td>
<td>5033.2</td>
<td>1679</td>
</tr>
<tr>
<td>T1-05</td>
<td>3-2S-0F</td>
<td>5252.4</td>
<td>1752</td>
</tr>
<tr>
<td>T1-06</td>
<td>3-3S-0F</td>
<td>5543.6</td>
<td>1849</td>
</tr>
<tr>
<td>T1-07</td>
<td>3-3S-1F</td>
<td>5689.8</td>
<td>1898</td>
</tr>
<tr>
<td>T2-01</td>
<td>(2-2S-0F)+(2-1S-0F)</td>
<td>5580.3</td>
<td>1861</td>
</tr>
<tr>
<td>T2-02</td>
<td>(2-2S-0F)+(2-2S-0F)</td>
<td>5871.6</td>
<td>1958</td>
</tr>
<tr>
<td>T2-03</td>
<td>(2-2S-0F)+(3-2S-0F)</td>
<td>6236.9</td>
<td>2080</td>
</tr>
<tr>
<td>T2-04</td>
<td>(2-2S-0F)+(3-2S-1F)</td>
<td>6383.0</td>
<td>2129</td>
</tr>
<tr>
<td>T2-05</td>
<td>(2-2S-0F)+(3-3S-0F)</td>
<td>6528.1</td>
<td>2177</td>
</tr>
<tr>
<td>T2-06</td>
<td>(2-2S-0F)+(3-3S-1F)</td>
<td>6674.3</td>
<td>2226</td>
</tr>
<tr>
<td>T2-07</td>
<td>(2-2S-0F)+(4-3S-0F)</td>
<td>6893.5</td>
<td>2299</td>
</tr>
<tr>
<td>T2-08</td>
<td>(2-2S-0F)+(4-3S-1F)</td>
<td>7039.6</td>
<td>2348</td>
</tr>
<tr>
<td>T2-09</td>
<td>(2-2S-0F)+(4-4S-0F)</td>
<td>7184.7</td>
<td>2396</td>
</tr>
<tr>
<td>T2-10</td>
<td>(2-2S-0F)+(4-4S-1F)</td>
<td>7330.9</td>
<td>2445</td>
</tr>
<tr>
<td>T3-01</td>
<td>2-1S-0F</td>
<td>3710.8</td>
<td>1238</td>
</tr>
<tr>
<td>T3-02</td>
<td>2-2S-0F</td>
<td>4002.1</td>
<td>1335</td>
</tr>
<tr>
<td>T3-03</td>
<td>2-2S-1F</td>
<td>4148.2</td>
<td>1384</td>
</tr>
<tr>
<td>T3-04</td>
<td>3-2S-0F</td>
<td>4367.4</td>
<td>1457</td>
</tr>
<tr>
<td>T3-05</td>
<td>3-2S-1F</td>
<td>4513.6</td>
<td>1506</td>
</tr>
<tr>
<td>T3-06</td>
<td>3-3S-0F</td>
<td>4658.7</td>
<td>1554</td>
</tr>
<tr>
<td>T3-07</td>
<td>3-3S-1F</td>
<td>4804.8</td>
<td>1603</td>
</tr>
<tr>
<td>T3-08</td>
<td>4-2S-0F</td>
<td>4732.7</td>
<td>1579</td>
</tr>
<tr>
<td>T3-09</td>
<td>4-3S-0F</td>
<td>5024.0</td>
<td>1676</td>
</tr>
</tbody>
</table>
Table 4-2. Amino acid compositions with nominal masses of 204, 274 and 366*.

