ADVANCES IN PROTEIN POST-TRANSLATIONAL MODIFICATIONS (PTMS) USING LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY

A dissertation presented

by

Wenqin Ni

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, Massachusetts

February, 2013
ADVANCES IN PROTEIN POST-TRANSLATIONAL MODIFICATIONS (PTMS) USING LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY

by

Wenqin Ni

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry and Chemical Biology in the College of Science of Northeastern University

February, 2013
Abstract

Protein post-translational modifications (PTMs) play significant roles in affecting the physical, chemical, and biological properties of proteins. PTMs can regulate protein functions by modulating protein activity, turnover, cellular location, and protein-protein interaction. In addition, the efficacy and safety of therapeutic protein drugs can be dramatically affected by PTMs. Structural elucidation of complex PTMs provides invaluable insights for the potential roles of PTMs. However, comprehensive characterization of PTMs presents a significant analytical challenge, often requiring the use of multiple orthogonal methods to ensure high confidence in the resulting analytical data. Liquid chromatography coupled with mass spectrometry (LC-MS) is one of the most powerful platforms for determination of PTMs due to its high sensitivity, resolution, and accuracy. LC-MS tools usually can achieve the goal of identification of PTMs, including modified proteins, specific modification sites, and structures of the modifications. This thesis focuses on the methods for the determination of several important PTMs, including deamidation of asparagine, isomerization of aspartic acid, disulfide linkages, and glycosylation using LC-MS approaches.

In chapter 2, a method with differentiation and enrichment of isoaspartic acid (isoAsp) by Asp-N digestion combined with detection and quantification of isoAsp by LC-MS/MS with electron transfer dissociation (ETD) is presented.

In Chapter 3, a proteome-level analysis of isoAsp in urine proteins from the wild type and protein L-isoaspartyl O-methyltransferase (PIMT, a highly conserved isoAsp repair enzyme) deficient mice is reported.
In Chapter 4, a successful and robust methodology for complete characterization of disulfide linkages, including cystine knots and nested disulfides in recombinant human arylsulfatase A (rhASA), is developed using multi-enzyme digestion and LC-MS methods with ETD and sequential collision induced dissociation (CID-MS$^3$) analysis.

In Chapter 5, the method development for comprehensive characterization of oligosaccharide structures with glycan composition, sequence, linkage, and positional information is described. The newly developed method employs fluoride-mediated negative ionization LC-MS/MS using a microfluidic chip packed with porous graphitized carbon (PGC) for N-glycan separation and an accurate mass quadrupole time of flight (Q-TOF) tandem mass spectrometer for characterization. An application to the characterization of N-glycans released from polyclonal human and murine IgG is described.
Acknowledgements

It is my great pleasure to thank all of the people who have supported and helped me fulfill this dissertation although I know it is hard for me to identify all of them.

First of all, I am greatly indebted to my advisor, Dr. Barry L. Karger, for his guidance, support, encouragement and patience with me. His scientific intuition and thinking will significantly impact my career.

I would like to express my deepest gratitude to: Dr. Jonathan Bones and Dr. Shiaw-Lin Billy Wu. I would like to thank both for the patient supervision and enthusiastic support. I would like to offer my great appreciation to Dr. Zhaohui Sunny Zhou and Dr. Shujia Daniel Dai for their help and support.

My special thanks go to my committee members: Drs. Graham Jones, William Hancock and Roger Giese for reviewing my thesis and the valuable suggestions.

I particularly thank the former and present people in Barnett Institute: Dr. Tianzhu Zang, Joshua Klaene, Dr. Yi Wang, Dr. Dongdong Wang, Dr. James Glick, Dr. Alexander Ivanov, Zhenke Liu, Chen Li, Siyang Li, Victoria Berger, Somak Ray, Siyuan Liu, Yuanwei Gao, Min Liu, Dr. Marina Hincapie, Dr. Alexander Ivanov, Dr. Sangwon Cha, Dr. Dipak Thakur, Dr. Zhi Zeng, Dr. Haitao Jiang, Dr. Qiaozhen Lu, Dr. Sandor Spisak, Xianzhe Wang, Suli Liu, Yue Zhang, Fangfei Yan and Fan Zhang for their help not only in science but also in my life.

Thanks are extended to my collaborators: Melanie Lin, Paul Salinas and Philip Savickas from Shire Human Genetic Therapies; Alexander Patananan and Steven Clarke
from Department of Chemistry and Biochemistry and the Molecular Biology Institute in UCLA. They have generously provided us samples, shared data and given helpful discussion.

I want to acknowledge the help from the staff in Department of Chemistry and Chemical Biology and Barnett Institute: Nancy Carbone, Jean Harris, Andrew Bean, Cara Shockley and Richard Pumphrey.

Finally and most importantly, I wish to dedicate this dissertation to my family members: my husband, son, parents, sister and parents-in-law. I am sincerely grateful for their generous understanding, endless love and selfless support.
# Table of Contents

Abstract ........................................................................................................................................... 2

Acknowledgements .......................................................................................................................... 4

Table of Contents ............................................................................................................................. 6

List of Figures .................................................................................................................................... 14

List of Tables ..................................................................................................................................... 21

Chapter 1: Introduction: Overview of Protein Post-Translational Modifications and Their LC-MS Determination .................................................................................................................. 22

1.1 Abstract ..................................................................................................................................... 23

1.2 Post-Translational Modifications of Proteins (PTMs) .................................................................... 23

1.2.1 Background ............................................................................................................................ 23

1.2.2 Formation .................................................................................................................................. 25

1.2.2.1 Deamidation of Asparagine and Isomerization of Aspartic Acid .............................................. 26

1.2.2.2 Disulfide Linkage .................................................................................................................. 27

1.2.2.3 Glycosylation ....................................................................................................................... 29

1.2.3 Biological Significance ............................................................................................................. 34

1.2.3.1 Deamidation of Asparagine and Isomerization of Aspartic Acid .............................................. 34

1.2.3.2 Disulfide Linkage .................................................................................................................. 35

1.2.3.3 Glycosylation ....................................................................................................................... 37

1.2.4 Importance of PTMs in Biopharmaceuticals ............................................................................. 40

1.2.5 Challenges in PTM Identification ............................................................................................. 43

1.3 Liquid Chromatography and Mass Spectrometry Based Techniques .............................................. 44
1.3.1 Liquid Chromatography .................................................................................................................. 45
  1.3.1.1 Reversed Phase Chromatography (RPC) .................................................................................... 45
  1.3.1.2 Porous Graphitized Carbon (PGC) ............................................................................................. 47
  1.3.1.3 Hydrophilic Interaction Chromatography (HILIC) ...................................................................... 48
  1.3.1.4 Ion Exchange Chromatography (IEC) ....................................................................................... 49
  1.3.1.5 Chromatography in a Microfluidic Chip Format ......................................................................... 50

1.3.2 Mass Spectrometry ............................................................................................................................ 51
  1.3.2.1 Soft Ionization Methods .............................................................................................................. 51
    1.3.2.1.1 Matrix-Assisted Laser Desorption Ionization (MALDI) ......................................................... 51
    1.3.2.1.2 Electrospray Ionization (ESI) ............................................................................................... 52
  1.3.2.2 Ion Activation ............................................................................................................................... 55
    1.3.2.2.1 Collision-Induced Dissociation (CID) .................................................................................. 56
    1.3.2.2.2 Electron Transfer Dissociation (ETD) .................................................................................. 59
  1.3.2.3 Mass Analyzers and Tandem Mass Spectrometry ....................................................................... 60
    1.3.2.3.1 Quadrupole ............................................................................................................................ 63
    1.3.2.3.2 Ion Trap .................................................................................................................................. 63
    1.3.2.3.3 Time of Flight (TOF) ............................................................................................................. 64
    1.3.2.3.4 Orbitrap ................................................................................................................................... 64
    1.3.2.3.5 Fourier Transform Ion Cyclotron Resonance (FTICR) ......................................................... 65
    1.3.2.3.6 Quadrupole-Time-of-Flight (Q-TOF) ..................................................................................... 65
    1.3.2.3.7 Triple Quadrupole (QQQ) ...................................................................................................... 66
  1.3.2.4 Quantitative Mass Spectrometry .................................................................................................. 67
    1.3.2.4.1 Label Free Quantitation ......................................................................................................... 67
Chapter 2: Theoretical Analysis of IsoAsp Peptide Identification

2.4.2 Limit of Detection for IsoAsp Peptide when Co-eluted with Asp Peptide .................................................. 109
2.4.3 Differentiation and Enrichment of IsoAsp Peptides via Asp-N Digestion ..................................................... 111
2.4.4 Unbiased Identification of IsoAsp ................................................................................................................. 112
2.4.5 Specificity of Asp-N ...................................................................................................................................... 116
2.4.6 Elution Order of IsoAsp and Asp Peptides ............................................................................................... 117
2.4.7 Detection of IsoAsp in Aged Beta-amyloid Peptide ............................................................................... 120
2.5 Conclusions ...................................................................................................................................................... 122
2.6 Appendix .......................................................................................................................................................... 123
   2.6.1 Monitor Sample Preparation Induced IsoAsp Using $^{18}$O Labeling ...................................................... 123
2.7 References ...................................................................................................................................................... 125

Chapter 3: An Integrated Proteomic Analysis of Major Isoaspartyl-Containing Proteins in the Urine of
Wild Type and Protein L-Isoaspartate O-Methyltransferase-Deficient Mice ......................................................... 131

3.1 Abstract .......................................................................................................................................................... 132
3.2 Introduction ................................................................................................................................................... 133
3.3 Experimental .................................................................................................................................................. 136
   3.3.1 Materials ................................................................................................................................................ 137
   3.3.2 Wild Type and PIMT Knockout Mice and Urine Collection ................................................................. 137
   3.3.3 Urine Fractionation Procedure ............................................................................................................ 137
   3.3.4 Quantitation of Isoaspartyl Content in Mouse Urine ........................................................................ 138
   3.3.5 Proteolytic Digestion ............................................................................................................................ 139
   3.3.6 LC-MS Analysis .................................................................................................................................... 140
   3.3.7 Data Processing .................................................................................................................................... 141
3.4 Results and Discussion .................................................................................................................................. 142
3.4.1 Secreted Proteins/Peptides in Urine of PIMT-Knockout Mice Contain Significant and Elevated Amounts of isoAsp ..............................................................142
3.4.2 Proteomic Analysis of Mouse Urine Proteins ..................................................................................144
3.4.3 Mapping of isoAsp Sites in MUPs by ETD-SRM Analysis .................................................................144
3.4.4 IsoAsp Formation during Sample Preparation ...................................................................................154
3.4.5 IsoAsp in MUPs from Wild Type and PIMT-Knockout Mice ...............................................................154
3.5 Conclusions .......................................................................................................................................155
3.6 Appendix ............................................................................................................................................156
3.6.1 Protein Identification by In-Gel Digestion and LC-MS Analysis .........................................................156
3.7 References ..........................................................................................................................................157

Chapter 4: Complete Mapping of a Cystine Knot and Nested Disulfides of Recombinant Human Arylsulfatase A by Multi-Enzyme Digestion and LC-MS Analysis Using CID and ETD ..........................163
4.1 Abstract ...........................................................................................................................................164
4.2 Introduction ......................................................................................................................................164
4.3 Experimental ....................................................................................................................................167
4.3.1 Samples ......................................................................................................................................167
4.3.2 Reagents ......................................................................................................................................167
4.3.3 Enzymatic Digestion ......................................................................................................................167
4.3.4 LC-MS .........................................................................................................................................168
4.3.5 Disulfide Assignment .....................................................................................................................169
4.4 Results and Discussion .......................................................................................................................170
4.4.1 Digestion Strategy ..........................................................................................................................171
4.4.2 Unpaired Cysteines: Cys20, Cys51, Cys276 ..................................................................................173
6.2.3 Glycosylation..........................................................................................................................247

6.2.3.1 Characterization of Anionic Glycans: Sulfated and Phosphorylated Glycans ..........247

6.2.3.2 Determination of Sialylated Glycans ...........................................................................247

6.2.3.3 Characterization of O-linked Glycans...........................................................................248

6.2.3.4 Automation by Chip Format..............................................................................................249

6.2.3.5 Characterization of Glycans on Different Mass Spectrometers. .............................250

6.3. References....................................................................................................................................250

Appendix............................................................................................................................................253

Publications........................................................................................................................................253

Copyright from American Chemical Society (ACS) .................................................................255

Copyright from Springer..................................................................................................................257
## List of Figures

| Figure 1-1 | The diversity of gene products due to alternative splicing on post-transcriptional regulation and co/post-translational modifications | 24 |
| Figure 1-2 | Disulfide formation mechanism by thiol-disulfide exchange | 28 |
| Figure 1-3 | The types of N-linked glycans and the core structure with Man3GlcNAc2 is highlighted in red | 32 |
| Figure 1-4 | Core structures of O-linked glycans | 32 |
| Figure 1-5 | In vivo N-glycosylation synthesis. (A) synthesis of Glc3Man9GlcNAc2-PP-dolichol as the glycan donor; (B) transfer of Glc3Man9GlcNAc2 to Asn residues; (C) process of N-glycan chain | 33 |
| Figure 1-6 | (A) Charge induced dipole at the PGC stationary phase; (B) analyte alignment on the PGC surface | 48 |
| Figure 1-7 | The scheme of ESI process | 53 |
| Figure 1-8 | Formation and interpretation of a, b, c, x, y and z fragments in peptides | 58 |
| Figure 1-9 | Formation of b and y ions in peptide dissociation | 58 |
| Figure 1-10 | Formation and interpretation of A, B, C, X, Y and Z fragments in oligosaccharides | 58 |
| Figure 1-11 | Three major pathways for oligosaccharides with metal adducts | 59 |
| Figure 1-12 | The mechanism for 2,4A cross-ring fragments in negative ion ESI analysis of oligosaccharides | 59 |
| Figure 1-13 | Chemical structure of (A) ICAT, (B) ITRAQ and (C) TMT | 70 |
| Figure 1-14 | Reaction of free cysteine with DTNB | 75 |
| Figure 1-15 | Released oligosaccharides and their common derivatizations | 81 |
| Figure 2-1 | Formation of Isoaspartic Acid via Deamidation of Asparagine or Isomerization of Aspartic Acid. The peptide backbones are highlighted in bold | 98 |
| Figure 2-2 | Mechanism of fragmentation of the Asp and isoAsp peptides in ETD MS. (a) formation of c and z fragment ions of the Asp peptides, which is the same for isoAsp peptides (not shown); (b) formation of the c+57 and z-57 diagnostic ions of the isoAsp peptides | 99 |
| Figure 2-3 | Structures of Asp and isoAsp peptides. Asp-N cleaves Asp at its N-terminal side but isoAsp is resistant to proteolysis by Asp-N | 100 |
| Figure 2-4 | Overview of the Approach Combining the Differentiation and Enrichment of IsoAsp Peptides by Asp-N Digestion and the Identification of IsoAsp Species by ETD-MS | 101 |
| Figure 2-5 | ETD spectra of Cyt-C (a) and isoCyt-C (b) | 107 |
| Figure 2-6 | ETD spectra of Aβ1-16 fragments from Lys-C digests of Aβ1-42 (a) and isoAβ1-42 (b) | 108 |
| Figure 2-7 | Enrichment of isoAsp-containing peptide in cytochrome c via Asp-N digestion. The dashed line indicates the average value of peak area ratios of z10 - 57 to z10 observed in the synthetic isoCyt-c peptide | 109 |
| Figure 2-8 | Correlation between the percentage of isoAsp in Lys-C digested Aβ1-42 and isoAβ1-42 and the peak area ratio of (z10-57)/z10 (a) and (c6+57)/c6 (b). The insets show the lower percentage region | 110 |
| Figure 2-9 | Correlation between the percentage of isoAsp in Cyt-c and isoCyt-c and the peak area ratio of (z10-57)/z10 (a) and (c5+57)/c5 (b) | 111 |
| Figure 2-10 | Enrichment of isoAsp containing peptide in Lys-C digested Aβ1-42 via Asp-N digestion | 112 |
| Figure 2-11 | Extracted ion chromatogram and ETD spectra of the Aβ1-16 peptide | 114 |
| Figure 2-12 | Extracted ion chromatogram and ETD spectra of synthetic Cyt-c peptide | 115 |
| Figure 2-13 | Extracted ion chromatogram and ETD spectra of synthetic Aβ1-42 peptide. | 116 |
| Figure 2-14 | Extracted ion chromatograms of Asp/isoAsp containing Aβ1-16 fragments generated by Lys-C digested synthetic Aβ1-42 or isoAβ1-42. (a) and (b) demonstrate the XICs of fragments of the Aβ1-16 Asp form (a) and isoAsp form (b), respectively, before Asp-N digestion. (c) and (d) show the XICs of Asp form (c) and isoAsp form (d) of the Aβ1-16 fragments and its Asp-N digested fragments. | 117 |
| Figure 2-15 | HPLC/UV chromatograms of synthetic Cyt-c and isoCyt-c using (a) 0.1 % TFA and (b) 0.1 % formic acid as ion pairing reagents in the mobile phase. (c) Total ion chromatogram of synthetic Cyt-c and isoCyt-c using 0.1 % formic acid as ion pairing reagent in nanoLC-MS. | 119 |
| Figure 2-16 | HPLC/UV chromatograms of Lys-C digested synthetic Aβ1-42 and isoAβ1-42 using (a) 0.1 % TFA and (b) 0.1 % formic acid as ion pairing reagents in the mobile phase; (c) total ion chromatogram of Lys-C digested synthetic Aβ1-42 and isoAβ1-42 using 0.1 % formic acid as ion pairing reagent in nanoLC-MS | 120 |
| Figure 2-17 | Extracted ion chromatograms of aged A1-16 peptide before and after Asp-N digestion. (a) and (b) show the XICs of the triply charged precursor ion from aged Aβ1-16 peptide before and after Asp-N digestion, respectively; (c) and (d) demonstrate the XICs of product ions z10 and z10 - 57 of aged A1-16 peptide before Asp-N digestion, respectively | 122 |
## List of Figures (Continued)

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2-18</td>
<td>(a) Extracted ion chromatograms of Aβ1-16 peptide after incubation in 18O water; (b) MS of Aβ1-16 peptide after incubation in 18O water; (c) ETD-MS/MS spectrum of Aβ1-16 peptide after incubation in 18O water and the expanded parts in the inset</td>
</tr>
<tr>
<td>Figure 3-1</td>
<td>Formation of isoAsp via deamidation of asparagine or isomerization of aspartic acid and the repair of isoAsp by PIMT</td>
</tr>
<tr>
<td>Figure 3-2</td>
<td>Enzymatic detection of isoAsp in mouse urinary species via PIMT-catalyzed methylation followed by fluorography of 3H-methyl ester after SDS-PAGE</td>
</tr>
<tr>
<td>Figure 3-3</td>
<td>Examination of isoAsp sites in MUP1 peptide (50-73). (a) ETD spectrum of peptide IED52N53GN55FRLFLEQIHVLEN68SLVLK and its isomers from MUP2; (b and c) the expanded ETD spectra of the peptide IED52N53GN55FRLFLEQIHVLEN68SLVLK and its isomers, leading to the identification of deamidation sites</td>
</tr>
<tr>
<td>Figure 3-4</td>
<td>Extracted ion chromatograms and ETD spectra of MUP1 peptide (50-73), IE52D53NG55NFRLFLEQIHVLE68NSLVLK. (a) Extracted ion chromatograms of MUP1 peptide and its isomers from wild type and PIMT-deficient mice; (b) ETD spectrum of native MUP1 peptide (50-73); (c,d) the reporter ions of isoAsp (c3+57, z6-57) in ETD MS/MS spectra from the selected precursors of above peptide and its isoaspartyl isoforms</td>
</tr>
<tr>
<td>Figure 4-1</td>
<td>Primary structure of rhASA with disulfide linkages and unpaired cysteines</td>
</tr>
<tr>
<td>Figure 4-2</td>
<td>(a) Mass and charge of the pepsin-digested peptide with an unmodified (free) Cys20, and (b) CID-MS2 spectrum of the precursor from (a)</td>
</tr>
<tr>
<td>Figure 4-3</td>
<td>(a) Mass and charge of the pepsin-digested peptide with a modified Cys51, which is converted to formylglycine, and (b) CID-MS2 spectrum of the precursor from (a)</td>
</tr>
<tr>
<td>Figure 4-4</td>
<td>(a) Mass and charge of the pepsin-digested peptide with an unmodified (free) Cys51, and (b) CID-MS2 spectrum of the precursor from (a)</td>
</tr>
<tr>
<td>Figure 4-5</td>
<td>(a) Mass and charge of the pepsin-digested peptide with an unmodified (free) Cys276, and (b) CID-MS2 spectrum of the precursor from (a)</td>
</tr>
<tr>
<td>Figure 4-6</td>
<td>(a) Mass and charge of the pepsin-digested peptide with one single disulfide linkage (Cys282–Cys396); (b) CID-MS2 spectrum of the precursor from (a); and ETD-MS2 spectrum of the precursor from (a)</td>
</tr>
<tr>
<td>Figure 4-7</td>
<td>The sequence of nested disulfide in trypsin digestion</td>
</tr>
<tr>
<td>Figure 4-8</td>
<td>The digestion strategy for nested disulfides with multi enzyme digestion</td>
</tr>
<tr>
<td>Figure 4-9</td>
<td>The sequence of nested disulfides after Lys-C, trypsin, PNGase F and Asp-N digestion</td>
</tr>
<tr>
<td>Figure 4-10</td>
<td>(a) Mass and charge of the Lys-C plus trypsin plus Asp-N plus PNGaseF-digested peptide with two disulfides (Cys138 with Cys154, and Cys143 with Cys150), and (b) CID-MS2 spectrum of the 2+ charged precursor from (a); (c) ETD-MS2 spectrum of the 3+ charged precursor</td>
</tr>
<tr>
<td>Figure 4-11</td>
<td>(a) Chromatographic retention for correct and scrambled disulfides; (b) Precursor ion of scrambled disulfide (Cys138 with Cys143, and Cys150 with Cys154), and CID-MS2 spectrum of the 2+ charged precursor</td>
</tr>
</tbody>
</table>
### List of Figures (Continued)