<table>
<thead>
<tr>
<th>HexNAc oxonium ion</th>
<th>204.084</th>
</tr>
</thead>
<tbody>
<tr>
<td>y₂ AN</td>
<td>204.098</td>
</tr>
<tr>
<td>y₂ GQ</td>
<td>204.098</td>
</tr>
<tr>
<td>y₂ GK</td>
<td>204.135</td>
</tr>
<tr>
<td>y₃ AGG</td>
<td>204.098</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NeuAc oxonium ion</th>
<th>274.065</th>
</tr>
</thead>
<tbody>
<tr>
<td>y₂ RV</td>
<td>274.187</td>
</tr>
<tr>
<td>y₃ GPT</td>
<td>274.139</td>
</tr>
<tr>
<td>y₃ APS</td>
<td>274.139</td>
</tr>
<tr>
<td>y₃ GVV</td>
<td>274.176</td>
</tr>
<tr>
<td>y₃ AAI(L)</td>
<td>274.176</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hex-HexNAc oxonium ion</th>
<th>366.139</th>
</tr>
</thead>
<tbody>
<tr>
<td>y₃ EMS</td>
<td>366.133</td>
</tr>
<tr>
<td>y₃ DMT</td>
<td>366.133</td>
</tr>
<tr>
<td>y₃ AEF</td>
<td>366.166</td>
</tr>
<tr>
<td>y₃ PSY</td>
<td>366.166</td>
</tr>
<tr>
<td>y₃ AI(L)Y</td>
<td>366.203</td>
</tr>
<tr>
<td>y₃ FI(L)S</td>
<td>366.203</td>
</tr>
<tr>
<td>y₃ FTV</td>
<td>366.203</td>
</tr>
<tr>
<td>y₃ HI(L)P</td>
<td>366.214</td>
</tr>
</tbody>
</table>

*Note only y-type ions are listed, since, for peptide fragmentation, these ions are the major type observed in a quadrupole instrument [Vinzenz Lange 2008]. Also, it should be noted that the listed item represents amino acid composition, but not sequence. For example, y₂ AN represents either “y₂ AN” or “y₂ NA”.

186
Figure 4-7. Glycopeptide and peptide SRM. (A) Glycopeptide SRM for glycoforms T3-03 and T3-04. For each precursor, 3 product ions (oxonium ions 204, 274 and 366) are monitored. Any interfering peak can be easily identified by correlating the signals in all three SRM channels and excluded from quantitation (see the red arrow for the 366 oxonium fragment ion of T3-03). (B) Peptide SRM for nonglycosylated peptide T3 (VVLHPNYSQVDIGLIK, m/z = 599.2, z = 3). The 3 product ions were determined using targeted qToF-MS/MS analysis. The product ions are $y_6$ DIGLIK (658.4, z = 1), $y_7$ VDIGLIK (757.5, z = 1), and $y_{12}$ PNYSQVDIGLIK (1346.7, z = 1), respectively.
4.4.4 Glycopeptide SRM optimization

Table 4-3 presents the full Hpt glycopeptide SRM method. The LC gradient was divided into 4 time segments, with glycopeptides T2, T3 (together with nonglycosylated T3) and T1 eluting in segments 1, 3 and 4, respectively, and 2 Hpt-β-chain peptides eluting in segment 2. The two Hpt-β-chain peptides were monitored to normalize glycopeptide SRM signals, in order to minimize the variation caused by sample injection, instrument fluctuation, etc. The normalization was performed using the equation: $P_{\text{Anormalization}} = \frac{P_{\text{Aglycopeptide}}}{P_{\text{Apeptide}}} \times 1000$. For all transitions, the first quadrupole was set at unit resolution (approximately ± 0.5 Da) and the third quadrupole at approximately ±1 Da. The QQQ instrument parameters, including fragmentor voltage, collision energy and dwell time, were investigated in detail to achieve high selectivity and sensitivity (see below). The fragmentor voltage was set at 160 V for all glycopeptides and 135 V for peptides (default value). The CE was 30 V for all glycopeptides and 15 V for peptides, and the dwell time was 100 ms for all transitions. The transitions used for quantitation are highlighted in red, as shown in Table 4-3. For glycopeptide SRM optimization studies, 10 fmol Hpt digest from the pooled Hpt sample was used for each LC-SRM analysis.
### Table 4-3. Glycopeptide SRM with all the monitored transitions and key instrument parameters