| Figure 4-12 | Precursor ion of nested disulfide (Cys138 with Cys154, and Cys143 with Cys150) peptide without glycosylation in pepsin digestion, and CID-MS2 spectrum of the 4+ charged precursor | 187 |
| Figure 4-13 | (a) Mass and charge of the pepsin-digested peptide with cystine knot (Cys470 with Cys482, Cys471 with Cys484 and Cys475 with Cys481), and (b) CID-MS2 spectrum of the 3+ charged precursor from (a) | 189 |
| Figure 4-14 | (a) ETD-MS2 spectrum (using the Orbitrap) of the cysteine knot precursor (m/z 656.30, 4+), (b) ETD-MS2 spectrum (using the Orbitrap) of the same peptide from (a) but with different charged precursor (m/z 525.20, 5+), and (c) CID-MS3 spectrum (using the Orbitrap) of m/z 1312.6 from (a) | 190 |
| Figure 4-15 | CID-MS3 spectrum (using ion trap) of m/z 1312.6 from ETD-MS2 spectrum (using ion trap) of the cysteine knot precursor (m/z 656.30, 4+) | 191 |
| Figure 5-1 | Positive and negative ESI ion CID analysis of infused maltohexaose glucose-α-1-4-Glucose (Dp6). Precursor ions of [M-H]- and adducts of Dp6 in negative ion analysis with (A) 10 mM ammonium bicarbonate and (B) 10 mM ammonium fluoride in 50% water and 50% acetonitrile; (C) annotated negative ion CID-MS/MS of the pseudomolecular anion of Dp6 [M-H]- in 10 mM ammonium fluoride in 50% of water and 50% of acetonitrile; (D) annotated positive ion CID-MS/MS of the protonated ion of Dp6 [M+H]+ in 0.1% formic acid in 50% of water and 50% of acetonitrile | 208 |
| Figure 5-2 | Annotated ESI negative ion CID-MS/MS analysis of [M-H]- ions of infused (A) cellohexaose (glucose-β-1-4-glucose), (B) laminarihexaose (glucose-β-1-3-glucose) and (C) glucose-α-1,6-glucose hexamer | 210 |
| Figure 5-3 | Annotated ESI negative ion CID-MS/MS analysis of infused Dp6: (A) [2M-H+F]2- ion, and (B) [M+F]- ion | 211 |
| Figure 5-4 | Extracted ion chromatogram and negative CID-MS/MS spectra of (A) FA2G2 and (B) A2G2S2 | 215 |
| Figure 5-5 | Negative CID-MS/MS spectra of (A) the N-glycan with a free reducing terminal GlcNAc residue and (B) the corresponding alditol for the triantennary trigalactosyl N-glycan (A3G3) | 216 |
| Figure 5-6 | (A) Extracted ion chromatogram of F(6)A2G[6]1, more commonly known as G1 and F(6)A2G[3]1, more commonly known as G1'. (B) Annotated ESI negative ion CID MS/MS spectrum for F(6)A2G[6]1 at the retention time of 19.6 min. (C) Negative ion CID-MS/MS spectra of F(6)A2G[3]1S(6)1 at the retention time of 20.7 min, indicating diagnostic fragments for comprehensive structural annotation | 221 |
| Figure 5-7 | (A) Base peak chromatogram of murine IgG; annotated ESI negative ion CID-MS/MS spectra of (B) FA2G2Gal[6]1 at the retention time of 21.3 min and (C) F(6)A2G2Gal[6]1S(Neu5Gc)[3]1 at the retention time of 22.4 min | 223 |
| Figure 5-8 | Anion exchange chromatogram for glycan fraction | 230 |
| Figure 5-9 | Extracted ion chromatogram and negative CID spectra of A2G2S1; (A) LC-MS elution profile, (B) precursor mass of doubly charged A2G2S1 in negative ion Q-TOF MS analysis, (C) annotated negative ion CID-MS/MS spectrum of Peak 3. (D) peak annotation of A2G2S1 | 234 |
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1-1</td>
<td>Important post-translational modifications.</td>
<td>25</td>
</tr>
<tr>
<td>Table 1-2</td>
<td>Common and important monosaccharides.</td>
<td>31</td>
</tr>
<tr>
<td>Table 1-3</td>
<td>The major functions of glycans on glycoproteins</td>
<td>38</td>
</tr>
<tr>
<td>Table 1-4</td>
<td>Comparison of commonly used mass spectrometric instrumentation.</td>
<td>62</td>
</tr>
<tr>
<td>Table 1-5</td>
<td>Strategies for glycosylation analysis.</td>
<td>78</td>
</tr>
<tr>
<td>Table 2-1</td>
<td>Sequences and abbreviations of the synthetic peptides.</td>
<td>102</td>
</tr>
<tr>
<td>Table 3-1</td>
<td>Identified proteins (&gt;2 distinct peptides) in mouse urine by discovery proteomics.</td>
<td>146</td>
</tr>
<tr>
<td>Table 3-2</td>
<td>Selected candidates for targeted ETD-SRM detection.</td>
<td>147</td>
</tr>
<tr>
<td>Table 3-3</td>
<td>Identification and quantitation of isoAsp peptides in MUPs.</td>
<td>153</td>
</tr>
<tr>
<td>Table 4-1</td>
<td>Unpaired cysteine and disulfide linkages in rhASA</td>
<td>173</td>
</tr>
<tr>
<td>Table 5-1</td>
<td>N-linked oligosaccharides identified by fluoride-mediated negative ion chip-LC-MS/MS in the human polyclonal IgG oligosaccharide pool.</td>
<td>220</td>
</tr>
<tr>
<td>Table 5-2</td>
<td>N-linked oligosaccharides identified by fluoride-mediated negative ion chip-LC-MS/MS in the murine polyclonal IgG oligosaccharide pool.</td>
<td>222</td>
</tr>
<tr>
<td>Table 5-3</td>
<td>N-linked oligosaccharides identified by fluoride-mediated negative ion chip-LC-MS/MS in α-1-acid glycoprotein.</td>
<td>233</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction: Overview of Protein Post-Translational Modifications and Their LC-MS Determination
1.1 Abstract

This chapter provides an introduction to the background and technologies of this thesis. First, the background of protein post-translational modifications, especially for isoaspartic acid, disulfide linkages and glycosylation, is described. Their formation, biological significance and analytical challenges for identification are presented. Second, current liquid chromatography and mass spectrometry based techniques are overviewed. Finally, the commonly used techniques for characterization of isoaspartic acid, disulfide linkage and glycosylation are discussed in detail.

1.2 Post-Translational Modifications of Proteins (PTMs)

1.2.1 Background

The human proteome is considerably more complex and larger than human genome, for example, ~1.8 million of protein species have been predicted from the reported 21,000 genes, thus far annotated in the human genome\(^1\). Diversity within the human proteome is achieved via two predominant mechanisms-mRNA splicing and post-translational modification of proteins, as shown in Figure 1-1\(^1\). The specific focus of this thesis is the determination of protein post-translational modifications using liquid chromatography-mass spectrometry (LC-MS) methods.

Post-translational modification is the covalent addition of a specific functional group to the protein after its translation, in which the polypeptide is biosynthesized in the ribosome. Usually, covalent bonds between the attached chemical functionality and the protein are formed at specific amino acid residues. PTMs occur on most amino acid residues in proteins except the side chains of leucine, isoleucine, valine, alanine and
phenylalanine\textsuperscript{2}. In addition, there is one other type of PTM generated by protein backbone cleavage, termed proteolysis\textsuperscript{2}. To date, approximately 300 different PTMs have been identified in physiological conditions\textsuperscript{1}. Table 1-1 lists some of the common PTMs, such as glycosylation, oxidation, deamidation, disulfide linkage, phosphorylation, ubiquitination, and acetylation. In this thesis, the characterization of deamidation of asparagine, isomerization of aspartic acid, disulfide linkages and N-glycosylation profiling are specifically described.

**Figure 1-1.** The diversity of gene products due to alternative splicing on post-transcriptional regulation and co/post-translational modifications\textsuperscript{1}. 
Table 1-1. Important post-translational modifications.

<table>
<thead>
<tr>
<th>PTM type</th>
<th>Modified amino acid residue</th>
<th>ΔMass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycosylation</td>
<td>Asn (N-linked)</td>
<td>&gt;800</td>
</tr>
<tr>
<td></td>
<td>Ser, Thr (O-linked)</td>
<td>203, &gt;800</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>Tyr, Ser, Thr</td>
<td>+80</td>
</tr>
<tr>
<td>Disulfide linkage</td>
<td>Cys</td>
<td>-2</td>
</tr>
<tr>
<td>Methylation</td>
<td>Arg, Lys</td>
<td>+14</td>
</tr>
<tr>
<td>Deamidation</td>
<td>Asn, Gln</td>
<td>+1</td>
</tr>
<tr>
<td>Oxidation</td>
<td>Met</td>
<td>+16, +32</td>
</tr>
<tr>
<td>Ubiquitination</td>
<td>Lys</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>Sulfation</td>
<td>Tyr</td>
<td>+80</td>
</tr>
<tr>
<td>Hydroxyproline</td>
<td>Pro</td>
<td>+16</td>
</tr>
<tr>
<td>Nitration</td>
<td>Tyr</td>
<td>+45</td>
</tr>
<tr>
<td>Formylglycine</td>
<td>Cys, Ser</td>
<td>-18</td>
</tr>
<tr>
<td>Pyroglutamic acid</td>
<td>Gln, Glu (N-terminus)</td>
<td>-17, -18</td>
</tr>
<tr>
<td>Acetylation</td>
<td>N-terminus</td>
<td>+42</td>
</tr>
<tr>
<td>Isomerization</td>
<td>Asp, Glu</td>
<td>0</td>
</tr>
</tbody>
</table>

1.2.2 Formation

The mechanism for formation of PTMs varies based on the modification. The majority of PTM formation is regulated by a variety of specific enzymes including glycosyltransferases that govern glycosylation processing and kinases that generate protein phosphorylation\textsuperscript{3,4}. However, some proteins can undergo spontaneous modifications...
without enzyme involvement\textsuperscript{5}. Here, a brief introduction of the formation of isoaspartic acid, disulfide bond and glycosylation, which are the subjects of this thesis, is presented in the following sections.

1.2.2.1 Deamidation of Asparagine and Isomerization of Aspartic Acid

Isoaspartic acid (isoAsp, isoD), aspartic acid in a beta-peptide linkage, is a ubiquitous post-translational modification. To date, this modification has been reported in more than 200 proteins\textsuperscript{6}. IsoAsp can be spontaneously generated through the non-enzymatic deamidation of asparagine or isomerization of aspartic acid, as shown in Figure 3-1\textsuperscript{7}. The loss of ammonia from asparagine or the dehydration of aspartic acid leads to the formation of a labile intermediate of succinimide which readily hydrolyzes to isoaspartic acid and aspartic acid\textsuperscript{7,8}. The limiting step in the physiological condition is formation of succinimide\textsuperscript{6}. Succinimide formation from asparagine is about 10 times faster than that from aspartic acid\textsuperscript{8}. The dependence of the deamidation or isomerization rate on pH, temperature, and protein sequence, secondary/tertiary structures, solution buffer and ionic strength have been widely reported\textsuperscript{6}. Usually, deamidation occurs in neutral and basic conditions (pH\textsuperscript{\textgreek{g}}7) while isomerization happens more readily in acidic conditions\textsuperscript{6}. Furthermore, the deamidation or isomerization rate is affected by C-terminal neighboring amino acids to asparagine or aspartic acid\textsuperscript{6}. The fastest deamidation sequence motif is an asparagine-glycine (NG) sequence; next is asparagine-serine (NS) sequence and the slowest is asparagine-proline (NP) sequence. The rate is slowed by steric hindrance and accelerated by the charged or hydroxyl groups of the adjacent C-terminal amino acids\textsuperscript{6}. Protein secondary and tertiary structures also play important roles in deamidation. Generally speaking, deamidation depends 60\% on the protein sequence and 40\% on the secondary and
tertiary structure. The half-life of deamidation in physiological conditions varies from 1 day to 1000 days. As deamidation and isomerization occur spontaneously without enzymatic catalysis, these modifications can also occur in vitro.

Isoaspartic acid can be partially repaired by protein L-isoaspartyl O-methyltransferase (PIMT or PCMT, EC 2.1.1.77), as shown in Figure 3-1. PIMT is a highly conserved isoAsp repair enzyme which recognizes and transfers the methyl group from S-adenosyl-L-methionine to isoAsp, but not Asp, to form the methyl ester. The labile methylester is rapidly converted back to succinimide and subsequently hydrolyzes into Asp and isoAsp. The enzymatic conversion by PIMT of the abnormal isoAsp residue to the Asp residue in proteins prevents the accumulation of a potentially dysfunctional protein. Studies in PIMT deficient mice show that there is increased isoaspartyl accumulation in tissues, such as brain, and in fluids, such as urine, in PIMT-deficient mice compared to wild type mice. Furthermore, the average life span for PIMT-deficient mice (42 days) is much shorter than for wild type mice (22-26 months), demonstrating the harmful consequences of isoAsp accumulation.

Deamidation of glutamine (Gln) or isomerization of glutamic acid (Glu) has also been reported. The deamidation of asparagine is about 2 orders of magnitude faster than that of glutamine as the five-membered ring of succinimide in Asn deamidation is much more stable than the six-membered ring in glutarimide of Gln deamidation. However, the formation of pyroglutamic acid is very fast when Gln or Glu is at the N-terminus due to the formation of the five-membered ring of pyrrolidone.

1.2.2.2 Disulfide Linkage
A disulfide bond (also called disulfide linkage or bridge) is formed by the oxidation of cysteine and is one of the most common post-translational modifications. The molecular mechanism for formation of the disulfide bond has been studied extensively. The main pathway of formation in prokaryotes and eukaryotes is similar. Briefly, the disulfide bond is formed via oxidation of sulfhydryl (SH) group with a thiol-disulfide exchange reaction in vivo, as shown in Figure 1-2. In the reaction, a thiol anion (S-) attacks a sulfur atom in one disulfide bond, resulting in mixed disulfide bonding. In the following step, the remaining thiol anion replaces the sulfur in the mixed disulfide and resolves the corrected disulfide linkages. This in vivo thiol-disulfide exchange reaction is catalyzed by a class of proteins called thio-disulfide oxidoreductases. Protein disulfide isomerase (PDI) is one of the first identified thio-disulfide oxidoreductases. PDI can catalyze the formation and reduction of the disulfide bond by transfer of electrons and hydrogen. PDI with the oxidized form of cysteine (disulfide) can oxidize dithiols and transfer disulfide to proteins while, in the reduced form (dithiol), it can catalyze reduction of mispaired thiols. The activity of this type of protein depends on a Cys-X-X-Cys motif, where X represents any amino acid. Disulfide formation or scrambling can also form in vitro without catalysis of enzymes. Usually, high pH deprotonates thiols and forms a thiol anion, which can initiate thiol disulfide exchange.

![Disulfide formation mechanism](image)

**Figure 1-2.** Disulfide formation mechanism by thiol-disulfide exchange.
1.2.2.3 Glycosylation

Glycosylation is a diverse co/post-translational modification that involves the attachment of carbohydrates to a protein through either asparagine residue (Asn) (N-glycosylation) or serine/threonine residue (Ser/Thr) (O-glycosylation). The carbohydrates are made of multiple monosaccharides via glycosidic linkages. The predominant monosaccharides in glycosylation are listed in Table 1-2. Glycosidic linkage is a covalent bond, which links one monosaccharide to another. For example, 1-4 glycosidic linkage is that C-1 of the first monosaccharide is bonded to C-4 of the second monosaccharide with loss of water. The anomeric definition (alpha or beta) is based on the configuration between the anomeric center and the reference atom. In hexoses, C-1 and C-5 act as anomeric and reference carbon, respectively. Alpha anomer is defined as the opposite stereochemistry at C-1 and C-5 while beta anomer is the same stereochemistry. Glycosylation occurs in the consensus sequence of Asn-X-Ser/Thr, wherein X represents any amino acid except proline. All N-glycosylations have a common core structure consisting of three mannose residues and two N-acetyl glucosamine residues (Man$_3$GlcNAc$_2$), to which varieties of monosaccharides are attached to the core structure to form three major subgroups: high mannose type, complex type and hybrid type, as shown in Figure 1-3 $^{17}$. O-linked glycans are attached to Ser/Thr residue without any consensus amino acid sequences. There are eight different types of core structures in O-glycosylation, as shown in Figure1-4 $^{17}$.

The N-glycosylation process occurs in two cell organelles, the endoplasmic reticulum (ER) and the Golgi, wherein a complex series of enzymes are involved in the sequential assembly of monosaccharides by adding or removing specific monosaccharides, as shown in Figure1-5 $^{18}$. N-glycosylation starts at the cytoplasmic side of the ER.
membrane, by synthesis of Glc3Man9GlcNAc2-PP-dolichol as the glycan donor. The first stage is to attach seven monosaccharides (2 GlcNAc and 5Man residues) via uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) and guanosine diphosphate mannose (GDP-Man) donors, followed by the flipping of the entire oligosaccharide structure to the luminal side of ER. The second stage is to attach seven more monosaccharides (4 mannose and 3 glucose) utilizing the mannose-PP-dolichol and glucose-PP-dolichol as donors to build the Glc3Man9GlcNAc2-PP-dolichol. Next, the Glc3Man9GlcNAc2 in the glycan donor of Glc3Man9GlcNAc2-PP-dolichol is transferred to asparagine in the consensus sequence of Asn-X-Ser/Thr of the nascent peptide with catalysis by OSTase. The beta-hydroxyl (β-OH) group in Ser/Thr interacts with the side chain amino group of Asn via a hydrogen bond, enhancing the nucleophilicity of the amino group\(^\text{18}\). After the Glc3Man9GlcNAc2 is transferred to an asparagine residue, the three glucose sugars are removed by glucosidases to form a glycan of Man9GlcNAc2. The ER mannosidase removes one mannose to form Man8GlcNAc2 in ER lumen. The glycan of Man8GlcNAc2 are further trimmed by Golgi mannosidases to form the core structure of Man3GlcNAc2 in Golgi complex. Then the addition of the Man or GlcNAc or other monosaccharides to the core structure sequentially lead to the diverse final products.

The synthesis of O-linked glycans is simpler than that of N-linked glycan. Neither lipid-linked intermediates nor glycosidases are involved in O-glycan synthesis\(^\text{19}\). The first sugar (GalNAc) is directly transferred from UDP-GalNAc to a serine or threonine residue by the catalysis of polypeptide-N-acetyl-galactosaminyltransferase\(^\text{19}\). Then the formation of the core structure is catalyzed by several specific glycosyltransferases in a stepwise manner\(^\text{19}\).
Table 1-2. Common and important monosaccharides.

<table>
<thead>
<tr>
<th>Monosaccharide</th>
<th>Structure</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mannose (Man)</td>
<td><img src="image" alt="Mannose Structure" /></td>
<td><img src="image" alt="Green" /></td>
</tr>
<tr>
<td>Galactose (Gal)</td>
<td><img src="image" alt="Galactose Structure" /></td>
<td><img src="image" alt="Yellow" /></td>
</tr>
<tr>
<td>Glucose (Glu)</td>
<td><img src="image" alt="Glucose Structure" /></td>
<td><img src="image" alt="Blue" /></td>
</tr>
<tr>
<td>Fucose (Fuc)</td>
<td><img src="image" alt="Fucose Structure" /></td>
<td><img src="image" alt="Red" /></td>
</tr>
<tr>
<td>N-acetylgalactosamine (GlcNAc)</td>
<td><img src="image" alt="N-Acetylgalactosamine Structure" /></td>
<td><img src="image" alt="Blue" /></td>
</tr>
<tr>
<td>N-acetylgalactosamine (GalNAc)</td>
<td><img src="image" alt="N-Acetylgalactosamine Structure" /></td>
<td><img src="image" alt="Yellow" /></td>
</tr>
<tr>
<td>N-acetyleneuraminic acid (NeuAc)</td>
<td><img src="image" alt="N-Acetyleneuraminic Acid Structure" /></td>
<td><img src="image" alt="Purple" /></td>
</tr>
<tr>
<td>N-glycolyl neuraminic acid (Neu5Gc)</td>
<td><img src="image" alt="N-Glycolyl Neuraminic Acid Structure" /></td>
<td><img src="image" alt="White" /></td>
</tr>
<tr>
<td></td>
<td>R=H, NeuAc; R=OH, Neu5Ge</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1-3. The types of N-linked glycans and the core structure with Man$_3$GlcNAc$_2$ is highlighted in red$^{17}$.

Figure 1-4. Core structures of O-linked glycans$^{17}$.
Figure 1-5. *In vivo* N-glycosylation synthesis. (A) synthesis of Glc3Man9GlcNAc2-PP-dolichol as the glycan donor; (B) transfer of Glc3Man9GlcNAc2 to Asn residues; (C) process of N-glycan chain\(^{18}\).
1.2.3 Biological Significance

The biological significance of PTMs for protein structure and function has been widely recognized. PTMs regulate enzyme activity, cellular location of proteins, stability and turnover, as well as dynamic interactions of the protein\textsuperscript{20,21}. Here, a brief discussion of the significance of isoaspartic acid, disulfide bond and glycosylation in fundamental biology is presented.

1.2.3.1 Deamidation of Asparagine and Isomerization of Aspartic Acid

Deamidation of asparagine introduces an additional negative charge on a protein by converting asparagine into aspartic acid and isoaspartic acid\textsuperscript{8}. Furthermore, formation of isoaspartic acid results in the insertion of a methylene group into the protein backbone, generating a “kink”\textsuperscript{7}. These changes may dramatically alter the protein structure and function thus, resulting in diseases. In this section, isoaspartic acid associated diseases such as ageing, cancers, Alzheimer’s disease and immunogenicity are briefly discussed.

\textit{Ageing:} Ageing is a complex process, the fundamental mechanisms of which remain largely unknown. Many modifications such as deamidation, oxidation, glycosylation and ubiquitination are proposed to be involved\textsuperscript{22}. In 1970, Robinson and coworkers proposed a molecular clock mechanism of deamidation for ageing\textsuperscript{6,23}. To date, increased deamidation of proteins has been reported with increased age, e.g., aged-dependent deamidation in human lens\textsuperscript{6,24}. Furthermore, it is reported that deamidation in subunits of crystallins in lens can also induce protein aggregation\textsuperscript{24}.

\textit{Cancers:} B-cell lymphoma-extra large (Bcl-xl) is an anti-apoptotic protein, which plays an important role in the survival of tumor cells. The deamidation of Bcl-xl at Asn52
and Asn 66 contributes to the protein function loss; thus, tumor cells are able to resist to apoptosis.

**Alzheimer’s disease (AD):** AD is a group of age-dependent brain diseases. While the fundamental mechanism for this disease is not clear, it has been reported that there are elevated deamidation products (isoaspartic acid and aspartic acid) in Alzheimer’s patients, and isomerization of Asp is found in amyloid plaques in Alzheimer’s disease. Furthermore, it has been determined that isoaspartic acid changes the solubility of beta-amyloid peptide, which plays an important role in AD.

**Immunogenicity:** Peptides containing isoaspartic acid are not recognized as self-peptides in cells, leading to immunogenicity. Mamula and co-workers have found that the digestion by cathespsin D (an enzyme involving in antigen process) of Asp/isoAsp containing cytochrome c peptide 88–104 (PCC 88-104) is different for the two isoforms. Furthermore, T cells, which recognize PCC 88-104, are observed to proliferate to a greater extent in response to isoAsp than Asp. These data suggest that isoAsp containing peptides as a non-self peptide may potentially trigger immunogenicity.

**1.2.3.2 Disulfide Linkage**

Disulfide bonds play vital roles for protein folding, stability and activity. The failure to form correct disulfide bonds may lead to protein misfolding and loss of activity. Furthermore, disulfide bonds often contribute to the formation of specific motifs such as cystine knots and catalytic motifs such as Cys-X-X-Cys, which play an important role in protein function. Unpaired cysteine (free or modified) is also critical for protein
stability and activity\textsuperscript{31}. In the following, the disulfide linkage and unpaired cysteine as a function of protein stability and activity is discussed.

**Disulfide and protein stability:** It is well known that disulfide bonds contribute to the thermal and chemical stability of proteins\textsuperscript{32}. For example, disulfide bond can stabilize protein structure by holding two portions of protein together, such as connecting heavy chain and light chain together in immunoglobulin G (IgG). In addition, disulfide bond can also help to form hydrophobic core to stabilize protein structure. The function of disulfide bond in stabilizing protein structure can be examined by two experimental approaches: disruption of naturally occurring disulfides or introduction of an extra disulfide bond. Disruption of the disulfide bond in the C\textsubscript{H}3 domain of IgG1, for example, leads to a significant instability to denature by heat or chemical reagent such as guanidinium chloride\textsuperscript{33,34}. Another example is the increased stability observed by insertion of a new disulfide bond. For example, in subtilisin, a disulfide bond of Cys 61-Cys 98 is engineered with site mutation of Gly 61 and Ser 98 into Cys, which enables a three-fold increase in the half-life of enzyme activity and an increase in the protein melting point\textsuperscript{32}.

**Disulfide and protein activity:** Many proteins lose activity once a disulfide bond is reduced due to structural changes and unfolding. For example, experiments show that reducing a disulfide can inactivate the protein of Aspergillus niger phytase\textsuperscript{32}. Another example is arylsulfatase A, which contains a three-disulfide intertwined cystine knot structure at the C-terminus. It is found that mutation of one out of six cysteine residues in the cystine knot structure, resulting in the disruption of cystine knot structure, will lead to enzyme activity loss\textsuperscript{35}. 
**Unpaired cysteine:** The importance of unpaired cysteines (free or modified) for protein stability and function is increasingly recognized. For example, free cysteines are very reactive and easily produce disulfide scrambling, thus affecting the stability of a protein by affecting protein conformation. For example, site mutation of Cys 6 and 11 to alanine and serine, respectively, in human and bovine superoxide dismutase, singly or in combination, have been shown to lead to more resistance to thermal inactivation than wild type superoxide dismutase\(^3\). Cysteine modification also plays an important role in protein function. In arylsulfatase, the free cysteine in CXP/AXR motif (X represents any amino acid residue) is required to be converted to formylglycine, which is essential for protein catalytic activity\(^3\). Failure of conversion of this free cysteine to formylglycine will result in inactive enzyme, which is associated with a disease of metachromatic leukodystrophy (MLD) in humans\(^3\).

### 1.2.3.2 Glycosylation

Protein glycosylation is one of the most complex PTMs, with important roles in diverse biological processes. Tables 1-3 lists the functions of the glycans\(^4\). Here, protein glycosylation in modulating protein activity, immunogenicity and recognition in protein interaction is briefly discussed. Furthermore, the alteration of glycosylation in diseases is also introduced.
Table 1-3. The major functions of glycans on glycoproteins.  

<table>
<thead>
<tr>
<th>Type</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical and Chemical</td>
<td>• Modify electrical charge, size, mass,</td>
</tr>
<tr>
<td></td>
<td>solubility and viscosity</td>
</tr>
<tr>
<td></td>
<td>• Regulate protein folding</td>
</tr>
<tr>
<td></td>
<td>• Effect protein stability</td>
</tr>
<tr>
<td></td>
<td>• Protect protein against proteolysis</td>
</tr>
<tr>
<td>Biological</td>
<td>• Regulate protein traffic and localization</td>
</tr>
<tr>
<td></td>
<td>• Control lifetime in circulation</td>
</tr>
<tr>
<td></td>
<td>• Modify immunogenicity</td>
</tr>
<tr>
<td></td>
<td>• Regulate interaction</td>
</tr>
<tr>
<td></td>
<td>• Modify protein activity</td>
</tr>
</tbody>
</table>

*Modulation of biological activity:* Tissue plasminogen activator (tPA) is a serine protease with the function of converting plasminogen into plasmin, an important enzyme for clot breakdown. Thus, tPA plays a significant role in the removal of clots. TPA has four potential N-glycosylation sites: Asn 103,117,184 and 448. It is reported that glycosylation on Asn184 and Asn448 is essential for the catalytic activity of tPA. As another example, human granulocyte-macrophage colony stimulating factor (GM-CSF) plays an important role in the immune system by stimulating stem cells to produce granulocytes and monocytes. It is reported that increasing N-linked or O-linked glycosylation will greatly decrease the activity of this protein due to the conformational change and inactivation of the protein. Additionally, the glycosylation of immunoglobulin
G (IgG) has been extensively studied. Each heavy chain of IgG has the N-glycosylation site at Asn 297. The N-glycosylation involves the stability and modulation of physiological functions\textsuperscript{44-46}. The stability results in the non-covalent bonding between the oligosaccharide and protein backbone\textsuperscript{45}. Depletion of the oligosaccharides at Asn 297 is reported to lead to a conformational change, reducing the interaction between the FcγRIIIa receptor and Fc portion of an IgG\textsuperscript{45,47}. The binding is also reduced by the core fucose on the reducing terminal GlcNAc residue\textsuperscript{48}. Alteration of the N-glycosylation at Asn297 is associated with autoimmune disease and cancer\textsuperscript{49,50}. In addition, the glycan linkage also plays an important role for protein function. It is reported that α-2,6-sialylation inhibits the binding of N-glycans to galectin 3, thus is a negative regulator for binding and function while unsialylation or α2,3-sialylation demonstrated much higher binding to galectin 3\textsuperscript{51}.

**Immunological properties:** Carbohydrates play important roles in the immune system. The ABO human blood-type immune-determinant is one example with different oligosaccharides in the A, B and O type (GalNAc for A type, Gal for B type and Fuc residue for O type, respectively) on antigens\textsuperscript{19}. Furthermore, it is reported that in some instances monoclonal antibodies generated against animal cells also recognize glycans\textsuperscript{40,52}. This result further suggests that glycans can affect antigenicity. Another example is natural killer cells, which are responsible for lysis of tumor cells and play an important role against cancers with the immune system. It has been found that tumor cells with mannose-type and hybrid-type carbohydrates are more susceptible to lysis than those with complex-type carbohydrates\textsuperscript{40}. This result suggests that carbohydrate plays a selectivity role in tumor cell lysis.
**Recognition determinants:** The diversity of carbohydrates on proteins also modulates protein targeting or protein-protein interaction\(^{40,52}\). For example, proteins with sulfated carbohydrates are cleared more rapidly than the proteins expressed in CHO cells with sialylated carbohydrates\(^{40}\). It is hypothesized that sulfated carbohydrate is the signal tag for fast clearance of proteins to prevent overloading of receptors\(^{40}\).

**Alterations in glycosylation with diseases:** As carbohydrates play diverse roles in protein activity, clearance, signal transduction and cell adhesion, alterations in glycosylation are considered as potential biomarkers for diseases such as cancer and arthritis\(^{52-55}\). One example is the inclusion-cell disease, a lysosomal storage disease caused by failure to produce oligosaccharides containing mannose 6-phosphate in the Golgi due to phosphotransferase defects\(^{54,55}\). As a consequence of this defect, the protein is secreted outside the cells. Consequently, lysosomes cannot function normally as catabolic enzymes to break down the waste materials and cellular debris, resulting in accumulation of these species within the cells and the resultant physiological manifestation\(^{55}\). Glycosylation changes are also associated with cancer\(^{50,53}\). Beta-1, 6 GlcNAc branching of N-glycans, potentially results in cancer progression due to glycan structural changes and subsequent loss of protein activity\(^{19}\). Also, glycoproteins are reported as cancer biomarkers, such as alpha-fetoprotein (AFP) for liver cancer and prostate-specific antigen (PSA) for prostate cancer, although their biomarker activity is a result of expression changes and not solely the presence of or alteration in the oligosaccharides attached\(^{53}\).

### 1.2.4 Importance of PTMs in Biopharmaceuticals

The majority of protein therapeutics bears some forms of post-translational modifications, which may affect the physical and chemical properties of proteins, and thus
potentially affect efficacy and safety\textsuperscript{41,43}. The International Conference for Harmonization Guideline, ICH Q6B, outlines requirements for the characterization of biopharmaceutics, in which the disulfide bridges and carbohydrate structures are required to be determined to the best extent possible. Scrambled disulfide linkages, glycosylation isoforms and deamidated or isomerized protein products need to be detected and characterized as product-related impurities. The advent of follow-on (biosimilar) therapeutics further highlights the importance for complete characterization of PTMs as such modifications are heavily affected by different manufacturing processes such as upstream and downstream processing\textsuperscript{56,57}. On one side, PTMs address the issue of drug safety such as immunogenicity\textsuperscript{58}. Efforts have been made to eliminate aberrant PTMs and thus to reduce concerns regarding product safety. On the other side, efforts have been made for manipulation of PTM-engineered proteins to increase product circulating half-life or efficacy\textsuperscript{41,43}. In this section, PTMs in modulating the immunogenicity, efficacy and stability of recombinant protein drugs are discussed.

**PTMs and Immunogenicity:** Several factors contribute to immunogenicity, e.g., non-human products, impurities, contaminants and aggregates\textsuperscript{56}. Aberrant PTMs can also affect immunogenicity by converting proteins into a non-self human structure that can subsequently trigger an immune response\textsuperscript{56,59}. As mentioned previously, isoaspartic acid has been shown to be immunogenic as it is recognized as a non-self peptide or protein\textsuperscript{27,28}. Oxidation of methionine and cysteine residues in human recombinant interferon α-2b also triggers immunogenicity due to a resulting change in protein folding\textsuperscript{60}. As proteins are expressed in non-human cells such as CHO and murine cells, certain potentially immunogenic non-human glycans such as galactose-α-1, 3-galactose and N-
glycolylneuraminic acid, are produced\textsuperscript{61,62}. These non-self products due to PTMs in recombinant proteins can potentially trigger immunogenicity. The determination of these antigenic glycans is discussed in the following chapters.

\textbf{PTMs and Drug efficacy:} It has been reported that glycosylation impacts the secretion, bioactivity, folding, pharmacokinetics and immunogenicity for biotherapeutics\textsuperscript{41,43}. Erythropoietin (EPO) is a 36 kDa glycoprotein with three N-glycans (Asn24, 38 and 83) and one O-glycan (Ser 126)\textsuperscript{41}. The glycosylation sites at Asn 38 and 83 are critical for the secretion of the protein from the producer cells, and removal of any of these two glycosylation sites will lead to poor secretion\textsuperscript{41}. Furthermore, the removal of N-linked glycans results in low activity \textit{in vivo} due to the shorter plasma half-life of the protein\textsuperscript{41}. Sialic acid also plays an important role for the plasma half-life of EPO, which decreases by 150 fold after desialylation\textsuperscript{41}. In addition, technology of glyco-engineered therapeutics has been developed to produce drugs with superior efficacy and safety. The Aranesp version of EPO in Amgen with two extra engineered N-glycan sites is reported to increase the overall level of sialylation therefore increasing the circulating half-life and potency\textsuperscript{63-65}.

Recombinant IgG is a major class of approved monoclonal antibodies that contain one N-glycosylation site at Asn 297 in the C\textsubscript{H}2 domain on each heavy chain. As noted above, the antibody bearing afucosylated N-glycans (without a core fucose) show a substantial increase of ADCC activity (>100 fold) compared to antibodies bearing fucosylated glycans due to the increased affinity binding of FcgRIIIa on natural killer cells (NK cells) to the Fc portion of an IgG\textsuperscript{66}. Thus, a new generation of glyco-engineered IgGs is in development focusing on the inactivation of the core fucosyltransferase\textsuperscript{48,66}. For
examples, glycol-engineered anti-CD20 antibody and anti-EGFR antibody lacking core fucose in Roche has demonstrated increased affinity binding and ADCC activity\textsuperscript{66}. These two antibodies are undergoing clinical trial studies now.

**PTMs and drug stability:** Deamidation in sequence motifs such as -NG- is fast and is one of the major factors contributing to protein instability\textsuperscript{6}. Considerable effort has been made to minimize deamidation, such as protein formulation in acidic conditions and site mutation of the –NG- sequence \textsuperscript{7,13,67,68}. Cysteine is a highly reactive amino acid and often causes disulfide scrambling, which also leads to the instability of proteins\textsuperscript{69}. There have been several strategies developed for modification or mutation of free cysteine to improve protein stability \textit{in vivo} or \textit{in vitro}. For example, it is reported that free cysteine is removed by site mutation in human fibroblast factor 1 to increase the stability, and cysteine PEGylation has been used to improve the \textit{in vivo} stability of interferon α-2b and antibodies\textsuperscript{70,71}. Much effort has been made to develop engineered cell lines which can enhance disulfide formation such as over-expression of Dsb proteins in \textit{E.coli} and knockout of the thioredoxin reductase gene, thus facilitating disulfide formation\textsuperscript{43}.

**1.2.5 Challenges in PTM Identification**

The generation of novel information regarding PTMs in various biological processes, human diseases and pharmaceutical proteins can lead to further insight toward understanding the mechanistic roles of PTMs and therefore, improve the production and quality of protein drugs. Currently, the commonly used methods for identifying PTMs are Western blot analysis and mass spectrometry (MS) based strategies\textsuperscript{21}. Western blot analysis, which depends on the known structure of PTM and the availability of specific antibodies, has limitations for the site-specific analysis. On the other hand, MS based
analysis can provide increased information about the PTMs, such as the modification-associated proteins, the specific sites of modification in proteins, occupancy of the modified sites and structures of the modifications.

MS based PTM analysis is still challenging. First, biological samples, such as plasma, urine, cell lysate are very complex and full characterization of PTMs on the proteome level is still challenging. Second, proteins usually contain multiple and different types of PTMs, and this usually complicates the data analysis in mass spectrometry. Third, many PTMs are present in low stoichiometry and therefore enhanced detection sensitivity is required. Fourth, some PTMs are labile during sample preparation and MS analysis. For example, the phosphate group usually undergoes neutral loss during MS analysis, thus limiting the peptide sequence information. It usually requires multiple tandem mass spectrometry such as MS3 or a variety of ion activation methods such as electron transfer dissociation or electron capture dissociation to obtain more sequence information. Fifth, sample preparation may introduce artificial modifications to a protein, such as deamidation, oxidation and disulfide scrambling. Sixth, some modified peptides during shot-gun peptide centric proteomics are often not recovered by liquid chromatography and thus are not accessible for MS analysis. Finally, data annotation of many PTMs needs to be manual checked for verification and data quality. For example, it is reported that false positive data are generated from automatic searches of diagnostic ions for isoaspartic acid as these ions are of relatively low intensity. Thus, these aspects of PTM characterization present a formidable analytical challenge.

1.3 Liquid Chromatography and Mass Spectrometry Based Techniques
Liquid chromatography and mass spectrometry were developed in the mid and late 1900s, respectively. The LC-MS analytical platform has the advantages of high sensitivity, robustness and reproducibility. As mentioned above, these techniques are widely used for determination of PTMs. In this section, a brief introduction to liquid chromatography and mass spectrometry is presented as these tools have been widely used in this thesis.

1.3.1 Liquid Chromatography

Liquid chromatography was discovered in early 1900s by Tswett, who investigated the separation of colored pigments using polar sorbents and nonpolar organic solvents. In the early stages of the development of this technique, normal phase chromatography (NPC) was the dominant mode for separation. Here, the stationary phase is a polar sorbent, such as silica or alumina, and the mobile phase is a non-polar solvent such as hexane (relatively less polar) and chloroform (relatively more polar). In this separation mode, hydrophilic compounds are adsorbed stronger than hydrophobic compounds, and thus the hydrophobic compounds elute more quickly.

To date, different kinds of liquid chromatography have been developed based on a variety of separation mechanisms, such as reversed phase chromatography, hydrophilic interaction chromatography, size exclusion chromatography and ion exchange chromatography. In this section, the introduction to several types of liquid chromatography is presented, such as reversed phase chromatography, porous graphite carbon, hydrophilic interaction chromatography and ion exchange chromatography. Furthermore, chromatography in chip-based format also is discussed.

1.3.1.1 Reversed Phase Chromatography (RPC)
In RPC, the stationary phase is a nonpolar sorbent such as octadecyl silica (C\textsubscript{18}) and the mobile phase is a polar solvent, commonly a water/methanol or water/acetonitrile mixture. Hydrophobic compounds (less polar) interact with the nonpolar stationary phase stronger than hydrophilic compounds (more polar). Retention in RPC is a result of the interaction of the sample with the mobile phase and stationary phase. Decreasing of the polarity of the mobile phase, i.e. increasing the percentage of the organic modifier of the mobile phase, is helpful for the elution of hydrophobic sample components. The separation mechanism for RPC is a combination of solvophobic interaction (the solute is aligned and attached to ligand), partition (the stationary phase is considered as the liquid phase, in which the solute is dissolved) and adsorption (solute is retained on the surface of stationary phase)\textsuperscript{74}.

The stationary phase for RPC consists of the porous matrix beads, on which the hydrophobic ligands are covalently bonded. Common matrices include silica and synthetic polymers\textsuperscript{75}. Silica beads are stable between pH 2 to 7.5 while the synthetic polymers are stable over the range of pH 1 to 12. The ligands, which are grafted to the surface of the bead, are usually linear hydrocarbon chains (alkyl group, C\textsubscript{1}-C\textsubscript{30}), phenyl or cyano groups. More hydrophobic samples will be readily separated when a less hydrophobic ligand is used. For example, typical sorbents for peptide separation are C\textsubscript{18} bonded phases and for protein C\textsubscript{8} or C\textsubscript{4}.

In RPC, the mobile phase is a mixture of aqueous and organic solvents (methanol or acetonitrile). The addition of ion pairing reagents may help improve the retention of charged samples and also modify the elution profile of samples, for example, by decreasing
tailing. The function of the ion pairing reagent is to bind the solute by ionic interaction, thus the hydrophobicity of solute will be modified and will therefore be more suitable for subsequent RPC separation\textsuperscript{76}. However, a major drawback associated with the addition of ion pairing reagents is signal suppression when MS detection is used\textsuperscript{77}. In the following sections, several types of liquid chromatography for separation of polar or charged analytes are discussed.

\subsection{1.3.1.2 Porous Graphitized Carbon (PGC)}

Although the polar compounds can be separated by ion pairing RPC as mentioned above, the separation of highly polar compound is still challenging. To date, porous graphitized carbon, hydrophilic chromatography (HILIC) and ion exchange chromatography (IEX) are particularly useful for separation of highly polar or charged compounds such as glycan, which is one topic of this thesis.

PGC was developed in 1980s by Knox\textsuperscript{78}. PGC is extremely stable over a wide pH range of 0 to 14 and at temperatures lower than 200 °C\textsuperscript{79}. PGC is composed of intertwined graphitic ribbons and has a relative homogenous structure. The covalent bonds with in the sheet and van der Waals interaction between sheets contribute to the stability of the packing. The separation mechanism of PGC is complicated and dependent on both polarity and shape of the molecule approaching the graphite surface\textsuperscript{78,79}. In general, the retention of analyte on PGC depends on disperse interactions between the solute and stationary phase, which is similar as in RPC: the polar retention effect on graphite (PREG), and the charge induced polarization on the surface of PGC, as shown in Figure 1-6 (A)\textsuperscript{79}. Dipole-dipole interaction between the analyte and stationary phase makes PGC highly suited for the separation of
polar compounds. In addition to its ability to separate polar compounds, PGC also facilitates isomer separation such as stereo, diastereo and linkage isomers as the analyte can have intense interaction with PGC and strong retention if it gets a close fit to PGC, as shown in Figure 1-6 (B). Charged analytes such as sialylated or sulphated glycans usually have strong retention on PGC and it usually requires certain ionic strength of the mobile phase for complete elution from the stationary phase.

Figure 1-6. (A) Charge induced dipole at the PGC stationary phase; (B) analyte alignment on the PGC surface.

1.3.1.3 Hydrophilic Interaction Chromatography (HILIC)
HILIC is an aqueous normal phase chromatography in which the elution order is reversed to that of RPC. HILIC employs the polar stationary phases such as silica, amino or cyano, and organic-aqueous mobile phases such as acetonitrile and water. Consequently, the separation of analytes can be achieved by adjusting the ratio of water to acetonitrile, water being the stronger elution solvent. Although the complete separation mechanism on HILIC remains unclear, its major mechanism is partitioning of polar analytes in and out of water layer on HILIC surface while the mechanism in NPC is adsorption. Given the aqueous mobile phase used in HILIC, it shows several advantages by overcoming the drawbacks of poor solubility of polar samples in NPC and weak retention of highly polar or charged compounds in RPC. In addition, elution from HILIC usually starts with high organic phase, which is a weaker elution solvent and is followed by increasing the amount of water, which is a stronger elution solvent. The high organic solvent also helps electrospray in electrospray ionization; and thus improves sensitivity for mass spectrometry. To date, HILIC is widely used for glycan separation and analytical strategies for glycan separation using HILIC is discussed in the following sections.

1.3.1.4 **Ion Exchange Chromatography (IEC)**

Ion exchange chromatography separates charged analytes based on the electrostatic interaction of ions, which includes two types of ion exchange chromatography: anion exchange chromatography (AEC) and cation exchange chromatography (CEC). In AEC, the stationary phase has the positive charges covalently connected to the resin, which can retain analytes with negative charges. The order of elution can be achieved by controlling net charge of analyte through adjusting solution pH or by adding salt to compete the ionic interactions. In contrast to AEC, the resin in CEC has the negative charges, which interact
with the positive charges in analyte. In this thesis, AEC is used for pre-fractionation of glycan pools based on their charges.

1.3.1.5 Chromatography in a Microfluidic Chip Format

As mentioned above, LC-MS is one of the most popular workhorses in analytical labs. Chip based LC-MS has been reported with many advantages: low sample and solvent consumption, high sensitivity, reproducibility, fast separation and simple operation. However, the development of LC-MS chip has faced two pitfalls. The first challenge is development of on-chip electrospray, and the second challenge is connection of a flat chip to an LC pump. In later 1990s, our group and Ramsey’s group reported MS analysis via ESI using a microfluidic chip. The electrospray was directly performed in the open edge of the monolith. Later, the insertion of a capillary into the end of chip was also studied. To date, the most commonly used method is the integrated emitter, which can eliminate dead volume, and thus dramatically improve sensitivity. The connection of a flat chip to an LC pump has been fulfilled in an Agilent commercial available microfluidic chip. This device provides a stable sealed connection and an on-chip switching valve. This type of commercial chip is fabricated from polyimide, which has the compatibility with most organic solvents. The laser ablation processed chip is composed of an enrichment column, switching valves, a separation column and an integrated nanospray tip. In addition, the most recent version of the chip has included an immobilized enzyme reactor (PNGase F) facilitating online deglycosylation. To date, chip-based LC-MS is widely used in metabolic, proteomic and glycomic studies. In chapter 5, the application of the PGC chip with negative ionization mass spectrometry in characterization of glycan structure is presented.
1.3.2 Mass Spectrometry

Mass spectrometry is an important analytical technique by measuring the mass to charge ratio of analyte. The mass spectrometer is composed of three modules: an ion source, a mass analyzer and a detector. In this section, the introduction to the ionization method, ion activation, mass analyzer and tandem mass spectrometry is described. Furthermore, quantitative mass spectrometry is also discussed.