<table>
<thead>
<tr>
<th>Segment</th>
<th>Precursor ID</th>
<th>Precursor Ion (MS1)</th>
<th>Product Ions (MS3)</th>
<th>Fragmentor (V)</th>
<th>Collision Energy</th>
<th>cycle time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (0-8 min)</td>
<td>T2-01 (+4)</td>
<td>1396.1</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td>3105</td>
</tr>
<tr>
<td></td>
<td>T2-02 (+4)</td>
<td>1468.9</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2-03 (+4)</td>
<td>1560.2</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2-04 (+4)</td>
<td>1596.8</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2-05 (+4)</td>
<td>1633.1</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2-06 (+4)</td>
<td>1669.6</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2-07 (+4)</td>
<td>1724.4</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2-08 (+4)</td>
<td>1760.9</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2-09 (+4)</td>
<td>1797.2</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2-10 (+4)</td>
<td>1833.7</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>2 (8-11.5 min)</td>
<td>IS Peptide 1 (+2)</td>
<td>673.8</td>
<td>319.1/1099.3/928.5</td>
<td>135</td>
<td>15</td>
<td>621</td>
</tr>
<tr>
<td></td>
<td>IS Peptide 2 (+2)</td>
<td>602.7</td>
<td>1003.5/803.4/675.4</td>
<td>135</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>3 (11.5-15.5 min)</td>
<td>T3-01 (+3)</td>
<td>1237.9</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td>3105</td>
</tr>
<tr>
<td></td>
<td>T3-02 (+3)</td>
<td>1335.1</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3-03 (+3)</td>
<td>1383.7</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3-04 (+3)</td>
<td>1456.8</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3-05 (+3)</td>
<td>1505.5</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3-06 (+3)</td>
<td>1553.9</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3-07 (+3)</td>
<td>1602.6</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3-08 (+3)</td>
<td>1578.6</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3-09 (+3)</td>
<td>1675.7</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nonglycosylated T3 (+3)</td>
<td>599.2</td>
<td>1346.7/757.5/658.4</td>
<td>160</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>4 (15.5-20 min)</td>
<td>T1-01 (+4)</td>
<td>1149.9</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td>2173</td>
</tr>
<tr>
<td></td>
<td>T1-02 (+4)</td>
<td>1186.5</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T1-03 (+4)</td>
<td>1222.8</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T1-04 (+4)</td>
<td>1259.3</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T1-05 (+4)</td>
<td>1314.1</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T1-06 (+4)</td>
<td>1386.9</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T1-07 (+4)</td>
<td>1423.4</td>
<td>366.0/274.0/204.0</td>
<td>160</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
4.4.4.1 Collision energy

Collision energy is one of the most important parameters for SRM method optimization. For a fixed precursor ion, the fragmentation will increase with the amplification of collision energy, however, at too high collision energy, the generated fragments could undergo secondary fragmentation and the resulting SRM signal may drop\(^1\). For SRM of peptides, the optimal collision energy is approximately linearly correlated with the precursor m/z; nevertheless, many peptide precursor ions deviate considerably from the predicted value\(^1\). For glycopeptides, the influence of CE on fragmentation was first studied from 10 to 50 V, in 10 V intervals. A CE of 30 to 40 V was found to generate strong oxonium ions for all the 3 glycopeptides. This result also correlates well with qToF data (see Figure 4-3). An example of glycoform T1-03 is shown in Figure 4-8-A. Among the 3 monitored oxonium ions, the one at 366 was chosen as the preferred ion for quantitation due to its generally highest intensity. To fine tune the CE voltage for each individual glycopeptide precursor ion, a further study was carried out by increasing CE from 30 to 40 V, in 1 V increments for each experiment. A linear relationship was not observed between the optimal CE voltage and the glycopeptide m/z(s). Thus, for the 26 glycoforms, a higher mass of the glycopeptide precursor ion did not require a higher CE voltage to achieve a stronger SRM signal. This was a result of the fact that the glycosidic linkages are generally weaker than peptide bonds, and a CE of 30 V was sufficient to break the glycosidic linkage. Glycan oxonium ions 204 and 274 were monitored for qualification/specificity purposes.
Figure 4-8. Collision energy (CE) influence on oxonium ion intensity. For all 26 glycoforms from the Hpt glycopeptides, the optimal CE value ranged between 30 and 40 V. Glycoform T1-03 is shown as an example. Glycan oxonium ion 366 is the preferred product ion for quantitation because of its generally highest intensity.
4.4.4.2 Fragmentor voltage

Fragmentor voltage is the voltage used for desolvation / declustering of electrospray droplets. This voltage is applied on the ion transfer capillary. If fragmentor voltage is too low, the generated ESI droplets may not be sufficiently desolvated / declustered, resulting in decreased sensitivity. On the other hand, if a fragmentor voltage is too high, the precursors could be fragmented in the ESI source, resulting in in-source dissociation. This early fragmentation must obviously be avoided, in order to achieve correct quantitation. It has been found that, for SRM of peptides, the fragmentor voltage usually displays a broad optimum, and optimization on individual peptide would not significantly enhance SRM signal. For the glycopeptides studied here, a similar trend was observed (Figure 4-9). For example, the signal for all 9 glycoforms on glycopeptide T3 (glycan structures ranging from bi- to tetra-antennary and fully or partially sialylated) remained stable in intensity from 100 V to 240 V (Figure 4-9). Reproducible quantitative results were achieved, as evidenced by the small relative standard deviation (7-16 %, triplicates). The SRM signal decreased at fragmentor voltage 280, probably due to glycopeptide in-source fragmentation. For the SRM of the Hpt glycopeptides, a fragmentor voltage of 160 V was selected.
Figure 4-9. The influence of fragmentor voltage on glycopeptide SRM. Fragmentor voltage was found to have a wide optimal value range for all 26 Hpt glycoforms (glycopeptide T3 shown here). For the glycopeptide SRM method, a fragmentor voltage of 160 V was selected for all 3 glycopeptides. The instrument default value was 135 V. Error bars indicate the standard deviation from 3 repeated injections.
4.4.4.3 Dwell time

Dwell time is the time (millisecond scale) the QQQ instrument spent on a specific transition to accumulate the SRM signal\(^1\). To achieve high sensitivity, the dwell time must be sufficiently long to accumulate signal. However, for SRM measurement, usually many transitions are monitored by QQQ in a sequential and cycled manner. At too long dwell time, few data points will be sampled over the chromatographic profile of the analyte, and the peak cannot be correctly reconstructed with too few data points (for example, < 7 points)\(^1\).

The influence of dwell time on SRM signal was studied by increasing the dwell time from 10 to 500 ms. Stable SRM signals could only be observed when the dwell time was equal to or longer than 50 or 100 ms. When the dwell time was less than 50 ms, electronic noise frequently resulted. Longer dwell times did not increase the peak area since the signal was normalized (counts per second). A dwell time of 100 ms was chosen for all the transitions, and at this value, the longest cycle time for the SRM method was ~3 seconds (Table 4-3). The chromatographic peak width was 1 min (see Figure 4-7), which meant that approximately 20 data points per peak could be collected, a number sufficient to define the peak shape.

4.4.4.4 Charge state

The charge states with the most intense ions for the 3 glycopeptides were first screened using the Agilent Chip LC-qToF-MS. The most intense charge states for glycopeptides T1, T2 and T3 were found to be +4, +4, and +3, respectively. The same result can be observed in the Chip LC-QQQ, as shown in Figure 4-10. The charge state of a precursor can be influenced by the spray voltage, flow rate etc\(^1\). The use of the same HPLC Chip across the qToF and QQQ
instruments facilitates method transfer and assures that the relative intensities of individual charge states remain similar on both instruments.
Figure 4-10. Most intense charge state determination. The most intense charge states for glycopeptides T1, T2 and T3 were +4, +4 and +3, respectively. Glycoforms from T1 and T2 are shown here as an example. Error bars indicate the standard deviation from 3 repeated injections.
4.4.4.5 Linearity

For the intense glycoforms T2-02 and T3-02, at least four orders of magnitude linear dynamic range was achieved using the developed glycopeptide SRM method (Figure 4-11). The peak areas of glycoform T2-02 and T3-02 were linear from 50 attomole to at least 500 fmol with \( r^2 \) greater than 0.99 (0.998 for glycoform T2-02, 0.992 for glycoform T3-02). Even in the low concentration range (50 attomole, 500 attomole and 5 fmol), the linearity was still acceptable (\( r^2 = 0.985 \) for T2-02, 0.965 for T3-02), as shown in the inset in Figure 4-11.