1.3.2.1 Soft Ionization Methods

There are a variety of ionization methods, such as electron ionization, chemical ionization, field ionization, fast atom bombardment, field desorption, matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI). Among these methods, MALDI and ESI are most popular ionization methods for biomolecules as these two methods are relatively soft, which means very little energy is transferred to the analyte and limited in-source fragmentation occurs, making them suitable for the ionization of thermally labile large molecules. Tanaka and Fenn were awarded 2002 Nobel Prize for chemistry for development of MALDI and ESI, respectively. In this following part, MALDI and ESI are briefly discussed.

1.3.2.1.1 Matrix-Assisted Laser Desorption Ionization (MALDI)

MALDI was introduced in the mid-1980s by Hillenkamp, Karas, Tanaka and other workers. Ionization in MALDI consists of three steps. The first step is co-crystallization of analyte with matrix. The second step is laser triggered desorption. The last step is ionization, which protonates or deprotonates the analyte. The mechanism for MALDI is still not well established. However, the most accepted model is that the matrix is desorbed
and ionized by laser beam and then a proton is transferred to the analyte from the matrix resulting in the protonation of the analyte\textsuperscript{91}. In addition, the matrix can prevent the clustering of analyte by separating the molecules and minimizing sample damage by absorption of energy from the laser pulse. Generally, single charged ions dominate in MALDI, which make it suitable for intact analysis of large biological molecules, such as proteins. The time of flight (TOF) mass analyzer is widely coupled to MALDI using a pulsed laser beam, and the mass spectrometer can analyze the large mass range generated from MALDI. However, MALDI-TOF MS/MS has drawbacks, such as low shot-to-shot reproducibility due to the matrix selection and sample preparation.

1.3.2.1.2 \textbf{Electrospray Ionization (ESI)}

ESI was developed in late\textsuperscript{1980s} by Fenn\textsuperscript{92}. In contrast to MALDI, ESI produces the ions with multiple charges, thus reducing their mass-to-charge ratio to a lower range that is a suitable match for common mass analyzers. In addition, ESI can continuously transfer the analyte from the solution phase to gas phase, which makes it suitable for coupling with high performance liquid chromatography or capillary electrophoresis to achieve online separations. To date, HPLC-ESI-MS has become a sensitive, rapid and high throughput analytical platform for various samples.

The ionization process in ESI, as shown in Figure 1-7, consists of three major steps: generation of charged parent droplets, small charge droplet formation due to solvent evaporation and droplet disintegration, and production of gas-phase ion from small charged droplet\textsuperscript{77}. The mechanism for the first two steps is well studied while the mechanism for the last step remains unclear. In the first step, electrochemical oxidation and reduction reactions occur in the positive and negative ion mode, respectively \textsuperscript{77}. Particles with accumulated
charges from these redox reactions are then extracted from a capillary, in which a high voltage has been applied and charged parent droplets are subsequently produced. In the second step, two forces act oppositely on charged parent droplets: surface tension and electrostatic Coulomb force of repulsion. Surface tension tries to keep the droplet shape spherical while Coulomb force tries to break up the droplet. The charged parent droplets becomes extremely small (about 1-10 nm of radius) due to solvent evaporation. Then, charged parent droplet breaks down into small highly charged droplets when the Coulomb force is stronger than surface tension. The small charged droplet is not only smaller than the charged parent ion but also has larger mass to charge ratio.

Figure 1-7. The scheme of ESI process.

In the last step, two major models for ion formation are proposed: the charged residue model and the ion evaporation model. In the charged residue model, it is proposed that desolvation of charged small droplets occurs until droplets are formed (about 1 nm of radius), that contain only one analyte molecule. This results in the charge to be retained on the analyte in the gas phase thereby forming a gas-phase analyte ion. In the ion evaporation model, it is proposed that ion emission occurs directly due to the electric field.
caused by the surface charges when the radius of droplet is less than 10 nm. Generally, large molecules such as proteins follow the charged residue model while small molecules such as organic or inorganic ions follow ion evaporation mode.

Solvent composition in ESI can affect the response by enhancement or suppression of signals. Generally speaking, a good ESI solution can generate a stable spray and minimize the chemical noise from itself. In the positive ion model, a weak acid such as formic acid (pKa 3.77) is usually added to the solvent such as water/acetonitrile to protonate the analyte. However, strong acids such as trifluoroacetic acid (pKa 0.23) are not recommended for use with ESI as they cause strong signal suppression in MS. In addition, cations such as sodium are commonly used to analyze polar or neutral analyte by facilitating the formation of adducts in ESI. In the negative ion mode, ammonium hydroxide is used to adjust the solution to the basic region, resulting in the deprotonation of the analyte to form negative charges. However, this method usually generates poor spray stability. In order to maintain a stable spray in the negative ion mode, halogenated solvents are preferred as stable negative ions can be created by electrochemical reduction. For example, fluorinated solvents at neutral pH are used to generate stable negative ion electrospray. Furthermore, halogenated solvents can suppress corona discharge, which creates significant background in negative ESI spectra and poor spray stability.

The characteristics of analyte can also affect ESI response. There are four types of methods to charge the analyte in ESI: charge separation, formation of adducts, gas-phase reaction and electrochemical oxidation and reduction. Charge separation is the primary method in ESI. For example, proteins or peptides form positive charges by protonation in the positive ion mode and oligonucleotides form negative charges by deprotonation in the
negative ion mode. Adduct formation also occurs in the positive and negative ion modes. For example, cationic adducts, such as sodium or potassium, are used to analyze carbohydrates in the positive ion mode and anionic adducts, such as chloride, are used in the negative ion mode. Low concentrations of salt can help ionization by adduct formation; however, high salt concentration usually suppresses analyte signal. Gas-phase reactions, such as gas-phase proton-transfer, charge the analyte in the gas phase. Gas-phase basicity is in terms of gas phase proton affinity\textsuperscript{92}. Analytes with high gas phase proton affinity scavenge a proton from species with low gas phase affinity; and thus dominate ESI spectra\textsuperscript{92}. Furthermore, if the solvent used has a higher gas phase proton affinity than the analyte, the signal of the analyte will be suppressed in the positive ion mode. Electrochemical oxidation and reduction can also generate charges for analyte to help ionization in ESI.

To date, nano-ESI is becoming much more popular due to its higher sensitivity and lower sample requirement, compared to conventional ESI\textsuperscript{93}. Nano-ESI uses low flow rates (about 100 nL/min), which further reduce droplet size. Furthermore, nano-ESI can analyze complex samples such as carbohydrates and glycoproteins, which are not particularly well suited for conventional ESI analysis\textsuperscript{93}. These surface-active compounds, which stay on the surface of primary droplets, can be enriched during offspring droplets in nano-ESI, thus enhancing the sensitivity \textsuperscript{93}. In addition, nano-ESI can maintain stable spray at low electrical voltage, thus avoiding electrical discharges that may occur at high voltage as used in conventional ESI\textsuperscript{93}.

1.3.2.2 Ion Activation
The precursor ions transferred in ESI provide the mass information of the analyte. However, this alone is not sufficient for the complete determination of the structure of analyte. Ion activation results in fragmentation of the precursor ion into product ions during tandem mass spectrometry. The product ions are required to create a fingerprint for analyte. In this thesis, the most commonly used methods for ion activation: collision-induced dissociation (CID) and electron transfer dissociation (ETD) are discussed.

### 1.3.2.2.1 Collision-Induced Dissociation (CID)

In CID, activation of ions occurs by collisions with inert gases, such as nitrogen and helium, and the energy is transferred to the precursors as vibration. Thus, the excited precursor ions are decomposed into product ions. There are two types of CID based on the energy: low energy CID (1-100 eV) and high energy CID (>keV). Low energy CID occurs in ion trap, FTICR or quadrupole analyzers while high energy CID usually uses magnetic or time of flight analyzer. In this thesis, CID fragmentation of peptides and oligosaccharides is discussed.

The nomenclature of sequenced ions after peptide fragmentation is shown in Figure 1-8. The peptide dissociation in CID is based on the mobile proton model. In positive ESI, peptides are protonated and the protons are attached to the basic residues. After the collision with inert gases, mobile protons are transferred to the amides in the peptide backbone. The corresponding carbon center is attacked by the oxygen of the N-terminal neighbor amide bond to form the direct sequence ions of b and y, as shown in Figure 1-9. The b and y sequence ions can be further dissociated to produce the internal fragments. Peptide fragmentation is affected by the charges on the peptide. In general,
double or multiple charged peptides easily produce more fragments because of more
dissociation pathways and stronger Coulomb repulsion\textsuperscript{77}.

CID fragmentation of oligosaccharides is discussed in three categories: low energy
CID with the positive ionization mode, high energy CID with the positive ionization mode
and CID with the negative ionization mode. The nomenclature of product ions is shown in
Figure 1-10\textsuperscript{96}. In low energy CID with positive ionization, the major fragments are B- and
Y-type glycosidic ions, which provide important information for composition and sequence
of carbohydrates\textsuperscript{77}. However, there are no cross-ring fragments as the energy is not
sufficient for cleavage of two covalent bonds. In contrast, cross-ring fragments can be
produced in high energy CID with the positive ionization mode. However, the cross-ring
fragments are reduced as lower collision energy is used. Metal cations such as sodium or
magnesium are used to form adducts, thus facilitating the ionization and fragmentation in
high energy CID. Generally, there are three pathways for metal-cation adduct carbohydrate
fragmentation: metal loss, glycosidic bond cleavage and cross-ring cleavage, as shown in
Figure 1-11\textsuperscript{97}. In CID with negative ionization, cross-ring fragments and glycosidic ions are
produced in either deprotonated or anion adduct forms. One example of the cross-ring
fragmentation mechanism (\textsuperscript{2,4}A ion) is shown in Figure 1-12\textsuperscript{98}. The cross-ring
fragmentation is initiated with deprotonation of the hydroxyl at the 3-position of the
reducing end of the monosaccharide, and then electron shifts lead to the cross-ring
cleavage\textsuperscript{98}. Cross-ring fragments cannot be created when a hydrogen at 3-position is absent,
i.e. if the carbon 3 forms part of a glycosidic bond\textsuperscript{99}. However, full determination of
linkage and positional isomers of oligosaccharides still remains challenging for mass
spectrometry.
Figure 1-8. Formation and interpretation of a, b, c, x, y and z fragments in peptides.

Figure 1-9. Formation of b and y ions in peptide dissociation\textsuperscript{95}.

Figure 1-10. Formation and interpretation of A, B, C, X, Y and Z fragments in oligosaccharides\textsuperscript{96}.
Figure 1-11. Three major pathways for oligosaccharides with metal adducts.

Figure 1-12. The mechanism for $^{2,4}$A cross-ring fragments in negative ion ESI analysis of oligosaccharides.

1.3.2.2.2 Electron Transfer Dissociation (ETD)

Although CID is the most commonly used fragmentation method, it has limitations for the characterization of protein post-translational modifications. The labile modification is usually the preferred site for cleavage during CID fragmentation; and thus sufficient peptide sequence and modification site information will not be generated. In contrast to
CID, electron capture dissociation (ECD) is a low energy reaction and does not involve the energy redistribution, thus the labile PTMs can be preserved\(^{100}\). However, ECD requires expensive Fourier transform ion cyclotron resonance (FTICR) instrumentation to store thermal electrons, thus it is not suitable for widely used instruments such as a quadrupole linear ion trap.

Similar to ECD, ETD is an alternative fragmentation method in which electrons (4-5.5 eV) are transferred to the protonated peptides by anions as the electron donors and then the electron shift and rearrangement result in the peptide backbone cleavage of the N-C\(\alpha\) bond to produce c and z ions as shown in Figure 1-8\(^{101,102}\). However, ETD analysis is not suitable for all peptides. The decision tree for ETD analysis, as reported by Coon and co-workers, recommends using ETD fragmenting peptides with m/z ≤ 850 and z≥3\(^{103}\). Supplemental activation is used to facilitate ETD efficiency, especially for doubly charged peptides\(^{103,104}\). In this process, the non-dissociated ETD product ([M+2H]\(^{++}\)) is collided with inert gas, as used in low energy CID, to dissociate the charge-reduced species. Another approach to improve dissociation of charge-reduced species was developed by our group: CID of the isolated ETD product ([M+nH](n-1)+)\(^{105}\). This strategy is widely used to characterize protein post-translational modifications including phosphorylation, glycosylation and disulfide linkages with less neutral loss than that in the supplemental activation process\(^{105-108}\). In this thesis, application of ETD to determine isoaspartic acid and disulfide linkages is presented in the following chapters.

**1.3.2.3 Mass Analyzers and Tandem Mass Spectrometry**
After gas-phase ions are produced in the ion source, the ions need to be sorted by their mass to charge ratios, which occurs in the mass analyzer. There are several types of mass analyzers and all have their own advantages and limitations. Furthermore, tandem mass spectrometry has been developed by combining different analyzers together to meet the requirements for different applications. There are several critical properties of analyzers: resolution, accuracy, mass range, linear dynamic range and speed. Resolution is the ability to differentiate two peaks with a slight mass difference. Mass accuracy is the ratio of error between the measured mass and theoretical mass. Mass range is the mass to charge (m/z) range amenable for a specific analyzer. Linear dynamic range is a range in which the signal is linear to a certain concentration. Scan speed is the time required for the determination of spectra. Transmission is the ratio of the ions arriving at the detector to the ions entering the analyzer. The comparison of these characteristics of commonly used mass analyzers is listed in Table 1-4. Here, a brief introduction to mass analyzers and tandem mass spectrometry instrumentation is presented.
### Table 1-4. Comparison of commonly used mass spectrometric instrumentation\(^\text{109}\).

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Ion source</th>
<th>Mass accuracy (ppm)</th>
<th>resolution</th>
<th>Mass range</th>
<th>Dynamic range</th>
<th>sensitivity</th>
<th>speed</th>
<th>MS/MS capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTQ</td>
<td>ESI</td>
<td>500</td>
<td>2000</td>
<td>50-2000,200-4000</td>
<td>1E4</td>
<td>fmol</td>
<td>Fast</td>
<td>(\text{MS}^{\text{ind}})</td>
</tr>
<tr>
<td>Q-q-Q</td>
<td>ESI</td>
<td>1000</td>
<td>1000</td>
<td>10-4000</td>
<td>6E6</td>
<td>amol to fmol</td>
<td>Moderate</td>
<td>(\text{MS}^2)</td>
</tr>
<tr>
<td>Q-q-LIT</td>
<td>ESI</td>
<td>500</td>
<td>2000</td>
<td>5-2800</td>
<td>4E6</td>
<td>fmol</td>
<td>Fast</td>
<td>(\text{MS}^{\text{ind}})</td>
</tr>
<tr>
<td>TOF</td>
<td>MALDI</td>
<td>20</td>
<td>20000</td>
<td>No upper limit</td>
<td>1E4</td>
<td>fmol</td>
<td>Fast</td>
<td>n/a</td>
</tr>
<tr>
<td>TOF-TOF</td>
<td>MALDI</td>
<td>20</td>
<td>20000</td>
<td>No upper limit</td>
<td>1E4</td>
<td>fmol</td>
<td>Fast</td>
<td>(\text{MS}^2)</td>
</tr>
<tr>
<td>Q-q-TOF</td>
<td>ESI/MALDI</td>
<td>20</td>
<td>20000</td>
<td>No upper limit</td>
<td>1E4</td>
<td>fmol</td>
<td>Moderate to fast</td>
<td>(\text{MS}^2)</td>
</tr>
<tr>
<td>FTICR</td>
<td>ESI/MALDI</td>
<td>2</td>
<td>750000</td>
<td>50-2000,200-4000</td>
<td>1E3</td>
<td>fmol</td>
<td>slow</td>
<td>(\text{MS}^{\text{ind}})</td>
</tr>
<tr>
<td>Orbitrap</td>
<td>ESI/MALDI</td>
<td>5</td>
<td>1000000</td>
<td>50-2000,200-4000</td>
<td>4E3</td>
<td>fmol</td>
<td>Moderate to fast</td>
<td>(\text{MS}^{\text{ind}})</td>
</tr>
</tbody>
</table>
1.3.2.3.1 Quadrupole

The quadrupole analyzer consists of four parallel rods on which direct and radiofrequency (rf) voltages are applied to separate ions based on their m/z values. Only ions with defined m/z values can travel through the region between the quadrupole rods to arrive at the detector while other ions collide with the rods\textsuperscript{110}. This characteristic helps to select the ions with specific m/z values and scan an m/z range by varying the voltage. Furthermore, a quadrupole is frequently used as the ion guide or collision cell in tandem mass spectrometers because of its ability to focus ion trajectory. Similar to the quadrupole, a hexapole (consisting of six rods) and an octopole (consisting of eight rods) have been developed as ion guides and collision cells.

1.3.2.3.2 Ion Trap

The operation of the ion trap is similar to that of the quadrupole mass analyzer. The ion trap, as the name suggests, can trap ions and then eject them sequentially according to their m/z ratio via an oscillating electric field\textsuperscript{110}. The major advantage of an ion trap is to perform tandem mass spectrometry (MS\textsuperscript{n}) because ions are trapped both in space and time. To facilitate fragmentation, precursor ions are trapped in the ion trap and collided with a low pressure inert gas. Resulting product ions are either expelled from the trap to the detector for signal analysis or trapped further and sequentially collided with inert gas over several cycles to generate MS\textsuperscript{n} spectra. Such multi-stage tandem MS spectra help with structural characterization and elucidation of analytes. Furthermore, the combination of an ion trap with ETD can facilitate fragmentation by transferring electrons from the ETD reagent ion.
source to the trapped analytes. Ion traps can also be utilized as the front end of hybrid high mass accuracy instruments including the orbitrap and FTICR mass analyzers.

1.3.2.3.3 Time of Flight (TOF)

A TOF separates ions based on the time taken by an analyte to fly through the analyzer, which is proportional to its m/z ratio. Larger analytes take a longer time to reach the detector compared to smaller analytes. The resolution of the TOF mass analyzer can be improved by delayed pulsed extraction or by the inclusion of reflectrons\textsuperscript{110}. Both methods can correct the kinetic energy dispersion, and thus improve resolution\textsuperscript{110}. TOF requires a bundle of ions and thus is suitable for MALDI. A TOF is also used for tandem mass spectrometry such as TOF/TOF and Quadrupole-TOF. The advantages of a TOF are high sensitivity due to high transmission efficiency, fast scan speed and broad mass range.

1.3.2.3.4 Orbitrap

The orbitrap mass analyzer was recently introduced and offers many advantages, including high mass accuracy and high resolution comparable to FTICR\textsuperscript{111,112}. The orbitrap is a trapping analyzer. Product ions are trapped by an electrostatic field and forced to travel in a spiral pattern, instead of using radiofrequency or magnetic fields\textsuperscript{111,112}. Orbitraps are usually used in tandem with an ion trap for tandem mass spectrometry purposes. The ETD-LTQ-Orbitrap is a commercial instrument by Thermo. It consists of several major parts: an ion source, a linear ion trap, a C-trap with an RF-only quadrupole capable of high energy CID, an orbitrap and an ETD reagent ion source\textsuperscript{111}. To date, the Orbitrap has been widely used for proteomics and PTM analysis in both bottom-up and top-down strategies. In this
thesis, determination of isoAsp and disulfide linkage using Orbitrap with ETD is presented in the following chapters.

1.3.2.3.5 Fourier Transform Ion Cyclotron Resonance (FTICR)

In FTICR, charged ions are trapped in vacuum by combination of a constant magnetic and electric fields\textsuperscript{110}. The charged ions oscillate around the magnetic field with a cyclotron frequency that is inversely proportional to their m/z ratios\textsuperscript{110}. Similar to the Orbitrap mass analyzer, FTICR can also utilize a linear ion trap in the front end to form a hybrid instrument. FTICR is compatible with ECD technology as it is able to trap and store thermal electrons. The FTICR mass analyzer offers high resolution and mass accuracy and is a widely used technology for mass spectrometry based proteomics although the high cost of instrumentation and maintenance limits the application.

1.3.2.3.6 Quadrupole-Time-of-Flight (Q-TOF)

Q-TOF has the generic name of QqTOF, which consists of three main parts: Q1, Q2 and the TOF. In the Agilent Q-TOF instrument, which is used for glycan analysis in our projects, the first major part is Q1, a mass-resolving quadruple filter, which is operated to transmit the precursor ions of interest in MS/MS mode. However, when it is operated in RF-only mode, it serves as a transmission element in single MS mode. An octopole is placed before Q1, which serves as the ion guide to cool down and focus the ions. The second major part is Q2, which is replaced by a hexapole in the Agilent Q-TOF. Q2 serves as a collision cell, where the precursor ions are collided with neutral gas and produce fragment ions that are then cooled and focused by another octopole after Q2. The third major part is the TOF analyzer, which can record all ions in the spectrum, thus providing
high sensitivity, high resolution and high mass accuracy for both precursor and product ions.
To date, the Q-TOF is widely used in proteomics and small molecule analysis. In this thesis, characterization of glycan structure using Q-TOF is described in Chapter 5.

1.3.2.3.7 **Triple Quadrupole (QQQ)**

Similar to the Q-TOF, triple quadrupole (QQQ) also consists of three main parts: Q1, Q2 and Q3 while Q3 is replaced by a quadrupole in QQQ instead of a TOF in the Q-TOF. Q1 and Q2 serve as the mass filter and the collision cell, respectively, and Q3 is a mass analyzer for recording the fragment ions. The QQQ provides several advantages, such as precursor ion, neutral loss and product ion scans. In product ion scans, Q1 is set to separate or filter the specific precursor ions, which are further fragmented by collision with neutral gas in Q2. The fragment ions are transported to Q3 and are set to scan for all ions. In precursor ion scan, a fixed product ion m/z is set in Q3 and Q1 is set to scan all the precursor ions which are able to fragment into the fixed product ions. In neutral loss, Q1 and Q3 are set to scan with a constant mass offset. QQQ is now widely used for quantitation of peptides and small molecules using selected reaction monitoring (SRM), also called multiple reaction monitoring (MRM). In SRM, a pair of predefined precursor ions and one or several major products ion from the precursor ion are scanned by Q1 and Q3 over a specific time window to generate chromatographic traces with retention time and signal intensity. There is no full scan event in QQQ-based SRM and this non-full-scan mode increases the sensitivity by one or two orders of magnitude. In addition, SRM can result in a wide dynamic range of linear response, which facilitates the quantitation of low amounts of analyte even in a complex biological sample.
1.3.2.4 **Quantitative Mass Spectrometry**

In addition to proteomic characterization, quantitative assays in proteomics are also critical for evaluation and verification of differential expression and modification of proteins. The development of a quantitative mass spectrometry strategy is a challenging topic in the field. Currently, two major strategies have been developed for protein quantitation: label-free quantitation and isotope labeling quantitation. These two strategies are briefly discussed in this section.