4.5 Analysis of individual patient plasma Hpt by glycopeptide SRM

Haptoglobin from individual non-small cell lung cancer patient, matched healthy control and arthritis patient plasmas was subjected to glycopeptide SRM quantitation in order to examine further the feasibility of the SRM method. Ten fmol Hpt tryptic digest was used for each LC-SRM analysis, with each sample run in triplicate. Glycopeptide and nonglycosylated T3 SRM signal were normalized by 2 Hpt-β-chain peptides (VVLHPNYSQVDIGLIK and VTSIQDWVQK, see Table 4-3). Both peptides gave similar normalization patterns, and the results shown below were normalized by the peptide VTSIQDWVQK (602.7 → 1003.5). Figure 4-12 presents the results for glycopeptides T1 and T3. Several interesting aspects on the site-specific Hpt glycosylation can be summarized as follows:

1) A similar site-specific glycosylation pattern was shared by all 15 individuals. For example, the strongest signal on glycopeptide T1 resulted from glycoform T1-03, followed by T1-01. This trend remained the same for T3, for all the 15 individual samples.

2) Generally, larger individual variations were observed in disease (both lung cancer and arthritis), as indicated by the error bars (individual variations) in Figure 4-12. Examples
include glycoforms T1-05, T1-06 and T1-07, T2-01, T2-02 and T2-03, T3-01, T3-04 and T3-05. The technical variation ranged between 7-20% (triplicates).

3) Highly branched glycoforms T1-05 and T1-06 in lung cancer and arthritis gave similar SRM signal intensities (abundance) but lower than that for healthy controls. This result may be caused by the incomplete processing of complex glycan in the glycosylation machinery. Haptoglobin is a high abundant plasma protein and its concentration could increase by 3 fold or more in cancer or arthritis. This dramatic increase of Hpt polypeptide production (by ribosome) during disease may exceed the working capacity of the glycosylation machinery (Golgi); therefore, the glycans presented on Hpt may be incompletely processed. It is also possible that the Hpt was decorated with highly branched glycans, but the released catalytic domain of glycosidases actually cleaved the highly branched glycans from Hpt during circulation. Therefore, less abundance of highly branched glycans could be observed, as seen in this study.

4) Glycoforms T3-03, T3-05 and T3-07 were all antenna-fucosylated. These glycoforms were also present in higher abundance in healthy than in lung cancer and arthritic patients. This seems to suggest that decreased antenna fucosylation at this particular site. While increased fucosylation has been reported for various cancers, the reasons for decreased antenna fucosylation on this particular site, if the result holds up with the study of a larger number of patients, would be worth further investigation.

5) Glycosylation occupancy of site N^241 is decreased in both lung cancer and arthritis. Site occupancy on transferrin has been reported to decrease in cogenital disorder of glycosylation.
Figure 4-11. Four orders magnitude of linearity (50 amol to 500 fmol) is achieved using the developed glycopeptide SRM. A) Glycoform T2-02, structure in the inset; B) Glycoform T3-02, structure in the inset.
Figure 4-12. Site-specific glycoforms quantitation by SRM on individual Hpt samples.
4.6 Conclusions

A sensitive and rapid glycopeptide SRM method has been, for the first time, successfully developed for site-specific glycoform quantitation of an affinity purified glycoprotein. Using Hpt as a proof-of-concept model glycoprotein (26 site-specific glycoforms), the glycopeptide fragmentation pattern was systematically studied in a quadrupole instrument using a Chip LC-qToF and Chip LC-QQQ. The qToF and QQQ (both from Agilent) were shown to produce similar glycopeptide fragmentation patterns. In contrast to an ion trap instrument, glycan oxonium ions are found to be generated at higher intensities than Y ions. Three glycan oxonium ions were used as SRM product ions for glycopeptide SRM quantitation.

The SRM method featured high sensitivity (10 fmol / injection), high reproducibility and relatively large dynamic range (4 orders of magnitude). A total of 26 site-specific glycoforms and 3 peptides (nonglycosylated T3 and 2 Hpt-β-chain peptides) from Hpt were simultaneously monitored in a single 30 min LC-MS run. The potential of the glycopeptide SRM is demonstrated by the identification of differentially expressed site-specific glycosylation isoforms of Hpt, affinity purified from individual lung cancer, matched healthy and inflammation controls (rheumatoid arthritis).

In this work, Hpt was chosen as the model glycoprotein to develop the site-specific glycosylation quantitation strategy. The glycopeptide SRM method, together with sensitive glycosylation discovery (PLOT LC-MS. Chapter 3), could be also applied to other glycoproteins and therefore will contribute to the emerging glycol-biomarker field. In the future, several aspects of this research could be studied further. One aspect is the application of the techniques developed in this thesis (PLOT LC-LTQ-CID/ETD-MS and chip LC-QQQ ) to characterize other
target glycoproteins, such as glycoprotein markers (e.g. prostate specific antigen or alpha-fetoprotein) to improve the sensitivity or specificity or both for diagnostic purposes.

Another aspect is related to technique development. For example, PLOT LC column can be coupled to LTQ-Orbitrap-ETD-MS, the latter of which measures glycopeptide precursor m/z in the Orbitrap detector with high mass resolution and high mass accuracy (< 5ppm). The multiple fragmentation techniques (CID, HCD and ETD) available in LTQ-Orbitrap-ETD-MS will allow detailed structure characterization of glycopeptides. For glycopeptide quantitation, the recently introduced TripleTOF-MS from AB Sciex, which provides similar scan speed to the current triple quadrupole system, but with much higher resolution and mass accuracy, could be used for glycopeptide SRM measurement. The high resolution and mass accuracy of the TripleTOF system may eventually enable direct quantitation of targeted glycopeptides in moderate to complex matrices (for example, high abundant protein depleted plasma, or partially purified protein targets).
4.7 References


Chapter 5

Summary and Perspectives
In this dissertation, mAb proteomics has been shown to be able to generate high quality mAb targeting antigens in their native state with high affinity. The application of mAb proteomics against lung cancer patient plasma samples has been demonstrated to lead to the discovery of cancer specific mAbs and antigens, which may potentially be useful for lung cancer diagnostics. mAb proteomics can be readily applied to other diseases and therefore will contribute to future biomarker research efforts, especially for biomarker verification/validation. For protein glycosylation, an ultrasensitive PLOT LC-LTQ-CID/ETD-MS method has been developed to identify/discover protein glycoforms in a site-specific manner. The developed platform is shown to be straightforward, robust, highly resolving and have sufficient sensitivity to allow characterization of site-specific protein glycosylation from trace amounts of glycosylated proteins. Last, an LC-MS platform, namely glycopeptide SRM using chip LC-QQQ instrument, is developed to quantify site-specific protein glycoforms in a sensitive manner. The SRM glycopeptide is shown to be highly reproducible, sensitive and has a relatively large linear range (4 orders of magnitude). The PLOT LC-LTQ-CID/ETD-MS and chip LC-QQQ allow robust and sensitive screening of glycosylation isoforms from small amounts of protein sample in clinical specimens.

In the future, several aspects of this research can be studied further. This will contribute to the ongoing efforts for protein/glycoprotein biomarker research.

5.1 mAb proteomics against tissue sample/cell line lysate

In this work, an mAb library was generated against cancer plasma proteins, because of the easy availability of plasma samples. However, the tumor-originated antigen(s) are inevitably diluted by the large volume of blood, and the high abundant plasma proteins need to be depleted
using multiple affinity chromatography methods to increase the chance of generating mAbs against lower abundant cancer specific antigens. In the future, mAb proteomics should be applied to cancer tissue/cell line sample. Compared to plasma, cancer tissue or cell lines have significantly lower dynamic range for proteins than in plasma, leading to the increased chance of discovering cancer differentiating mAbs with tissue specificity.

5.2 Site-specific protein glycosylation using PLOT LC-LTQ-CID/ETD-MS

Many established biomarkers are glycoproteins, such as carcinoembryonic antigen, cancer antigen 125 and prostate-specific antigen. One drawback of these biomarkers is their lack of tissue specificity. The developed PLOT LC-LTQ-CID/ETD-MS can be adopted to discover unique biomarker glycosylation isoform(s) to disease samples. This type of study may increase the chance of identifying site-specific protein glycoforms, which can better differentiate normal from disease state than either protein or glycan alone. The PLOT LC-LTQ-CID/ETD-MS can also be useful for elucidation of the structure-function relationship of biologically important molecules such as epidermal growth factor receptor and the role of glycosylation in their biological functions.