1.3.2.4.1 **Label Free Quantitation**

Label free quantitation is a protein quantitation strategy based on the mass spectrometry signal without isotope labeling. There are several methods commonly used for label free quantitation, such as selected reaction monitoring, ion intensity and spectrum counting. Since label free methods do not require additional sample preparation for isotope labeling, they reduce sample loss and the overall assay cost. However, as the signal response factor (counts/mol) varies among different samples, only similar peptides or analytes with similar response factors on the mass spectrometer are comparable. In addition, this method is unable to correct for variation introduced from differences in sample preparation. To overcome the above limitations, a strategy involving the combination of selected reaction monitoring on a QQQ has been developed for relative and absolute quantitation. When an internal standard is used in selected reaction monitoring, a calibration curve can be established and absolute quantitation can be achieved. This strategy combines the liquid chromatography retention time and the mass-to-charge of precursor and product ions to generate time-based MS/MS transitions, thus eliminating false positive results and increasing sensitivity. Selected reaction monitoring can also be
performed in an ion trap instrument. We use selected reaction monitoring for verification and quantitation of protein PTMs which is discussed in the following chapters.

1.3.2.4.2 Isotope Labeling Quantitation

Isotope labeling strategies are widely used in quantitative proteomics, which include metabolic labeling, stable isotope labeling by amino acids (SILAC), proteolytic labeling and chemical labeling using isotope coded affinity tags (ICAT), isobaric tags for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT). Here, the brief introduction to these methods is described.

Metabolic labeling: In this strategy, a stable isotope is added into cell culture medium and incorporated into the newly synthesized proteins during cell growth. $^{15}$N labeling is commonly used in autotrophic cells. The advantage is that the isotopic incorporation rate is relatively high, often more than 90% $^{113}$. The disadvantage is that only two samples can be compared in a single experiment$^{114}$. Automatic data analysis is also complicated due to difficulties in predicting the mass shift in different peptides by software.

SILAC: In SILAC, isotopic labeled amino acids are incorporated into protein. The commonly labeled amino acids are lysine and arginine as the C-termini of these two amino acids can be cleaved during trypsin digestion and thus, the C-terminus of each tryptic peptide is labeled. The advantages of SILAC compared to metabolic labeling are that it can compare more than two samples in one experiment run and the mass shift can be predicted.

Proteolytic labeling ($^{18}$O): In this strategy, $^{18}$O is incorporated during digestion. For example, for trypsin digestion in $^{18}$O water, one or two $^{18}$O atoms are incorporated into the peptide C-terminus resulting in a 2 or 4 Dalton mass shift. The disadvantage is that the incorporation of 1 or 2 $^{18}$O atoms can complicate the quantitation. Back exchange of $^{18}$O
with $^{16}$O during the subsequent analysis is also an issue. $^{18}$O labeling has also been
developed for monitoring artificial deamidation induced by sample preparation\textsuperscript{115}. The
artificial deamidation due to sample treatment shows a 3 Dalton mass shift while the
intrinsic deamidation has a 1 Dalton mass shift compared to the non-deamidated peptide.

**ICAT:** This method consists of the use of a cysteine reactive moiety
(iodoacetamide), an isotope linker (0 or 8 deuterium atoms), and an affinity group (biotin) as
shown in Figure 1-13 (A)\textsuperscript{114}. Reduced cysteines in proteins react with iodoacetamide to
form the labeled proteins. Labeled proteins are then purified by exploiting the high natural
affinity of avidin for biotin. The labeled protein can be pooled for digestion with the
advantage of normalization correction for sample loss during digestion thereby increasing
the overall reproducibility. The disadvantages are that only two samples can be compared
in one experiment and this method cannot be used for quantitation of proteins without
cysteine residues.

**TMT and iTRAQ:** TMT and iTRAQ reagents also consist of three parts: the peptide
reactive group (amine specific), the reporter mass group and the mass balance group as
shown in Figure 1-13 (B) and (C)\textsuperscript{114}. Proteins are digested and then labeled \textit{via} the amine
group in lysine or the N-terminus. In the MS spectrum, differently labeled peptides have an
identical mass. During MS/MS spectrum generation, the resulting intensity of reporter ions
determines the abundance of samples. The advantage is that multiple samples, up to 16, can
be compared in a single experiment. The disadvantage is the variability during labeling and
digestion\textsuperscript{114}. 


Figure 1-13. Chemical structure of (A) ICAT, (B) ITRAQ and (C) TMT.

1.4 Applications of LC and MS Based Techniques in the Determination of PTMs

Liquid chromatography interfaced to mass spectrometry is widely used for characterization and quantitation of proteins in proteomics due to its high sensitivity, specificity, low sample consumption and robustness. LC-MS also proves extremely useful in determination of PTMs. The physiological and chemical changes due to modifications aid in the separation of peptides with different modifications. Furthermore, the mass shift due to PTMs in proteins can be detected by mass spectrometry. Mass spectrometry is able
to determine the identity of modified proteins, specific sites of modification, the modification structure and the modification occupancy. Furthermore, mass spectrometry has the capability to identify PTMs in complex samples and the discovery of novel PTMs. The following section discusses the determination of isoaspartic acid, disulfide linkages and glycosylation using liquid chromatography and mass spectrometry.

1.4.1 Identification of Deamidation of Asparagine and Isomerization of Aspartic Acid

The formation of isoAsp resulting from Asn deamidation or Asp isomerization is a common protein PTM. However, differentiation of Asp and isoAsp is still a challenging task due to the similar structures and identical molecular mass. So far, the main approaches for characterization and quantitation of Asn, Asp and isoAsp at the peptide or protein level are enzymatic, separation and mass spectrometric methods. Protein L-isoaspartyl methyltransferase (PIMT, PCMT) can convert isoAsp residue to L-Asp residue as a protein repair process, which has been utilized for an isoAsp detection assay as used in the ISOQUANT detection kit (Promega, Madison, WI) for global analysis of isoAsp. As shown in Figure 3-1, PIMT catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to isoAsp, generating the labile isoaspartic methyl ester, which spontaneously hydrolyzes to produce methanol and succinimide. In the radioactive format of ISOQUANT, the methyl group donor, SAM, is isotopically labeled with tritium and the resulting byproduct of $[^3]H$ methanol is used for isoAsp quantitation. In the HPLC format of ISOQUANT, a byproduct of S-adenosyl homocysteine(SAH) is separated by reverse phase HPLC and then quantified using a standard curve. However, this method has several limitations. First, the location of isoAsp sites cannot be achieved. Second, false positive
results may be generated by interference from certain substance such as sodium dodecyl sulfate (SDS). Finally, it usually requires a relatively high amount (at least pmol) of samples for the experiment due to the limited sensitivity. Therefore, more sensitive and site-specific analytical strategies such as liquid chromatography and mass spectrometry are preferred for the characterization and quantitation of these isomers.

The peptides or proteins with Asn, Asp and isoAsp have slight differences in structure and pI values. The pKa value of the Asp side chain is about 3.9 while the side chain of isoAsp has a pKa value of 3.1\textsuperscript{116}. Although the pKa value of the Asn side chain is not known, it is expected to be higher than that of Asp due to the presence of the amine group. A strategy based on the difference in pKa values has been used for the gel separation of the isomers. However, the application of this strategy is limited by its low resolving power. Ion exchange, such as cation exchange, has been reported to successfully separate the isoforms of monoclonal IgG based on the charge related heterogeneity resulting from isoAsp\textsuperscript{117}. In addition, the separation of the triad isomers based on slight structural differences can be achieved by reversed phase HPLC. However, due to the similar properties, separation of the isomers is usually time-consuming and a baseline separation is difficult to achieve. Furthermore, the elution order varies with different liquid chromatography conditions\textsuperscript{7,118}. For example, ion pairing reagents can change the elution order.

Mass spectrometry is a powerful analytical tool due to its high sensitivity, speed and low consumption of samples. FTICR, orbitrap and Q-TOF mass analyzers with high mass accuracy and high resolution can be used for deamidation detection by measuring the
+0.984 Dalton mass shift from Asn to Asp/isoAsp. However, it is difficult to differentiate the deamidated products or isomerization of aspartic acid and isoaspartic acid as they have an identical mass. Different ionization and dissociation methods such as positive CID/HCD and negative ion CID/HCD have been studied to differentiate Asp and isoAsp, but all fail to provide specific and reliable diagnostic ions. O’Connor et al have successfully developed ECD and ETD based methods to differentiate isoAsp and Asp\textsuperscript{119,120}. As shown in Figure 2-2, one pair of diagnostic ions (c+57, z-57) for isoAsp are generated by cleavage of the Ca-Cβ bond in isoAsp by ECD or ETD\textsuperscript{7,121}. Furthermore, the diagnostic ions for N-terminal isoAsp were also reported by O’Connor et al\textsuperscript{122}. However, high throughput analysis of isoAsp is still challenging due to the relative low intensity of diagnostic ions (c+57, z-57). Yang et al has combined other criteria such as chromatographic parameters to verify the results, thus avoiding certain false positive results that may be generated using only the diagnostic ions (c+57, z-57) in studying isoAsp at the proteome level\textsuperscript{72}.

To date, isoAsp characterization in mass spectrometry is usually performed by a bottom-up approach. However, this approach can generate artificial isoAsp residues induced by sample preparation. Sample preparation induced isoAsp can be monitored by \textsuperscript{18}O labeling during sample preparation by using \textsuperscript{18}O water instead of \textsuperscript{16}O water\textsuperscript{7,123}. Deamidation in \textsuperscript{18}O water leads to a +3 Dalton mass shift (2 Dalton from \textsuperscript{18}O labeling and 1 Dalton from the mass difference of the amino and hydroxyl group in Asn and Asp/isoAsp, respectively). Isomerization in \textsuperscript{18}O water leads to +2 Dalton mass shift. However this approach is complicated by 1 or 2 \textsuperscript{18}O labeling at C-terminus during the digestion. Recently, Liu et al have modified the detection method by measuring the b series of ions to eliminate the complexity induced by C-terminal \textsuperscript{18}O labeling\textsuperscript{124}. Unlike the bottom-up
approach, the top-down approach can also eliminate artificial results caused by sample preparation. Recently, O’Connor and co-workers used top-down ECD to study intact β2-microglobin\textsuperscript{125}. In this work, two different methods were used: ECD-MS\textsuperscript{2} and nozzle-skimmer dissociation (NSD)-ECD-MS\textsuperscript{3}. Three deamidation sites (Asn17, 42 and 83) were detected using the accurate mass of fragment ions in ECD-MS\textsuperscript{2}. One isoaspartic acid reporter ion ($c_{16}+57$ of isoAsp17) was observed to confirm the formation of isoAsp. However, no isoaspartic acid reporter ions at the other two sites were detected in ECD-MS\textsuperscript{2}. In the NSD-ECD- MS\textsuperscript{3} method, the protein underwent fragmentation by nozzle-skimmer dissociation and then the fragments of interest were selected for further fragmentation by ECD. The diagnostic ions of isoaspartic acid for all three deamidation sites (Asn17, 42 and 83) were observed and thus the formation of isoAsp was verified\textsuperscript{125}. Furthermore, this method is also suited for quantitative analysis of isoAsp formation using a calibration curve. Despite the aforementioned improvements in methodology, the top-down strategy is still challenging for the characterization of deamidation and its products in large proteins.

1.4.2 Determination of Disulfide Linkage

Determination of cysteine status in proteins includes free cysteine, modified cysteine and disulfide linkages. Determination of free or modified cysteine is determined by peptide mapping technology. However, free cysteine is highly reactive and readily participates in disulfide scrambling during sample preparation. Acidic conditions, such as pepsin digestion at pH 2, are preferred for sample digestion as cysteine is protonated and unable to participate in scrambling at low pH. Furthermore, 3-D protein structure is disrupted in strong acidic condition and thus easy to be digested. Another way to reduce scrambling is to alkylate the free cysteine before digestion in acidic conditions. Quantitation of total free
cysteine can be performed using Ellman’s reagent (also known as 5,5'-dithiobis(2-nitrobenzoate) DTNB). As shown in Figure 1-14, free cysteine reacts with DTNB to form a disulfide bond and 2-nitro-5-thiobenzoic acid (TNB). The amount of free cysteine can be estimated by measuring TNB absorbance at 412nm.

![Diagram of DTNB and TNB](image)

**Figure 1-14.** Reaction of free cysteine with DTNB.

The determination of disulfide linkages is still analytically challenging, especially for cysteine-rich domains. The approaches applied for characterization of disulfide bonds include X-ray of crystal structures, nuclear magnetic resonance (NMR), Edman sequencing, LC separation and mass spectrometry based methods. Crystallization is time-consuming and often met with limited success. NMR data can sometimes confuse disulfide assignment if cysteines are in close proximity in cysteine-rich domains. Edman sequencing requires a relatively large amount of purified proteins and the method is time consuming. To date, LC-MS has become a routine method to determine disulfide linkage due to its high sensitivity and low requirement of sample purity and amount. The determination of disulfide linkages can be performed by comparison of the LC chromatograms and mass spectra with and without reduction and alkylation. Linkage assignment can be performed
by matching the experimental mass and theoretical mass of the proposed disulfide linkages.

Tandem mass spectrometry is now more popular for disulfide characterization. The fragments are not only able to generate the peptide sequence information but also produce certain signature ions for disulfide linkage assignment. The disulfide bond is a strong bond with dissociation energy of 251 kJ/mol, about 20-fold higher than the dissociation energy of amide bond. Thus, disulfide linkages are not able to be broken by CID fragmentation as CID usually breaks the weakest bond in peptides. This limits the application of CID in characterization of complex disulfide linkages. Partial reduction and stepwise alkylation are also used to characterize complex disulfide linkages such as those in cysteine knot structures. The resulting different species after partial reduction and stepwise alkylation are then separated by reversed phase HPLC and detected by CID mass spectrometry. In contrast to CID with positive ion mode, CID in negative ion mode can dissociate disulfide linkages. The proton in CαH or CβH of the cysteine residue is abstracted in negative ion and then an α/β elimination reaction occurs to generate the dissociated peptides. However, there is no peptide sequence information generated in this method. Thus, it is not commonly used in disulfide characterization. Some other efforts have been made for resolving disulfide linkages, such as $^{18}$O or dimethyl labeling or ultraviolet photo dissociation. However, these methods are not suitable for characterization of intertwined disulfide linkages.

A more recent breakthrough is to use ECD and ETD to analyze disulfide bonds, which were shown to be preferentially cleaved over the peptide backbone. Wu et al established the method of using LC-MS with a combination of ETD and then CID of isolated charge-reduced ions (CRCID or CID-MS3) to detect complicated and intertwined
disulfides, such as those in recombinant tissue plasminogen activator and the scrambling disulfide linkages in monoclonal antibodies\textsuperscript{106-108}. In ETD, electrons prefer to be transferred to sulfur groups and thus lead to the breakage of disulfide linkages. The major peaks are charge reduced species, dissociated peptide peaks and c and z fragments. Further fragmentation of dissociated peptide fragments using CID can provide more sequence information. The charge reduced species consists of two populations: peptide dissociated species and disulfide dissociated species. These peptide or disulfide dissociated species are still held together by non-covalent forces\textsuperscript{106}. Further CID of the charge reduced species or the use of supplemental activation during ETD results in further peptide backbone cleavage and thus more sequence information.

1.4.3 Characterization of Glycosylation

Proteomics analysis of glycosylation is one of the most challenging fields as it requires achievement of the following goals: identification of glycoprotein, specific modification site, occupancy of glycosylation, and glycan structure including the composition and linkage information. To achieve these goals, orthogonal techniques are employed such as NMR, gas phase or liquid phase chromatography, capillary electrophoresis and mass spectrometry. This thesis focuses on glycosylation analysis by liquid chromatography and mass spectrometry.

In general, analysis of glycosylation can be divided into three categories: characterization of glycans in intact glycoproteins, analysis of glycopeptides, and structural analysis of native, reduced or derivatized glycans. The commonly used approaches for each category are summarized in Table 1-5 \textsuperscript{17}. In this section, the strategies for analysis of intact glycoproteins, glycopeptides and glycans will be discussed in detail.
Table 1-5. Strategies for glycosylation analysis.\textsuperscript{17}

<table>
<thead>
<tr>
<th>Type</th>
<th>Strategy</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intact glycoprotein</td>
<td>CE, LC-MS/MS, lectin analysis</td>
<td>Glycoform profiling</td>
</tr>
<tr>
<td>analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycopeptides analysis</td>
<td>CE, LC-MS/MS (CID,ETD,ECD)MALDI-MS/MS</td>
<td>Glycosylation site, compositional glycan structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaease glycan analysis</td>
<td>CE, HPLC (HILIC, WAX, RP, PGC), LC -MS /MS</td>
<td>compositional glycan structure, detailed glycan structural analysis</td>
</tr>
<tr>
<td></td>
<td>(CID,ETD,ECD), MALDI-MS/MS, MALDI-TOF</td>
<td></td>
</tr>
</tbody>
</table>

1.4.3.1. Intact Glycoprotein Analysis

Glycosylation analysis of intact glycoproteins usually can provide information on glycoform profiling. However, the site and structural information of glycans are usually hard to obtain due to the limitation of instrumentation.

Lectins are a group of proteins that have specific affinity for glycans, such as concanavalin A (Con A), wheat germ agglutinin (WGA) and jacalin lectin (JAC). Lectin affinity chromatography has been extensively used for glycoprotein characterization. Almost all glycoproteins can be recovered using lectins with broad specificities or a multi-lectin column. Also, the glycoforms can be differentiated using lectins with narrow
specificities. Lectin arrays consisting of 35 different lectins have been used for qualitative and quantitative analysis of glycoforms with high sensitivity and high throughput capabilities.

By virtue of the high resolution and high mass accuracy mass spectrometers, such as Q-TOF, Orbitrap, FTICR and Q-Exactive have been used to facilitate the top down analysis of intact glycoproteins. For example, FTICR has been used for glycosylation profile analysis of recombinant human chorionic gonadotrophin, in which many different glycoforms were observed and the mass difference of major peaks was also reported.

1.4.3.2 Glycopeptide Analysis

Since the analysis of intact glycoproteins provides limited information on glycosite and glycan composition, characterization of glycopeptides is more commonly used for analysis of peptide sequence, glycan attachment site and the glycan structure. Glycopeptides are generated by digestion of glycoproteins with a protease such as trypsin and then analyzed by RPLC-ESI-MS/MS using CID, ECD or ETD or by MALDI mass spectrometry. However, the sensitivity of glycopeptide analysis is relatively low due to their low ionization behavior and glycoform heterogeneity of glycopeptides with low concentrations. Thus, enrichment of glycopeptides is usually employed before mass spectrometry analysis. To date, multiple strategies have been developed for glycopeptide enrichment, such as lectin affinity, size exclusion chromatography, HILIC, sepharose, PGC and hydrazide bead enrichment.

After enrichment, glycopeptides are usually resolved by reversed phase HPLC. Glycopeptides usually elute earlier than the corresponding non-glycosylated peptides due to the high polarity induced by the hydroxyl group of the glycan. CID mass spectrometry, a
common fragmentation method in peptide mapping, is used extensively for glycopeptide analysis. In the low energy CID with the positive ion mode, the major peaks present in the spectra correspond to loss of monosaccharide residues resulting in B- and Y-type glycosidic ions. However, peptide sequence information is not available because the peptide backbone remains intact and unfragmented. Furthermore, rearrangement of monosaccharides is also observed, which complicates data interpretation \textsuperscript{135}. Unlike low energy CID, peptide backbone cleavage of b and y ions dominate in high energy CID using positive ion mode. However, these b and y ions are observed in their deglycosylation form of the peptide or by only retaining the N-linked N-acetylglusamine; thus glycan structural information is not retained \textsuperscript{136}. In contrast, analysis of glycopeptides using CID in negative ion mode can provide structural information of glycan and peptide sequence and linkage by glycosidic and cross-ring fragments of glycans and peptide backbone cleavage \textsuperscript{137}. Furthermore, detection of glycopeptides with sialic acid is more sensitive in negative ion mode than in positive ion mode \textsuperscript{137}. ECD or ETD is an alternative fragmentation strategy to characterize glycopeptides, in which fragmentation creates peptide backbone cleavage of c and z ions with an intact glycan moiety. This fragmentation provides complementary information to that generated in CID which generates glycan composition information. Thus, a combination of CID and ECD or ETD can produce both peptide and glycan information.

1.4.3.3 Glycan Analysis

The analysis of released glycans provides more information about glycan structure although it usually loses information on glycosylation sites. The released glycans can be grouped into three types: native glycan (with reducing end, also called aldose), alditol (with reduced end) and derivatized glycan as shown in Figure 1-15.
Native glycans are usually released by digestion with enzymes such as PNGase F digestion. Alditols are usually generated by the chemical release of glycans such as β-elimination or reduction of the native glycans. These glycans are relatively hydrophilic and are not retained well on reversed phase sorbents. However, they are well suited for HILIC and PGC separation. PGC has another advantage in glycan analysis as it has strong resolving power for isomer separation. For example, anomers and linkage isomers in native glycans can be separated by PGC. Separation on PGC is affected by the mobile phase composition such as pH and ionic strength. Certain ionic strength in mobile phase is required for separation of charged oligosaccharides such as sialylated or sulfated glycans. The native glycans or alditols are usually detected by mass spectrometry after online or offline separation. In CID with positive ion mode, glycosidic B- and Y-type ions are the major fragments, which provide the glycan composition and sequence information. However, no linkage information can be obtained. In CID using the negative ion mode, A- and X-type cross-ring fragments in addition to glycosidic C ions are generated; thus both composition and linkage information can be annotated. Furthermore, loss of sialic acid,
which has been observed in MALDI analysis using positive ionization mode, has not been observed in negative mode analysis \(^{139}\).

Released oligosaccharides can be further derivatized before analysis. For example, oligosaccharides with a free reducing end can be labeled by reductive amination. As shown in Figure 1-15, a Schiff base in the glycosyl imine is formed after the aldehyde group reacts with the amine group which is then reduced to form a secondary amine. The advantage of labeling is that it facilitates stoichiometric measurement. Rudd’s group performed the analysis of serum N-glycans with 2-aminobenzamide labeling and the labeled glycans were then separated with a HILIC column\(^45\). The glycan structure is annotated with the aid of the digestion of glycans by exoglycosidases, which cleave the specific glycosidic bonds of monosaccharides. Another strategy for the analysis of labeled glycans is to use CE-LIF. Similar to the strategy mentioned above, multi exoglycosidases are also required for glycan structure annotation. Permethylation as shown in Figure 1-15 is another widely used method for the chemical derivatization of glycans, which can stabilize sialic acid, increase the sensitivity of mass spectrometry analysis and facilitate detailed structure analysis by multiple MS\(^n\) scans. Permethylated glycans are usually separated by RPLC as the methyl group can increase the hydrophobicity and thus increase the retention on RPLC. Tandem mass spectrometry is usually used for analysis of permethylated glycans. Multiple stages of mass spectra can provide B- and Y-type glycosidic and A-type cross-ring fragments for elucidation of glycan structures in CID analysis. Recently, ECD and ETD are also employed for analysis of permethylated glycans. Costello \textit{et al} analyzed permethylated milk oligosaccharides by ETD and N-glycans by hot ECD \(^{140,141}\). Various cross-ring
fragments of A- and X-types are dominant in ETD and hot ECD spectra, which provide the glycan branch and linkage information.