To further improve the capability of the PLOT LC-MS for protein glycosylation analysis, the recently developed LTQ-Orbitrap-ETD can be readily coupled to PLOT LC column. LTQ-Orbitrap-ETD-MS measures glycopeptide precursor m/z in the Orbitrap detector with high mass resolution and high mass accuracy (< 5ppm), which could increase the confidence of glycopeptide identification. The multiple fragmentation techniques (CID, HCD and ETD) available in LTQ-Orbitrap-ETD-MS will allow detailed structure characterization of glycopeptides.
5.3 Site-specific quantitation using Chip LC-QQQ

In this work, glycopeptide SRM using Chip LC-QQQ was used to quantitate the glycosylation change of Hpt from lung cancer patient plasmas. In blood, the circulating Hpt mainly comes from liver. It will be interesting to explore the Hpt glycosylation status change in liver disease. Another application of the developed glycopeptide SRM is the in-process monitoring of the glyco-profile of biopharmaceutical products such as antibodies. For the technical development of glycopeptide SRM, the recently introduced TripleTOF-MS from AB Sciex, which provides similar scan speed to the current triple quadrupole system, but with much higher resolution and mass accuracy, could be used for glycopeptide SRM measurement. The high resolution and mass accuracy of the TripleTOF system may eventually enable direct quantitation of targeted glycopeptides in moderate to complex matrices (for example, high abundant protein depleted plasma, or partially purified protein targets).
Publications


2) Comprehensive characterization of the site specific glycosylation of affinity purified haptoglobin using 10 μm PLOT LC-LTQ-CID/ETD-MS, Dongdong Wang, Marina Hincapie, Tomas Rejtar, Barry L. Karger, *Analytical Chemistry*, under review.

3) Selected reaction monitoring to quantify site-specific glycosylation status of haptoglobin affinity purified from lung cancer, rheumatoid arthritis and healthy control plasma, Dongdong Wang, Marina Hincapie, Tomas Rejtar, Fateme Tousi, Barry L. Karger, In preparation.

4) Proteomic Analysis of 10,000 Cells from Laser Captured Microdissected Tumor Breast Tissue Using a Porous Layer Open Tubular (PLOT) Column LC-MS, Dipak Thakur, Tomas Rejtar, Dongdong Wang, Buffie Clodfelder-Miller, Elizabeth Richardson, Dennis Sgroi, Barry L. Karger, In preparation.
Appendix 1 for Chapter 2.

Exoglycosidase treatment to identify critical monosaccharides for the interaction between Hpt and anti-Hpt-β mAb

To identify the critical monosaccharide for the interaction between Hpt and anti-Hpt-β, the monosaccharides were cleaved sequentially from Hpt using exoglycosidases, and the resulting Hpt molecule was probed with anti-Hpt-β. Exoglycosidases used to deglycosylate haptoglobin are shown in panel (a). The specificity of each enzyme is indicated on the N-linked biantennary complex glycan. PNGase F was also included (negative control). Western blotting results for the deglycosylated haptoglobin are shown in panel (b). Lane 1: positive control, haptoglobin; Lane 2: negative control, haptoglobin + PNGase F; Lane 3: haptoglobin + neuramidase; Lane 4: haptoglobin + neuramidase + galactosidase; Lane 5: haptoglobin + neuramidase + galactosidase + N-acetylglucosaminidase.

Neuraminic acid has little effect on the mAb-Ag interaction (lane 3, similar intensity as lane 1). Galactose and N-acetylglucosamine contribute more to the Ab-Ag interaction (lanes 4 and 5, weaker intensity compared to lane 1). Adjusting the chemiluminescence acquisition parameters, the signals from lane 4 and lane 5 gradually disappear, as shown in row I to III, while the signals
from lanes 1 and 3 remain visible. Key to symbols: empty circle = galactose, filled circle = mannose, empty square = GlcNAc, empty diamond = NeuAc.
Appendix 2 for Chapter 2.

Plasma protein fractionation / separation and dot blotting analysis for antigen identification

Protein affinity fractionation / separation and dot blotting were performed with the aim of developing a general method for antigen identification for globally produced mAbs. (a) Pooled plasma protein depletion using multiple affinity removal system (MARS) column (Agilent Technologies, Santa Clara, CA). The flow through was collected and subjected to glycoprotein enrichment using multiple lectin affinity chromatography (M-LAC); (b) Chromatogram of glycoprotein enrichment using M-LAC. The bound fraction was collected and separated using mRP Hi-Recovery Protein Column (Agilent Technologies, Santa Clara, CA). (c) Chromatogram of protein separation using mRP Hi-Recovery Protein Column. Twenty-eight fractions were manually collected. (d) Dot blot analysis was performed by directly spotting of the collected fractions onto a nitrocellulose membrane, and probed by mAb #3 using standard blotting
procedures. Both cancer and healthy samples were subjected to fractionation and dot blotting analysis. The example shown here is from pooled lung cancer plasma sample. It was found that 20 out of 28 fractions gave positive signals during dot blot analysis, which makes it challenging to determine which fraction(s) actually contain(s) the antigen molecule.
Appendix 1 for Chapter 3.

Multistage CID-MS\(^n\) (n = 2, 3, 4) to characterize glycopeptides

One advantage of an ion trap mass spectrometer over other types of instruments is its ability for multistage CID-MS/MS analysis. This advantage was explored by repeated MS/MS analysis of the precursors from the same glycopeptide. Panel (a): CID-MS\(^2\) on glycopeptide m/z 1222.3 (z = 4). The glycan structure is shown in the inset. Panel (b): CID-MS\(^3\) on the dominant fragment (m/z 1410.5, z = 3) from CID-MS\(^2\). Panel (c): CID-MS\(^4\) on the dominant fragment (m/z 1313.1, z = 3) from CID-MS\(^3\).

The CID-MS\(^3\) and MS\(^4\) could provide fragments not observed by CID-MS\(^2\), such as 1786.4 in CID-MS\(^3\) (panel (b)) and 1083.1 in CID-MS\(^4\) (panel (c)). The extra fragmentation could
increase the confidence for glycopeptide identification. However, the advantage of performing multiple stage MS/MS on glycopeptide is limited, since most fragments are already observed in CID-MS$^2$ (panel (a)). Key to symbols: empty circle = galactose, filled circle = mannose, filled square = GlcNAc, filled diamond = NeuAc.
One disadvantage of electron transfer dissociation (ETD) is the low efficiency of precursor fragmentation: during ETD, only part of the activated precursor ions undergoes fragmentation, while others became charge-reduced species, as shown in panel (a). Glycopeptide m/z 1000.8 (z = 4), after ETD fragmentation, was charge reduced to m/z 1334.3 (z = 3) with limited fragmentation (indicated by the z ions). This charge reduced species (m/z 1334.3) was isolated and subjected to charge reduced-CID-MS/MS analysis to improve the ion utilization efficiency. Glycosidic linkage breakage was observed in CID-MS/MS and to provide information on glycan structure.
Appendix 1 for Chapter 4.

Precursor ion scan to identify glycans presented on the same peptide sequence

Precursor ion scan (also called fixed product ion scan) was performed using Chip LC-QQQ to test whether or not this type of MS operation is suitable for identification of the glycans presented on individual glycopeptides. Glycopeptide T1 is shown here as an example. Only 2 glycoforms (m/z 222 and m/z 1314) can be identified by monitoring glycan oxonium ion 366 under the precursor ion scan mode. As a comparison, PLOT LC-MS identified 7 glycoforms (Chapter 3). The low sensitivity of the precursor ion scan prevents the usage of QQQ for discovery purposes.
Appendix 2 for Chapter 4.
Neuraminic acids influence the glycopeptide retention on a C\textsubscript{18} column

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Glycopeptide Diagram" /></td>
<td><img src="image" alt="Glycopeptide Diagram" /></td>
</tr>
</tbody>
</table>

Glycopeptides, sharing the same peptide sequence but different glycan compositions, will closely elute, but not co-elute, with each other. An Agilent packed C\textsubscript{18} Chip-LC could slightly separate the individual glycopeptides, as shown in panel (a). The elution order is found to be correlated with the number of neuraminic acids present on individual glycopeptides (glycoforms on glycopeptide T1 are shown here). This result is reproducibly observed, as shown in panel (b). The structure of the glycan is indicated in the figure. The first number represents the number of antenna, the second indicates the number of neuraminic acids, “S” indicates “neuraminic acid/sialic acid”, and the third number indicates the number of fucoses, “F” indicates “fucose”.

218