1.5 References


(3) Spiro, R. G. Glycobiology 2002, 12, 43R.


(90) Bynum, M. A.; Yin, H.; Felts, K.; Lee, Y. M.; Monell, C. R.; Killeen, K. 


(95) Wysocki, V. H.; Tsapralis, G.; Smith, L. L.; Breci, L. A. *J Mass Spectrom* 
**2000**, *35*, 1399.


*21*, 1012.

4230.

(106) Wu, S. L.; Jiang, H.; Lu, Q.; Dai, S.; Hancock, W. S.; Karger, B. L. *Anal 


Chapter 2: Analysis of Isoaspartic Acid by Selective Proteolysis with Asp-N and Electron Transfer Dissociation Mass Spectrometry


Co-authors’ work in this chapter: Wenqin Ni: experimental design and perform, data analysis, manuscript writing and revision; Shujia Dai: experimental design, manuscript writing and revision; Barry L. Karger: idea contribution, manuscript writing and revision and grant support; Zhaohui Sunny Zhou: idea contribution, experimental design, manuscript writing and revision and grant support.
2.1 Abstract

A ubiquitous yet underappreciated protein post-translational modification, isoaspartic acid (isoAsp, isoD or β-Asp), generated via the deamidation of asparagine or isomerization of aspartic acid in proteins, plays a diverse and crucial role in ageing, as well as autoimmune, cancer, neurodegeneration and other diseases. In addition, formation of isoAsp is a major concern in protein pharmaceuticals, as it may lead to aggregation or activity loss. The scope and significance of isoAsp have, up to now, not been fully explored, as an unbiased screening of isoAsp at low abundance remains challenging. This difficulty is due to the subtle difference in the physicochemical properties between isoAsp and Asp, e.g., identical mass. In contrast, endoprotease Asp-N (EC 3.4.24.33) selectively cleaves aspartyl peptides but not the isoaspartyl counterparts. As a consequence, isoaspartyl peptides can be differentiated from those containing Asp and also enriched by Asp-N digestion. Subsequently, the existence and site of isoaspartate can be confirmed by electron transfer dissociation (ETD) mass spectrometry. As little as 0.5 % of isoAsp was detected in synthetic beta amyloid and cytochrome c peptides, even though both were initially assumed to be free of isoAsp. Taken together, our approach should expedite the unbiased discovery of isoAsp.

2.2 Introduction

Ubiquitous in biological systems and protein pharmaceuticals, isoaspartic acid (isoAsp, isoD or β-Asp) in peptides and proteins arises from either deamidation of asparagine (Asn) or isomerization of aspartic acid (Asp) via a common succinimide intermediate, as depicted in Figure 2-1. Formation of isoAsp inserts an extra methylene
group into the protein backbone, resulting in a beta-peptide linkage \(^1\text{-}^4\). As a consequence, a “kink” in the protein conformation is generated, potentially leading to altered functions of the protein. For example, isoAsp has been implicated in β-amyloid aggregation and neurodegenerative disorders \(^4\text{-}^6\). In most cases, isoAsp formation is associated with loss in protein function. Typically accumulated over time, isoAsp is naturally associated with ageing, perhaps acting as a molecular clock \(^2\text{-}^7\). For example, the age-dependent accumulation of isoAsp in numerous proteins, e.g. α-crystallin in the eye lens, has been observed \(^8\). On the other hand, signaling and regulatory functions have also been reported for isoAsp formation, for example, in the case of fibronectin \(^9\). Another well studied system is the deamidation of proteins of the Bcl-2 family, critical apoptotic regulators following DNA damage and cancer therapy \(^10\text{-}^11\). Furthermore, even present at low abundance, isoAsp may also trigger autoimmune responses \(^12\text{-}^13\), likely due to altered antigen presentation \(^14\). Hence, it is not surprising that all organisms possess mechanisms to reduce the levels of isoAsp, either via the action of protein isoaspartic acid methyltransferase (PIMT or PCMT, EC 2.1.1.77) or putative proteolytic pathways \(^1,^15,^16\). For example, isoAsp formation and repair is central to the survival and germination of plant seeds \(^17,^18\). IsoAsp is certainly much more prevalent and plays a much broader role than currently realized owing to the high occurrence of Asn (~4%) and Asp (~5%) in proteins and the propensity for many of these residues to be converted into isoAsp. As discussed below, approaches that enable unbiased screening for low abundant isoAsp will greatly expedite the systematic analyses of isoAsp in proteomics and basic biology.

Another important area for isoAsp characterization is in the biopharmaceutical industry \(^19\text{-}^22\). Once isolated from host cells, protein pharmaceuticals and reagents alike are
devoid of repair mechanisms and are thus prone to the accumulation of isoAsp upon long-term storage. In fact, the presence of isoAsp in protein pharmaceuticals is often the norm rather than the exception\textsuperscript{19-22}. The effects of isoAsp formation have also been examined in numerous cases, including aggregation and loss of activity\textsuperscript{23,24}. In a few cases, isoAsp formation was eliminated by protein engineering\textsuperscript{25}. But most commonly, for long-term storage, isoAsp formation is minimized by optimizing formulation conditions, e.g., under mildly acidic conditions and at low temperature\textsuperscript{19,21}. However, once administered into patients and thus exposed to physiological conditions (pH ~ 7 and 37 °C), protein pharmaceuticals, particularly those with long circulation time, may generate significant amount of isoAsp; indeed, several recent investigations have shown considerable deamidation of monoclonal antibodies in serum; in one case, around 15 % over 20 days\textsuperscript{22,26}. Clearly, these results call for comprehensive \textit{in vivo} studies, as \textit{in vivo} modifications and activities of protein drugs ultimately determine their clinical outcome. Because samples recovered from biological systems are much more heterogeneous and only available in limited amounts, there is again a pressing need for sensitive and facile methods to characterize isoAsp.

Approaches for isoAsp analysis range from instrumental, chemical, enzymatic to immunological\textsuperscript{1-4,27}. Since the early days of isoAsp studies, protein isoaspartic acid O-methyltransferase, which methylates isoAsp but not Asp, has been used for the detection and quantitation of isoAsp; the process has been commercialized under the name IsoQuant\textsuperscript{28}. We have previously extended this method by trapping the resulting methyl isoaspartate esters, which are intrinsically labile, with hydrazines and various tags, allowing affinity enrichment and orthogonal detection\textsuperscript{29}. 
Currently, the most frequently employed method for isoAsp analysis in peptides is reversed phase liquid chromatography (RPLC) combined with mass spectrometry (MS). If separation can be achieved, peak assignment is typically inferred from the elution order. However, as others have noted and we also report in this paper, the elution order of these peptides significantly varies depending on the chromatographic conditions, and hence by itself is insufficient for unambiguous assignment. Equally challenging can be mass spectrometric analysis since Asp and isoAsp have identical mass and formal charge. Although different fragmentation patterns have occasionally been observed for the Asp and isoAsp peptides in selected cases via collision induced dissociation (CID), high energy CID (HCD) or negative electrospray ionization, the reporter ions are non-specific for isoAsp and highly dependent on sequence and conditions, rendering the assignment unreliable. As such, reference peptides are usually required in order to identify unambiguously each species.

A recent advance in distinguishing Asp and isoAsp is the application of electron capture dissociation (ECD) or electron transfer dissociation (ETD) mass spectrometry. As shown previously and illustrated in Figure 2-2, both ETD and ECD can generate a single pair of reporter ions (c+57 and z-57) that are unique to isoAsp. Using LC-ECD MS, hundreds of human isoAsp peptides have been identified from cell lines. However, many of the initial hits based on the reporter ions for isoAsp were inclusive, and consequently, additional constraints and manual inspection were required for definitive assignment. The difficulty is mainly due to the intrinsic limitations of ETD/ECD MS: the peak intensities for both signature ions of isoAsp are significantly lower than those for the common fragmentation ions, typically around 5 % of the nominal level in ETD/ECD MS/MS.
Indeed, the two peaks characteristic of isoAsp are often indistinguishable from spurious noise peaks. The problem is even more exacerbated when Asp and isoAsp peptides co-elute. It is again to be emphasized that the precursor ions for both isoAsp and Asp have identical mass, so co-eluted Asp peptides essentially dilute the isoAsp species, further reducing the intensities of the characteristic peaks of isoAsp in the MS/MS spectra. As such, isoAsp peptides, particularly of low abundance, are likely to be not observed.

**Figure 2-1.** Formation of isoaspartic acid via deamidation of asparagine or isomerization of aspartic acid. The peptide backbones are highlighted in bold.
Figure 2-2. Mechanism of fragmentation of the Asp and isoAsp peptides in ETD MS. (a) formation of c and z fragment ions of the Asp peptides, which is the same for isoAsp peptides (not shown); (b) formation of the c+57 and z-57 diagnostic ions of the isoAsp peptides.

Given the current situation, a method to differentiate isoAsp peptides from the Asp counterparts prior to the LC-MS analysis would be highly desirable. Owing to the different chemical structure of isoAsp vs Asp, as illustrated in Figure 2-3, isoAsp linkage has been shown to be refractory to proteolytic cleavage by endoprotease Asp-N (EC 3.4.24.33), an enzyme widely used in protein analysis; in contrast and importantly, Asp-N efficiently cuts at the N-terminal side of Asp residues. Up to now, the resistance of isoAsp to cleavage by Asp-N protease has been explored, albeit only in a handful cases, to identify isoAsp candidates, but not with ECD or ETD analysis. The underutilization
of this straightforward and general sample preparation step may be attributed to several factors, such as whether all isoAsp residues are refractory to Asp-N digestion. However, the major concern up to now has been that the assignment is only tentative due to the uncertainty of incomplete digestion of the Asp species and the lack of direct identification of the isoAsp species.

![Figure 2-3. Structures of Asp and isoAsp peptides. Asp-N cleaves Asp at its N-terminal side but isoAsp is resistant to proteolysis by Asp-N.](image)

Altogether, we envisaged that differentiation and enrichment of isoAsp peptides via Asp-N digestion (sample preparation) and detection of isoAsp by ETD MS (sample analysis) should lead to definitive identification and quantitative determination of isoAsp peptides, even at trace levels in the presence of the corresponding Asp peptides. As illustrated in Figure 2-4, first, Asp-N is utilized to cleave Asp peptides into smaller peptides while the
isoAsp peptides remain intact, flagging and enriching the isoAsp species. Next, with little interference from the Asp species, the isoAsp peptides are detected by ETD MS with high confidence. Indeed, as we report herein, low abundant (as low as 0.5 %) isoAsp in beta-amyloid (Aβ) and cytochrome c peptides are detected and quantified, even when isoAsp and Asp peptides are co-eluted. Moreover, for samples that were initially thought to be free of isoAsp, trace amounts of isoAsp was observed, highlighting the utility of the approach for unbiased screening and discovery.

**Figure 2-4.** Overview of the approach combining the differentiation and enrichment of isoAsp peptides by Asp-N digestion and the identification of isoAsp species by ETD-MS.

2.3 Experimental Section

2.3.1 Chemicals

Sequences and abbreviations of the peptides are listed in Table 2-1. Aβ1-16 and Aβ1-42 were purchased from Bachem (Torrance, CA). Cyt-c, isoCyt-c and isoAβ1-42 were gifts from Drs. Mark Mamula and Hester Doyle of Yale University. Sequencing grade
endoprotease Asp-N (EC 3.4.24.33) from *Pseudomonas fragi* (mutant strain) was purchased from Sigma-Aldrich (Catalog No. P3303, St. Louis, MO), and mass spectrometry grade lysyl endopeptidase (Lys-C) was from Wako (Catalog No. 121-05063, Richmond, VA).

**Table 2-1.** Sequences and abbreviations of the synthetic peptides.

<table>
<thead>
<tr>
<th>Peptides</th>
<th>Abbreviations</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-amyloid (1-42) Asp7</td>
<td>Aβ1-42</td>
<td>DAEFRHDSGYEVHHQKLVFFAEDVSNKGAIIGLMVGVVIA</td>
</tr>
<tr>
<td>Beta-amyloid (1-42) isoAsp7</td>
<td>isoAβ1-42</td>
<td>DAEFRHisoDSGYEVHHQKLVFFAEDVSNKGAIIGLMVGVVIA</td>
</tr>
<tr>
<td>Beta-amyloid (1-16)</td>
<td>Aβ1-16</td>
<td>DAEFRHDSGYEVHHQK</td>
</tr>
<tr>
<td>Cytochrome c (46-60) Asp51</td>
<td>Cyt-c</td>
<td>GFSYTDANKNKGITW</td>
</tr>
<tr>
<td>Cytochrome c (46-60) isoAsp51</td>
<td>isoCyt-c</td>
<td>GFSYTisoDANKNKGITW</td>
</tr>
</tbody>
</table>

### 2.3.2 Sample Preparation

Peptide concentrations were determined from the absorbance at 280 nm measured on a NanoDrop UV-vis spectrometer (ND-1000, Wilmington, DE), using extinction coefficients of 1490 and 5500 M⁻¹·cm⁻¹ for tyrosine and tryptophan, respectively. Separately, Aβ1-42 and isoAβ1-42 were freshly dissolved in 0.16 % ammonium hydroxide at pH 9.5 to a final concentration of 111 μM. Then, each solution was diluted 4-fold into 50
mM ammonium bicarbonate in water and mixed to afford different percentages of isoAsp peptide. Separately, Cyt-c and isoCyt-c were dissolved in 50 mM ammonium bicarbonate in water to a final concentration of 70 μM. Then, the above solutions of Cyt-c and isoCyt-c were mixed to generate different percentages of isoAsp peptide.

2.3.3 Aging of Amyloid Peptide

Aβ1-16 was dissolved in 25 mM ammonium bicarbonate in water at pH 8.3 to a final concentration of 750 μM. The peptide solution was diluted 37.5-fold into 100 mM sodium acetate at pH 4.0 to a final concentration of 20 μM. The solution was then incubated at 37 °C for 1 month and stored at -80 °C.

2.3.4 Protease Digestion

Lys-C and Asp-N were dissolved in 100 mM ammonium bicarbonate in water (pH 8.3) to final concentrations of 0.5 and 0.05 μg/μL, respectively. Each solution of Aβ1-42 and isoAβ1-42 was digested with Lys-C at an enzyme:peptide ratio (w/w) of 1:10 for 6 hrs at 20 °C. All other solutions containing Asp and isoAsp mixtures were treated with Asp-N at an enzyme:peptide ratio (w/w) of 1:40 in 50 mM ammonium bicarbonate in water (pH 8.0) overnight at 37 °C. The samples were then stored at -80°C.

2.3.5 HPLC Separation of Asp and IsoAsp Containing Peptides

HPLC was performed on a Varian HPLC/UV-Vis system using a Vydac C18 column (4.6 mm i.d. × 250 mm, Grace, Deerfield, IL). Trifluoroacetic acid (TFA, 0.1 %) or 0.1 % formic acid in water were used as mobile phase A, and in acetonitrile as mobile phase B. The flow rate was set at 1.0 mL/min, and the wavelength for UV detection was 220 nm. A linear gradient from 2 % to 40 % mobile phase B over 30 minutes was used to separate Lys-
C digest of Aβ1-42 and isoAβ1-42. A gradient from 2% to 30% mobile phase B over 56 minutes was used to separate the peptide mixture of Cyt-c and isoCyt-c.

2.3.6 LC-MS Analysis

LC-MS experiments were performed on an LTQ-XL with ETD mass spectrometer (Thermo Fisher, San Jose, CA), consisting of a linear ion trap with an additional chemical ionization source to generate fluoranthene anions. An Ultimate 3000 nano-LC pump (Dionex, Mountain View, CA) was used to generate the gradient, and a self-packed reversed phase (RP) column (Vydac C18, 300 Å pore and 5 μm particle size, 75 μm i.d. × 10 cm) was coupled online to the mass spectrometer through a nanospray ion source (New Objective, Woburn, MA). Mobile phase A consisted of 0.1 % formic acid in water and mobile phase B of 0.1 % formic acid in acetonitrile. Initially, a linear gradient from 2 % to 10 % mobile phase B over 2 minutes, and then from 10 % to 40 % mobile phase B over 15 minutes was used to separate the mixtures of Cyt-c peptides and Lys-C digest of Aβ1-42 and isoAβ1-42. Due to the co-elution of Asp and isoAsp isoforms using the above condition, a shallower gradient from 2 % to 40 % mobile phase B over 30 minutes, at a flow rate of 200 nL/min, was applied to achieve a near baseline separation of these isoforms. At least three blank runs were used to minimize any carryover effects between sample injections. The temperature of ion transfer tube of the linear ion trap was held at 245 °C, and the electrospray voltage at 2.2 kV. The mass spectrometer was operated in the data-dependent mode to switch automatically between MS (1st scan event), CID-MS\(^2\) (2nd scan event), ETD-MS\(^2\) (3rd scan event). Briefly, a full MS scan at the mass range of 400-2000 m/z was followed by 2 sequential data-dependent MS\(^2\) scans (CID and ETD) of the most abundant precursor ion, with a 3 mass unit isolation width, from the full MS scan. Dynamic exclusion
was implemented with 2 repeat counts (repeat duration of 30 seconds, exclusion list 200, and exclusion duration of 30 seconds). The normalized collision energy in CID was set at 28%. The chemical ionization (CI) source parameters for fluoranthene, such as ion optics, filament emission current, anion injection time (anion target value set at $3 \times 10^5$ ions), fluoranthene gas flow, and CI gas flow, were optimized automatically following the standard procedure for tuning the instrument. The reagent transfer multipole RF amplitude was set at 300 V. The duration time of the ion/ion reaction was maintained constant throughout the experiment at 150 ms. The supplemental activation function was integrated into the ETD data acquisition method. To detect and quantify Asp/isoAsp containing peptides, selective reaction monitoring (SRM) was applied to achieve lower detection limits. The multiply charged precursor ions of Asp/isoAsp containing peptides were selected for SRM detection based on the quality of ETD spectra, such as the relatively high response of $c_n$ and $z_{l-n}$ ion series and reporter ions ($c_n+57$ and $z_{l-n}-57$, where $n$ is the position of Asp/isoAsp from the N terminus and $l$ is the total number of amino acids in the peptide).

2.3.7 Data Processing

LC-MS/MS data were analyzed using the Xcalibur 2.0 software (Thermo Fisher, San Jose, CA). For all isoAsp and Asp containing peptides, $c_n$, $z_{l-n}$, $c_n+57$, and $z_{l-n}-57$ ions were extracted using Qual Browser software (Thermo Fisher, San Jose, CA). The peak areas of selected ions were extracted manually to quantify the relative amount of isoAsp peptide and its Asp counterpart.

2.4 Results and Discussion

2.4.1 Distinguishing IsoAsp and Asp Peptides by ETD
We first characterized the synthetic peptides containing either isoAsp or Asp using LC-ETD-MS. For the Cyt-c/isoCyt-c peptides, the ETD spectra from the triply charged precursor ions \([m/z = 568.3]\) of both the isoAsp/Asp forms showed the complete sequence coverage of the peptides by the \(c\) and \(z\) series ions. For the isoaspartyl peptide, the diagnostic ions, \(c_5+57 (m/z = 630.3, 1+)\) and \(z_{10}-57 (m/z = 1073.6, 1+)\), were observed, as shown in Figure 2-5. Conversely, no such ions were detected from the corresponding aspartyl peptide, as shown in Figure 2-5. Comparable results were observed for the peptide fragment \([^{1}DAEFRH(D/isoD)SGYEVHHQK]^{16}\) from \(A\beta1-42/isoA\beta1-42\) digested by Lys-C, as shown in Figure 2-6. These observations are in agreement with those of others\(^{27,36-45}\). For all peptides studied, however, the intensity of the pair of diagnostic ions was found to be considerably lower than other fragmentation ions; for example, the intensity of the \(z_{10}-57\) peak was only about 7 % of the \(z_{10}\) peak in the case of isoCyt-c peptide (see Figure 2-7).
Figure 2-5. ETD spectra of Cyt-C (a) and isoCyt-C (b).
Figure 2-6. ETD spectra of Aβ1-16 fragments from Lys-C digests of Aβ1-42 (a) and isoAβ1-42 (b).
Figure 2-7. Enrichment of isoAsp-containing peptide in cytochrome c via Asp-N digestion. The dashed line indicates the average value of peak area ratios of $z_{10}$ - 57 to $z_{10}$ observed in the synthetic isoCyt-c peptide.

2.4.2 Limit of Detection for IsoAsp Peptide when Co-eluted with Asp Peptide

To simulate complex systems, mixtures of synthetic Asp and isoAsp peptides with varied amounts of isoAsp ranging from 0 to 100 % were prepared and analyzed by LC-ETD-MS/MS. The initial LC separation conditions led to, at best, only partial separation of all isoAsp/Asp isoforms. As discussed earlier, due to the subtle difference between the physiochemical properties of isoAsp and Asp peptide, these peptides often co-elute; one example is shown in Figure 4 in a recent paper. The precursor ions for the matching isoAsp and Asp peptides have identical mass and generate identical c and z ions; hence, the
co-eluted Asp peptide essentially dilutes the isoAsp species, further reducing the already low intensities of the \( z_{10} - 57 \) and \( c_{t+n} + 57 \) diagnostic ions of the isoAsp peptide. Indeed, as shown in Figure 2-8, the ratio of peak area for the \( z_{10} - 57 \) to \( z_{10} \) ions from Aβ1-16 fragment in Lys-C digest of Aβ1-42 peptides exhibited an overall linear correlation with the percentage of the isoAsp peptide in the mixture. However, as expected, the ratio significantly varied from one run to another when isoAsp was low (see the inset in Figure 2-8). Similar results were obtained for the \( c_6 \) and \( c_6 + 57 \) ions, as well for the Cyt-c peptide, see Figure 2-9. Therefore, the large variation of the peak area ratio makes it difficult to assign isoAsp when the abundance of this form is lower than 5 %; however, this level is commonly found in biological samples or protein pharmaceuticals.

![Figure 2-8](image)

**Figure 2-8.** Correlation between the percentage of isoAsp in Lys-C digested Aβ1-42 and isoAβ1-42 and the peak area ratio of \( (z_{10} - 57)/z_{10} \) (a) and \( (c_6 + 57)/c_6 \) (b). The insets show the lower percentage region.
Figure 2-9. Correlation between the percentage of isoAsp in Cyt-c and isoCyt-c and the peak area ratio of \((z_{10-57}/z_{10})\) (a) and \((c_{5+57}/c_{5})\) (b). The insets show the lower percentage region.

2.4.3 Differentiation and Enrichment of IsoAsp Peptides via Asp-N Digestion

As depicted in Figure 2-4, we devised a strategy to differentiate and enrich isoAsp-containing peptides, in which endoprotease Asp-N specifically cleaves Asp but not isoAsp peptides. Figure 2-7 illustrates the results for the cytochrome c peptides that contained low levels of isoAsp: the ratios of peak areas for the \(z_{10-57}\) to \(z_{10}\) ions were quite low before Asp-N digestion, but markedly increased after digestion, and in fact, approached that for the synthetic isoAsp peptides (marked by the dashed line), indicating near complete digestion of the Asp species. Similar results were observed for beta-amyloid peptides (see Figure 2-10). Significant lowering of the limit of detection for isoAsp makes feasible a broad array of
applications, e.g., the discovery of early disease biomarkers present at low abundance and shorten the stability tests of protein pharmaceuticals.

Figure 2-10. Enrichment of isoAsp containing peptide in Lys-C digested Aβ1-42 via Asp-N digestion. The dashed line indicates the average value of peak area ratios of ($z_{10}$-57)/$z_{10}$ observed in Aβ1-16 fragment from Lys-C digested isoAβ1-42.

2.4.4 Unbiased Identification of IsoAsp

Of importance in Figures 2-7 and 2-10, after Asp-N digestion, isoAsp was clearly detectable for the synthetic standard Asp peptide (0 % spike) of both beta-amyloid and cytochrome c. This result initially appeared puzzling, but turned out to be an excellent example of the power of our approach. As mentioned in the previous section, the synthetic standard Asp peptides were originally assumed to be free of isoAsp, as the initial LC-ETD-MS/MS analyses did not reveal signature ions above the noise level. In order to determine
whether the low level isoAsp was generated during sample preparation, we repeated the experiments in $^{18}$O water $^{30,40,53,54}$. No peaks corresponding to $^{18}$O isoAsp were detected (data not shown), suggesting isoAsp was present in the original samples and not due to sample preparation factors. The results were consistent with the observation that, under our conditions (pH above 7 and a few hours in solution), isomerization of Asp is negligible $^{1,3,7,19}$. To ascertain the level of isoAsp in the untreated standard peptides, we used ETD MS detection as a guide to optimize LC conditions (the shallower gradient in the Experimental Section) to fully resolve the isoAsp and Asp peptides. The chromatogram for the Aβ1-16 peptide is shown in Figure 2-11; the small peak (B) following the main peak (A) was identified as the isoAsp form by both ETD-MS/MS and comparison to the relative retention time of the isoAsp standard peptide. Assuming the response factor of isoAsp and Asp isomers are equal in the ETD-MS spectrum, about 0.5 %, 0.7 % and 1.0 % of isoAsp was present in the synthetic peptides, Aβ1-16, Cyt-c and Aβ1-42 (see Figures 2-11, 2-12 and 2-13), respectively. The presence of isoAsp species in synthetic Asp peptides is not surprising, since isoAsp formation has been frequently observed in solid-phase peptide synthesis $^{55,56}$. 

Figure 2-11. Extracted ion chromatogram and ETD spectra of the Aβ1-16 peptide. The reporter ions of isoAsp ($c_6 + 57$, $z_{10} - 57$) are observed only in peak B, not in peak A.
Figure 2-12. Extracted ion chromatogram and ETD spectra of synthetic Cyt-c peptide. The reporter ions of isoAsp ($c_5 + 57$, $z_{10} - 57$) are observed only in Peak B, not in Peak A.
Figure 2-13. Extracted ion chromatogram and ETD spectra of synthetic Aβ1-42 peptide.

The reporter ions of isoAsp (c₅⁺57, z₁₀⁻57) are observed only in Peak B, not in Peak A.

2.4.5 Specificity of Asp-N

Considering there had been limited comparison by others of the activities of Asp-N on matching isoAsp and Asp peptides, we tested and found that all synthetic isoAsp peptides were resistant to Asp-N digestion while their Asp counterparts were completely cleaved under the same conditions (see Figure 2-14). As illustrated in Figure 2-3, isoAsp introduces an extra methylene group into the peptide backbone and results in a beta-peptide linkage, compared to a typical alpha-peptide linkage rendered by Asp. Moreover, relative to peptide backbones, the stereochemistry of the chiral center next to the amide is inverted when Asp is converted into isoAsp, essentially generating a D-amino acid from the L-Asp as depicted in Figure 2-3. As such, isoAsp linkages have been shown to be resistant to other
proteases as well $^{57,58}$. Hence, isoAsp in other peptides are in all likelihood resistant to Asp-N cleavage as well. It should be noted that cleavage at the N-terminal side of glutamic acid (Glu11) in Aβ1-16 was also observed, albeit less than 5%. Slow cleavage at some glutamyl sites has been reported for Asp-N $^{51}$, but this digestion can be minimized by shorter incubation time, and importantly, the cleavage at glutamate does not affect the analysis of isoAsp.

![Figure 2-14](image)

**Figure 2-14.** Extracted ion chromatograms of Asp/isoAsp containing Aβ1-16 fragments generated by Lys-C digested synthetic Aβ1-42 or isoAβ1-42. (a) and (b) demonstrate the XICs of fragments of the Aβ1-16 Asp form (a) and isoAsp form (b), respectively, before Asp-N digestion. (c) and (d) show the XICs of Asp form (c) and isoAsp form (d) of the Aβ1-16 fragments and its Asp-N digested fragments.

### 2.4.6 Elution Order of IsoAsp and Asp Peptides
As discussed in the introduction, assignment of isoAsp and Asp peptides often has been based largely on elution order. The assumption is that isoAsp peptides elute earlier than their corresponding Asp forms on reversed phase liquid chromatography (RPLC)\textsuperscript{30,31}. However, both beta-amyloid and cytochrome c isoAsp peptides eluted after their Asp counterparts under the LC-MS conditions, see Figure 2-11 for beta-amyloid peptides and Figure 2-12 for cytochrome c peptides. To the best of our knowledge, this is the first report of such a scenario. In addition, similar to reported by others, the elution order was affected by many factors, including the ion pairing reagent in the mobile phase and the packing material\textsuperscript{30,31}. As shown in Figures 2-15 and 2-16, for both pairs of peptides, the isoAsp peptides eluted earlier than the Asp forms in RPLC (monitored by UV) with 0.1% TFA in the mobile phase, while the two forms co-eluted by using 0.1% formic acid in the mobile phase using the same gradient. The above results again emphasize that it is unreliable to assign the isoAsp and Asp forms solely based on the elution order in liquid chromatography.
Figure 2-15. HPLC/UV chromatograms of synthetic Cyt-c and isoCyt-c using (a) 0.1 % TFA and (b) 0.1 % formic acid as ion pairing reagents in the mobile phase. (c) Total ion chromatogram of synthetic Cyt-c and isoCyt-c using 0.1 % formic acid as ion pairing reagent in nano-LC-MS.
Figure 2-16. HPLC/UV chromatograms of Lys-C digested synthetic Aβ1-42 and isoAβ1-42 using (a) 0.1 % TFA and (b) 0.1 % formic acid as ion pairing reagents in the mobile phase. (c) Total ion chromatogram of Lys-C digested synthetic Aβ1-42 and isoAβ1-42 using 0.1 % formic acid as ion pairing reagent in nano-LC-MS.

2.4.7 Detection of IsoAsp in Aged Beta-amyloid Peptide

Finally, we applied our approach to an unknown sample that closely resembled a biological system. Beta-amyloid peptides are generally considered as a major causative factor in Alzheimer’s disease (AD) \(^{4-6,57}\). Noticeable levels of isoAsp have been detected at several positions, e.g., Asp-1 and Asp-7, in beta-amyloid from the senile plaques in the brains of Alzheimer’s patients \(^{4-6}\). To simulate aging, the Aβ1-16 peptide was incubated at
pH 4, 37 °C for a total of one month. As shown in Figure 2-17, LC-MS analysis of this aged peptide showed three peaks. First, to establish which peptide(s) contained isoAsp, the sample was digested with Asp-N. As shown in Figure 2-17, peak A disappeared after digestion, indicating this peptide contained an Asp residue at position 7. Interestingly, both peaks B and C remained unchanged, suggesting isoAsp was formed at the 7th position in both peptides. Subsequently, peaks B and C, before and after Asp-N digestion, were analyzed by ETD MS and found to produce the pair of characteristic ions for isoAsp (z_{10}-57 and c_{6}+57; XIC for the former is shown in Figure 2-17), confirming both contained isoAsp at the 7th position. In addition, in one of the peptides, the first Asp had the potential to be converted to isoAsp due to the DA sequence; however, the expected signature ions (M-57 and c_{1}+57) were observed at the near noise level, probably due to the lower fragmentation efficiency in N-terminus of the peptide. It has been reported that isomerization of Asp7 should be faster than that of Asp1^{4-6}, so peak B is likely to be D^{1}AEFRHi soD^{7}SGYEVHHQK^{16} and peak C as isoD^{1}AEFRHi soD^{7}SGYEVHHQK^{16}. In summary, even for this relative simple system, the partially-resolved peaks could not be assigned based on their elution order. On the other hand, the changes in peaks before and after Asp-N treatment were easily observed.
2.5 Conclusions

This paper presents a general workflow for isoAsp analysis that comprises the following steps. First, a sample, e.g., tryptic digest, is treated with endoprotease Asp-N that cleaves Asp. Second, the resulting mixture and the original sample are analyzed by LC-MS to determine all peptides which sequences contain Asp but are resistant to Asp-N cleavage. For this screening step, CID MS rather than ETD MS suffices and multiply charged peptides

---

**Figure 2-17.** Extracted ion chromatograms of aged A1-16 peptide before and after Asp-N digestion. (a) and (b) show the XICs of the triply charged precursor ion from aged Aβ1-16 peptide before and after Asp-N digestion, respectively. (c) and (d) demonstrate the XICs of product ions $z_{10}$ and $z_{10} - 57$ of aged A1-16 peptide before Asp-N digestion, respectively.
are not required. Third, candidate isoAsp peptides identified in the second step can be interrogated further by a combination of optimized separation and ETD MS/MS. Such analysis can be performed on either the original sample or Asp-N digest, or both, to further increase the confidence of assignment.

2.6 Appendix

2.6.1 Monitor Sample Preparation Induced IsoAsp Using $^{18}$O Labeling

It is reported that isoAsp can be artificially generated during sample preparation $^{40}$. To examine possible sample preparation induced isoAsp, we mimicked the enzymatic Lys-C digestion condition in $^{18}$O water. $^{18}$O water (97%) was obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA). Aβ 1-16 was incubated in 50 mM ammonium bicarbonate in $^{18}$O water at 37 °C for 40 hours (a longer incubation time than the digestion condition). After incubation, Aβ 1-16 was analyzed by LC-ETD-MS/MS.

Only one peak of Aβ 1-16 was extracted without a mass shift, as shown in Figure 2-18 (a). The observed mass (489.8, 4+, 652.7, 3+ and 978.1, 2+) was matched with the theoretical mass (1953.9 Da) without 2 Da mass shift in Figure 2-18 (b). The peptide sequence was further confirmed by ETD-MS/MS spectrum as shown in Figure 2-18 (c). The diagnostic ions for isoAsp (c+57, z-57) were not identified. As the +2 Da mass shift peak possibly overlapped with the isotopic peak, we also checked the diagnostic ions for artificial isoAsp (c+57+2, z-57+2). None of these ions were identified, as shown in the inset of Figure 2-18 (c). These data demonstrated that no isomerization from Asp to isoAsp was induced in our digestion conditions.
Figure 2-18. (a) Extracted ion chromatograms of Aβ1-16 peptide after incubation in $^{18}$O water; (b) MS of Aβ1-16 peptide after incubation in $^{18}$O water; (c) ETD-MS/MS spectrum of Aβ1-16 peptide after incubation in $^{18}$O water and the expanded parts in the inset.
2.7 References


Chapter 3: An Integrated Proteomic Analysis of Major Isoaspartyl-Containing Proteins in the Urine of Wild Type and Protein L-Isoaspartate O-Methyltransferase-Deficient Mice


Co-authors’ work in this chapter: The biochemical analysis was done by Prof. Clarks’ group in Department of Chemistry and Biochemistry and the Molecular Biology Institute, UCLA; the proteomics and target isoAsp analysis by LC-MS/MS was done by Prof. Karger’s group.

Wenqin Ni: experimental design and perform for proteomics and target isoAsp analysis, data analysis, manuscript writing and revision; Shujia Dai: experimental design, manuscript writing and revision; Alexander N. Patananan: sample collection, experimental perform for biochemical analysis and manuscript writing and revision; Steven G. Clarke: sample providing, manuscript revision and grant support; Barry L. Karger: idea contribution, manuscript writing and revision and grant support; Zhaohui Sunny Zhou, idea contribution, experimental design, manuscript revision and grant support.
3.1 Abstract

The formation of isoaspartyl residues (isoAsp or isoD) via either aspartyl isomerization or asparaginyl deamidation alters protein structure and potentially biological function. This is a spontaneous and non-enzymatic process, ubiquitous both in vivo and in non-biological systems, such as in protein pharmaceuticals. In almost all organisms, protein L-isoaspartate O-methyltransferase (PIMT, EC2.1.1.77) recognizes and initiates the conversion of isoAsp back to aspartic acid. Additionally, alternative proteolytic and excretion pathways to metabolize isoaspartyl-containing proteins have been proposed but not fully explored, largely due to the analytical challenges for detecting isoAsp. We report here the relative quantitation and site profiling of isoAsp in urinary proteins from wild type and PIMT-deficient mice, representing products from excretion pathways. First, using a biochemical approach, we found that the total isoaspartyl level of proteins in urine of PIMT-deficient male mice was elevated. Subsequently, the major isoaspartyl protein species in urine from these mice were identified as major urinary proteins (MUPs) by shotgun proteomics. To enhance the sensitivity of isoAsp detection, a targeted proteomic approach using electron transfer dissociation-selected reaction monitoring (ETD-SRM) was developed to investigate isoAsp sites in MUPs. Thirty-eight putative isoAsp modification sites in MUPs were investigated, with five derived from the deamidation of asparagine that were confirmed to contribute to the elevated isoAsp levels. Our findings lend experimental evidence for the hypothesized excretion pathway for isoAsp proteins. Additionally, the developed method opens up the possibility to explore processing mechanisms of isoaspartyl proteins at the molecular level, such as the fate of protein pharmaceuticals in circulation.
3.2 Introduction

The generation of isoaspartyl residues in proteins is one of the most common spontaneous, non-enzymatic post-translational modifications. Asparaginyl deamidation and aspartyl isomerization both lead to the generation of the isoaspartyl residue (isoAsp or isoD) as shown in Figure 3-1. Isoaspartyl residues contain an additional methylene group in the polypeptide backbone, resulting in a beta-peptide linkage. As such, the beta linkage can impart protein structural changes (kinks) that typically lead to alteration in protein function. Proteins containing isoAsp have been associated with the loss of function during aging, e.g., calmodulin and tubulin in aged brain, and β-amyloid in Alzheimer’s disease. Additionally, differences in the proteolytic degradation of isoaspartyl versus aspartyl peptides in antigen-presenting cells may render such peptides immunogenic. IsoAsp formation can also confer specific biological functions, such as in the regulation of the Bcl-2 protein family following DNA damage and cancer therapy. Additionally, isoAsp has been observed in protein therapeutics and vaccines during production, storage and even after administration, including tissue plasminogen activator and monoclonal antibodies. This modification can alter the immunogenicity and efficacy of the biopharmaceutical.
Figure 3-1. Formation of isoAsp via deamidation of asparagine or isomerization of aspartic acid and the repair of isoAsp by PIMT. The peptide backbone is highlighted in bold.

One well-established pathway for isoAsp processing involves the protein L-isoaspartate O-methyltransferase (PIMT or PCMT, EC 2.1.1.77). Using S-adenosylmethionine (AdoMet or SAM), the PIMT enzyme specifically methylates isoAsp, but neither Asp nor Asn, to form methyl isoaspartate ester that ultimately converts back to aspartic acid (see Figure 3-1). The over-expression of PIMT extends the lifespan of Drosophila and Escherichia coli under stress conditions. Conversely, levels of isoAsp in different cells and tissues in PIMT knockout (KO) mice are significantly elevated over that of the wild-type mice, but eventually plateau with age. However, isoaspartyl levels in the urine continue to increase in knockout animals with age, suggesting the existence of
alternative proteolytic and excretive isoAsp processing pathways \(^{21,22}\). The elucidation of these pathways may not only lead to a better understanding of isoaspartyl regulation, but also to potential new approaches for treatment of diseases. However, technical challenges associated with the identification and quantitation of isoAsp have hindered progress in these areas. To overcome these challenges, we report a novel workflow for the proteomic analysis of isoAsp containing proteins in mouse urine, followed by their targeted site-specific quantitation.

The analysis of isoAsp remains difficult due to the subtle differences among isoAsp, Asp, and Asn residues. For example, isoAsp and Asp have identical mass. Enzymatic, \(^{8,23,24}\) chemical, \(^{25,26,27,28}\) and chromatographic methods \(^{29}\) have been developed for the detection of isoAsp residues in proteins. In recent years, liquid chromatography tandem mass spectrometry (LC-MS/MS) has become the method of choice \(^{30,31}\). Importantly, the more recent development of electron capture dissociation (ECD) \(^{32}\) and electron transfer dissociation (ETD) \(^{33,34,35}\) make it possible to directly identify isoAsp peptides and their sites. In these methods, isoaspartyl, but not aspartyl residues, can generate site-specific reporter ions, \(c_{l,n}+57\) \((\text{C}_2\text{H}_2\text{O}_2)\) and \(z_n-57\) \((\text{C}_2\text{HO}_2)\) fragments, where \(n\) is the position of the isoAsp residue from the C-terminus of the peptide and \(l\) is the peptide length \(^{32,33}\). As a result, LC-MS/MS, using ECD or ETD fragmentation, can differentiate and quantitate Asp and isoAsp residues in peptides \(^{32}\). However, for large scale isoAsp analysis using a shotgun proteomics strategy, the specificity and sensitivity of detection by ECD or ETD MS remain inadequate \(^{36}\). For example, the MS/MS spectra are commonly acquired in a data-dependent mode in shotgun proteomics, and the randomly selected MS/MS scans do not ensure that fragmentation of precursor ions occurs at or close to the chromatographic peak.
apex. Therefore, poor sensitivity of isoAsp detection, low quality of MS/MS spectra, and missed identification of peptides can result \(^{37,38}\). On the other hand, once candidate isoAsp peptides are identified, ETD/ECD can be fine tuned to give higher quality data for confirmation.

In this paper, we describe an integrated proteomics workflow to investigate specific isoAsp containing proteins in mice urine. First, using an enzymatic assay based on PIMT, the total isoaspartyl level of proteins in the urine of PIMT-deficient mice was found to be elevated. Subsequently, to identify the major content of excreted isoaspartyl-containing protein species in the urine of wild type (WT) and PIMT knockout mice, proteomic analysis was first performed. Based on the results, the digested peptides of a family of major urinary proteins (MUPs) that could potentially form isoAsp were selected for subsequent targeted analysis. Selected reaction monitoring (SRM) with full-scan ETD MS/MS (ETD-SRM) was implemented to enhance the sensitivity of detection of low abundance isoAsp peptides \(^{35}\). Five isoaspartyl sites in MUPs were identified and their relative abundance determined. The targeted results indicated higher levels of isoAsp in the urine of PIMT-knockout mice than their wild type littermates, in agreement with the overall isoAsp content in urine, determined separately by an enzymatic method. As a complement to shotgun proteomics, the ETD-SRM targeted approach is demonstrated to generate reliable identification and significantly enhanced sensitivity for quantification for low abundance isoAsp/Asp peptides. The targeted method for isoAsp peptides can be applied as a high throughput assay for the determination of isoAsp modifications in low abundance on proteins in biological fluids.

### 3.3 Experimental
3.3.1 Materials

Mass spectrometric grade lysyl endopeptidase (Lys-C) was purchased from Wako (Richmond, VA) and $^{18}$O water (97%) from Cambridge Isotope Laboratories (Andover, MA). LC-MS grade water was obtained from J. T. Baker (Phillipsburg, NJ), and HPLC grade acetonitrile from Thermo Fisher Scientific (Fairlawn, NJ). Amicon ultrafiltration devices, Ultrafree-0.5, with 5 kDa or 10 kDa molecular weight cutoff were obtained from Millipore (Billerica, MA). All other reagents were from Sigma-Aldrich (St. Louis, MO).

3.3.2 Wild Type and PIMT Knockout Mice and Urine Collection

PIMT$^{-/-}$ mice of a background of 50% C57BL/6J and 50% 129/SvJae were generated from breeding heterozygous PIMT$^{+/-}$ mice $^{39}$. Mice were fed an NIH-31 Modified Diet #7013 (Teklad Diets, Madison, WI) and maintained on a 12-h light/dark cycle in a barrier facility. Genotypes were determined by PCR of genomic DNA from tail biopsies of 18-day old pups. The WT primers (5’-ACCCTCTTCCCATCCACATCGCCGAG and 5’-AGTGGCAGCGACGGCAGTAACAGCGG), upstream and downstream of exon 1, yield a product of 409 nucleotides in the absence of the neomycin cassette. The KO primers (5’-CGCATCGAGCGACGACGTACTCGG and 5’-GCACGAGGAAGCGGTCAGCCATTCA) are both specific for the neomycin resistance gene that is inserted within exon 1, yielding a product of 310 nucleotides. Mouse urine was collected on Parafilm and stored at -20 °C.

3.3.3 Urine Fractionation Procedure

Approximately 100 µL of mouse urine from WT or KO mice (male or female, three of each group) was applied to a Millipore Amicon Ultra-0.5 centrifugal filter device with a 10 kDa molecular weight cut off, and centrifuged at 14,000 × g for 15 min. The filtrate,
representing the peptide fraction, was collected and designated "<10". To collect the retentate (~40 µL), the filter was inverted, placed in a centrifuge tube and spun for 2 min at 1,000 × g. The recovered sample (~35 µL) contained proteins above approximately 10 kDa, (designated ">10"). Residual small molecules were depleted by adding 5 µL of this material to 245 µL of 0.2 M Bis-Tris (2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)-1,3-propanediol) buffer (minimum 98% titration), adjusted to pH 6.4 with HCl. After thorough mixing, 225 µL of the sample was applied to a new Amicon filter and centrifuged at 14,000 ×g for 15 min, leaving about 40 µL of retentate. The filter was transferred to a new centrifuge tube, inverted, and centrifuged for 2 min at 1,000 ×g to recover the protein fraction in a volume of about 35 µL.

3.3.4 Quantitation of Isoaspartyl Content in Mouse Urine

A plasmid encoding the recombinant human L-isoaspartyl protein methyltransferase (rhPIMT) with an N-terminal polyhistidine tag was a generous gift of Dr. Bruce Downie 40. The rhPIMT was purified using a His-Trap nickel column (GE Healthcare, Waukesha, WI) and FPLC (BioRad Biologic HR workstation, Hercules, CA) to a concentration of 0.86 mg/mL and a specific activity of 5039 pmol of methyl groups transferred/mg/min. For quantitation, isoAsp residues in whole urine, as well as in the peptide (<10 kDa) and protein (>10 kDa) fractions, were labeled with rhPIMT and S-adenosyl-[\textit{methyl}^{3}H]-L-methionine ([\textit{3}H]AdoMet; PerkinElmer Inc., 78.0 Ci/mmol, 0.55 μCi/μL in 10 mM H\textsubscript{2}SO\textsubscript{4}: ethanol (9:1)). Specifically, 5 µL of sample was incubated with 5 µL rhPIMT, 5 µL [\textit{3}H]AdoMet (0.35 µM final concentration) and 85 µL of 0.2 M Bis-Tris, pH 6.4 at 37°C. After a 2 hr incubation, 15 µL of the reaction was mixed with 15 µL of 2× SDS-PAGE loading buffer (100 mM Tris-HCl, pH 6.8, 200 mM 2-mercaptoethanol, 4% SDS, 20% glycerol, and 0.1%
bromophenol blue) and heated for 3 min at 100 °C. SDS gel electrophoresis was performed by loading 25 µL of sample onto a NuPAGE Novex Bis-Tris 4-12% Mini Gel (Invitrogen, Carlsbad, CA) in a running buffer of 50 mM MES, 50 mM Tris base, 0.1% SDS, 1 mM EDTA at pH 7.3 (NuPAGE MES SDS running buffer NP0002). The samples were separated by electrophoresis for approximately 35 min at 200 V, and the gel was stained with Coomassie Brilliant Blue R-250. To identify the isoAsp species, fluorography was performed on the gel by soaking the gel for 1 hr with EN³HANCE (PerkinElmer, Waltham, MA) after destaining, and incubating the gel for 30 min in water. The gel was subsequently dried for 3 h at 80 °C before being exposed to film.

3.3.5 Proteolytic Digestion

Prior to LC-MS, urinary proteins from WT or KO mice were separately isolated by ultrafiltration (10 kDa M.W. cut-off) and denatured with 6 M guanidine hydrochloride in 50 mM ammonium bicarbonate, reduced with 10 mM dithiothreitol (DTT) for 30 min at 37 °C, and then alkylated with 20 mM iodoacetamide (IAA) in the dark for 30 min at room temperature. The solvent was exchanged to the digestion buffer (50 mM ammonium bicarbonate, pH 8.3) by centrifuging at 10,000 ×g for 10 min for three cycles in a Millipore Ultrafree-0.5 ultrafiltration device (5 kDa cutoff). The sample was then digested with Lys-C at an enzyme:protein ratio of 1:50 (w/w) at 28 °C for 12 hrs, and next frozen at -80 °C to quench the digestion and minimize any potential deamidation. To investigate the level of deamidation and isomerization induced by the sample preparation process, a duplicate experiment was conducted in parallel following the above same procedure except that ¹⁸O water, and longer incubation times for digestion were investigated.
3.3.6 LC-MS Analysis

LC-MS experiments were performed on an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher, San Jose, CA) equipped with an ETD ion source. An Ultimate 3000 nano-LC pump ( Dionex, Mountain View, CA) was used to generate the gradient, and a self-packed reversed phase column (Vydac C18, 300 Å pore size, 5 μm particle size, 75 μm i.d. × 10 cm, Grace, Deerfield, IL) was coupled online to the mass spectrometer through a Digital PicoView® nanospray source (New Objective, Woburn, MA). Mobile phase A consisted of 0.1 % formic acid in water and mobile phase B of 0.1 % formic acid in acetonitrile. Initially, a linear gradient from 2 % to 10 % mobile phase B over 2 minutes, and then from 10 % to 40 % mobile phase B over 90 minutes, was used to separate the digested mixtures at a flow rate of 200 nL/min. At least three blank runs were conducted to minimize any carryover effects between sample injections. The temperature of the ion transfer tube of the linear ion trap was set at 245 °C and the electrospray voltage at 2.2 kV. The mass spectrometer was operated in the data-dependent acquisition mode, switching automatically between one full MS scan at the mass range of 400-2000 m/z and sequential CID MS/MS scans for the nine most abundant precursor ions with a 3.0 mass unit isolation width. Dynamic exclusion was implemented with 2 repeat counts (repeat duration of 30 seconds, exclusion list 200, and exclusion duration of 30 seconds). The normalized collision energy for CID was set at 28%. The chemical ionization (CI) source parameters for fluoranthene, such as ion optics, filament emission current, anion injection time (anion target value set at 3 × 10^5 ions), fluoranthene gas flow, and CI gas flow, were optimized automatically, following the standard procedure for tuning the instrument. The time of the ion/ion reaction was maintained constant throughout the experiment at 150 ms. The
supplemental activation function was integrated into the ETD data acquisition method, with the collision energy optimized based on the measured charge state of precursor ion. A list of peptides that contain asparagine or aspartic acid was generated from the survey shotgun proteomics. The multiply charged precursor ions that are favorable to ETD fragmentation were then selected for targeted proteomic analysis. Briefly, a full MS scan with resolution 30,000 (at 400 m/z) in the Orbitrap MS, followed by targeted full-scan ETD MS/MS events in the ion trap MS, was applied to acquire MS and ETD MS/MS spectra of selected precursors.

3.3.7 Data Processing

LC-MS/MS data were analyzed using the Xcalibur 2.0.7 software (Thermo Fisher). All files were searched against the mouse SwissProt annotated database (updated in June 2011) using the Sequest algorithm in Proteome Discover 1.2 (Thermo Fisher). Cysteine carbamidomethylation was included as a fixed modification. Deamidation of asparagine was set as a variable modification, and the deamidated residues were assigned based on the +0.984 Da mass shift compared to that of the native form. Full Lys-C enzyme specificity was applied with up to two missed cleavage sites. Mass tolerances were set at 1 Da or 10 ppm for the precursor ions generated in the ion trap or Orbitrap-MS, respectively, and 1.0 Da for the fragment ions in the ion trap. For MS2 spectral assignment, a false discovery rate (FDR %) of less than 1% at the peptide level was targeted by applying the target-decoy database searching strategy. For all potential isoAsp containing peptides from targeted proteins, ETD-MS$^2$ spectra and extracted ion chromatograms were manually investigated to confirm the occurrence of isoAsp and its site. Multiple criteria were manually applied as described in the Results and Discussion. Among the multiply charged ions of a given
peptide, the precursor with the highest intensity was used to extract the chromatographic peak area with a 1.0 Da extraction window in the full MS scan. The peak areas from the XICs were then employed to estimate the relative amounts of the Asn, Asp, and isoAsp isoforms.

3.4 Results and Discussion

3.4.1 Secreted Proteins/Peptides in Urine of PIMT-Knockout Mice Contain Significant and Elevated Amounts of isoAsp

The total isoAsp levels of the peptide (<10 kDa) and protein (>10 kDa) content in mouse urine were determined by enzymatic labeling coupled with SDS-PAGE, as described previously\textsuperscript{39}. As shown in Figure 3-2, judged by Coomassie staining, proteins at 21 kDa in male urine, and 66 and 21 kDa in female urine, appeared to be the dominant proteins. Based on autoradiography, for the urine of wild type male and female mice or PIMT knockout female mice, only small amounts of methylated protein species (i.e. isoAsp content) were observed at slightly higher than the background level. In comparison, significantly elevated isoAsp levels in urine proteins from PIMT-knockout male mice were observed, as shown in Figure 3-2 (bands with apparent molecular weight of 27, 21, and 11 kDa, respectively). All of these species were concentrated in the protein fraction of the urine (excretion fraction, > 10 kDa) as compared with the peptide fractions (degradation fraction, <10 kDa) or the unfractionated urine. The overall changes in isoAsp amount were consistent with our previous results\textsuperscript{20}. However, for the first time, the peptide and protein fractions were analyzed separately in this work. The data strongly suggest that both excretion (intact isoaspartyl-containing proteins) and proteolysis (degraded peptides, followed by excretion) are operative in the removal of isoAsp species from circulation in mice. Since the excretion
of intact damaged proteins was significantly elevated in knockout male mice, we focused on the fraction of urinary proteins (> 10 kDa) from male mice (WT or KO) and applied our proteomics workflow to identify isoAsp-containing proteins, their modification sites, and relative amounts.

Figure 3-2. Enzymatic detection of isoAsp in mouse urinary species via PIMT-catalyzed methylation followed by fluorography of $^3$H-methyl ester after SDS-PAGE. Isoaspartyl groups in unfractionated (whole) urine were compared to fractionated urine with species less than 10 kDa (<10; peptide fraction) and greater than 10 kDa (>10; protein fraction). The samples were labeled with $^3$H-methyl groups from $^3$H-AdoMet catalyzed by rhPIMT and separated by SDS-PAGE. The dried and EN$^3$HANCEd Coomassie-stained gel is shown on the left, and the fluorograph on the right (82 day exposure at -80 °C using Kodak BioMax XAR film). The asterisk (*) denotes the position of auto-methylated rhPIMT.
3.4.2 Proteomic Analysis of Mouse Urine Proteins

As described in the Experimental Section, urine protein fractions (> 10 kDa) from WT or KO male mice were digested with Lys-C. A non-biased proteomic analysis with data-dependent CID scans was first performed to identify proteins in mouse urine. Injection of approximately 5 µg of the Lys-C digest into the nano-LC column (the maximum loading amount) led to the identification of 26 proteins with at least two unique peptides. These proteins mainly were soluble proteins from epithelial cell secretion, such as epidermal growth factor (EGF), sulfated glycoprotein 1 (SAP), and kallikrein-1 (KLK1), or from glomerular filtration of plasma proteins, such as major urinary proteins (MUPs) and albumin (ALBU)\textsuperscript{42,43}.

The top five proteins identified belong to the mouse MUP family, including MUP1, MUP2, MUP3, MUP6 and MUP8 (see Table 3-1) with sequence coverage ranging from 54% to 98%. MUPs are a group of low molecular weight (~21 kDa) secreted proteins produced mainly in the liver\textsuperscript{44} that are abundant in the urine and other secretions of many mammals, especially males\textsuperscript{45,46}. Based on: a) the dominance of MUPs in male mouse urine and b) the elevated isoAsp level at the 21 kDa protein band in urine from KO male mice, we next targeted MUP species in male urine for comprehensive isoAsp analysis.

3.4.3 Mapping of isoAsp Sites in MUPs by ETD-SRM Analysis

To further investigate isoAsp in MUPs, a sensitive and targeted strategy was developed. SRM with ETD MS/MS (ETD-SRM) was implemented based on a full-scan MS with high resolution in the Orbitrap, followed by selected full-scan ETD MS/MS in the ion trap to continuously acquire MS and ETD MS/MS spectra of MUP peptides. Asn- or Asp-
containing peptides in MUPs (found in the above proteomic experiment), representing all possible peptide sequences that could contribute to isoAsp formation were selected, and their precursor mass and retention time were employed to establish SRM scanning events. In our study, peptides with only singly charged ions, very short peptide fragments (< 5 amino acid residues), miscleaved peptides by Lys-C, and sequence-redundant peptides were excluded from the SRM assay. Those candidate peptides that can generate triply (or higher) charged precursor ions are preferred for the ETD-SRM method. Based on these criteria, 13 candidate peptides were selected (see Table 3-2), with 38 Asn or Asp residues, i.e. potential sites for deamidation and/or isoaspartyl isomerization.
Table 3-1. Identified proteins (>2 distinct peptides) in mouse urine by discovery proteomics.

<table>
<thead>
<tr>
<th>SwissProt Accession</th>
<th>Description</th>
<th>Protein Symbol</th>
<th>M.W./kDa</th>
<th>Sequence Coverage</th>
<th>Distinct Peptides</th>
<th>Spectral counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>P11588</td>
<td>Major urinary protein 1</td>
<td>MUP1</td>
<td>20.6</td>
<td>91.4</td>
<td>15</td>
<td>1373</td>
</tr>
<tr>
<td>P11589</td>
<td>Major urinary protein 2</td>
<td>MUP2</td>
<td>20.7</td>
<td>88.3</td>
<td>13</td>
<td>1331</td>
</tr>
<tr>
<td>P02762</td>
<td>Major urinary protein 6</td>
<td>MUP6</td>
<td>20.6</td>
<td>91.4</td>
<td>12</td>
<td>1287</td>
</tr>
<tr>
<td>P04938</td>
<td>Major urinary proteins 11 and 8</td>
<td>MUP8</td>
<td>17.5</td>
<td>98.0</td>
<td>12</td>
<td>1287</td>
</tr>
<tr>
<td>P04939</td>
<td>Major urinary protein 3</td>
<td>MUP3</td>
<td>21.5</td>
<td>54.3</td>
<td>5</td>
<td>77</td>
</tr>
<tr>
<td>P15947</td>
<td>Kallikrein-1</td>
<td>KLK1</td>
<td>28.8</td>
<td>39.9</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>Q61207</td>
<td>Sulfated glycoprotein 1</td>
<td>SAP</td>
<td>61.4</td>
<td>12.8</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>P01132</td>
<td>Pro-epidermal growth factor</td>
<td>EGF</td>
<td>133.1</td>
<td>12.4</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>P35459</td>
<td>Lymphocyte antigen 6D</td>
<td>LY6D</td>
<td>13.4</td>
<td>22.1</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>P07724</td>
<td>Serum albumin</td>
<td>ALBU</td>
<td>68.6</td>
<td>19.9</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>P15945</td>
<td>Kallikrein-1-related peptidase b5</td>
<td>K1KB5</td>
<td>28.7</td>
<td>10.3</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>P28825</td>
<td>Meprin A subunit alpha</td>
<td>MEP1A</td>
<td>84.1</td>
<td>11.7</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>P28790</td>
<td>Granulins</td>
<td>GRN</td>
<td>63.4</td>
<td>11.0</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Q9DAU7</td>
<td>WAP four-disulfide core domain protein 2</td>
<td>WFDC2</td>
<td>18.0</td>
<td>27.0</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>P15501</td>
<td>Prostatic spermine-binding protein</td>
<td>SPBP</td>
<td>22.0</td>
<td>14.1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Q8BZH1</td>
<td>Protein-glutamine gamma-glutamyltransferase 4</td>
<td>TGM4</td>
<td>75.5</td>
<td>16.3</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>P09803</td>
<td>Cadherin-1</td>
<td>CADH1</td>
<td>98.2</td>
<td>8.4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Q06890</td>
<td>Clusterin</td>
<td>CLUS</td>
<td>51.6</td>
<td>8.7</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>P08228</td>
<td>Superoxide dismutase [Cu-Zn]</td>
<td>SODC</td>
<td>15.9</td>
<td>24.7</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>P10923</td>
<td>Osteopontin</td>
<td>OSTP</td>
<td>32.4</td>
<td>15.0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Q9WTR5</td>
<td>Cadherin-13</td>
<td>CAD13</td>
<td>78.2</td>
<td>3.1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P06281</td>
<td>Renin-1</td>
<td>REN1</td>
<td>44.3</td>
<td>6.2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P00796</td>
<td>Renin-2</td>
<td>REN2</td>
<td>44.3</td>
<td>6.2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P56480</td>
<td>ATP synthase subunit beta, mitochondrial</td>
<td>ATPB</td>
<td>56.3</td>
<td>4.4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P49183</td>
<td>Deoxyribonuclease-1</td>
<td>DNAS1</td>
<td>32.0</td>
<td>16.6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Q02596</td>
<td>Glycosylation-dependent cell adhesion molecule 1</td>
<td>GLCM1</td>
<td>16.2</td>
<td>21.2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 3-2. Selected candidates for targeted ETD-SRM detection. The investigated sites are underlined. The listed peptides containing potential isoAsp were selected for targeted ETD-SRM.

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Proteins</th>
<th>m/z</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEDNGNFRLFLEQIHVLEK</td>
<td>MUP2</td>
<td>772.1</td>
<td>3</td>
</tr>
<tr>
<td>HGILRENIDLSSANRCLQARE</td>
<td>MUP1</td>
<td>648.8</td>
<td>4</td>
</tr>
<tr>
<td>INGEWHTILLASDK</td>
<td>MUP1, 2, 6, 8</td>
<td>533.0</td>
<td>3</td>
</tr>
<tr>
<td>IEDNGNFRLFLEQIHVLSENSLVLK</td>
<td>MUP1</td>
<td>710.9</td>
<td>4</td>
</tr>
<tr>
<td>TDYDNLMAHLINEK</td>
<td>MUP1, 2, 6, 8</td>
<td>608.6</td>
<td>3</td>
</tr>
<tr>
<td>AGIYMNIDGFNTFSILK</td>
<td>MUP3</td>
<td>1059.0</td>
<td>2</td>
</tr>
<tr>
<td>AGEYSVTYDGFTFTIPK</td>
<td>MUP1, 2, 6, 8</td>
<td>670.7</td>
<td>3</td>
</tr>
<tr>
<td>DGETFQLMGLYGREPDSLSDIK</td>
<td>MUP1, 2, 6, 8</td>
<td>824.4</td>
<td>3</td>
</tr>
<tr>
<td>FHTVRDEECSELSMVADK</td>
<td>MUP1, 2, 6, 8</td>
<td>718.3</td>
<td>3</td>
</tr>
<tr>
<td>FHLIVNNECTEMTAIGEQTEK</td>
<td>MUP3</td>
<td>827.1</td>
<td>3</td>
</tr>
<tr>
<td>TFQLMELYGREPDLSLDIK</td>
<td>MUP3</td>
<td>756.7</td>
<td>3</td>
</tr>
<tr>
<td>LCEEHIIRENIIDLTNVNCLEARE</td>
<td>MUP3</td>
<td>792.4</td>
<td>4</td>
</tr>
<tr>
<td>TDYDNYIMIHLINK</td>
<td>MUP3</td>
<td>876.9</td>
<td>3</td>
</tr>
</tbody>
</table>
A 90 min LC-MS gradient was divided into several time-based segments. The SRM data-acquisition window for each precursor was set at 10 min, based on the retention time of the selected unmodified peptide. The baseline separation of the deamidation products ensured the purity of the ETD spectra, thus minimizing the potential interference from other isoAsp and Asp isoforms, which share high similarity of ETD MS/MS spectra.

Furthermore, the ETD-SRM method was optimized to increase the efficiency of ETD fragmentation for isoAsp analysis, in order to maximize the intensity of c and z product ions, including the specific reporter ions of isoAsp residues (the pair of c<sub>i-n</sub>+57 and z<sub>n</sub>-57 ions). First, we utilized the decision tree algorithm<sup>47</sup> to select precursor ions with relatively high intensity. The measured masses ranging at 650-950 (m/z) and at least triply charged were the preferred candidate precursor ions for ETD fragmentation. Secondly, supplemental activation was implemented to enhance ETD fragmentation of doubly charged precursors<sup>48</sup>. Thirdly, an isolation width at ± 1.5 Da/e for the ETD-SRM method was selected for the multiply-charged precursor ions, allowing the inclusion of precursor ions of isoforms with at least 3 deamidated Asn residues in one peptide. The candidate peptides and their precursor ions (m/z) and charge states are listed in Table 3-2.

One targeted peptide, the MUP1 peptide (IEDNGNFRFLFLEQIHVLENSLVLK, 50-73; also shared with MUP6 and MUP8), containing four potential Asn deamidation and aspartyl isomerization sites, can serve to illustrate the data interpretation for identifying the isoAsp site(s). As shown in Figures 3-2 and 3-3, database searching resulted in 5 forms of the MUP1 peptide: the native form (peak III) and its four-deamidated isoforms (peaks I, II, IV, and V). To increase the confidence of isoAsp site mapping, multiple criteria were
implemented for manual investigation, particularly, the unique reporter ions (the pair of \( c_l \), \( c_l+57 \) and \( z_n-57 \) ions, the loss of 60.0 Da from the reduced species for Asp form) from more selective SRM scans. In brief, based on a +0.984 Da mass shift (from Asn to Asp or isoAsp) in the precursor ions, the peaks I, IV, and V were identified as mono-deamidated forms of the peptide, and a +1.968 Da mass shift for peak II as a di-deamidated product. Further investigation of the ETD spectra and XICs revealed that peak I contained a deamidated modification of Asn 68 to isoAsp 68, according to the +0.984 Da mass shift. With the same approach, peaks IV and V were identified with the deamidation from Asn 53 to Asp 53 and isoAsp 53, respectively. Peak II had the simultaneous deamidation sites of Asn53 to isoAsp53 and Asn68 to isoAsp68.

It is important to note that we found it a challenge to confidently identify all Asn deamidation and Asp isomerization sites in a proteome scale using a conventional shotgun proteomics. The randomly selected MS/MS scans in a data-dependent mode would not ensure the fragmentation of precursor ions at the chromatographic peak apex, thus leading to lower sensitivity of detection and poor quality of MS/MS spectra. When the automated software was used for assignment of modification sites, these low-quality spectra could potentially result in a high false positive rate. An example for such a challenge is demonstrated by the above peptide, MUP1 peptide (50-73). As shown in Figure 3-4, from the shotgun proteomics, the unmodified MUP1 peptide (50-73) (peak III) was identified from both WT and KO mice while its deamidated isoforms (either Asp or isoAsp or the combination of the two, peaks I, II, IV, and V) only from KO mice. The isoforms (peaks I, II, IV, and V) in relatively low abundances in wild type mice and their deamidation sites could not be confidently identified. In addition, the occurrence of isoaspartyl and aspartyl
residues in deamidated peptides could not be differentiated. The implementation of targeted proteomics using ETD-SRM method complemented the above limitation by providing enhanced sensitivity for detection and the differentiating isoAsp and Asp, based on the unique ETD-fragment ions. For the above deamidated isoforms of MUP1 peptide (50-73), the specific isoAsp/Asp sites were assigned, with their relative abundance estimated at less than 3% of the total abundance in the urine of WT mice.

In total, of 38 putative isoAsp-containing sites from the 13 candidate MUP peptides, 6 Asn residues were found to be deamidated (5 confirmed to form isoAsp), as summarized in Table 3-3. The remaining 14 Asn residues were observed with no modification, and 18 Asp residues were found with no detectable isomerization. These findings revealed the isoAsp sites that contribute to the elevated isoAsp level in MUP species in KO-male mouse urine. For the wild type mice, very low level of isoAsp was observed in urine, indicating that PIMT-catalyzed conversion from isoAsp to Asp is an efficient process.
**Figure 3-3.** Examination of isoAsp sites in MUP1 peptide (50-73). (a) ETD spectrum of peptide IED$_{52}^N$N$_{53}^5$FRLFLEQIHVLEN$_{68}^N$SLVLK and its isomers from MUP2; (b and c) the expanded ETD spectra of the peptide IED$_{52}^N$N$_{53}^5$FRLFLEQIHVLEN$_{68}^N$SLVLK and its isomers, leading to the identification of deamidation sites.
Figure 3-4. Extracted ion chromatograms and ETD spectra of MUP1 peptide (50-73), IEDNGNFRLFLEQIHVLENSLVLK. (a) Extracted ion chromatograms of MUP1 peptide and its isomers from wild type and PIMT-deficient mice; (b) ETD spectrum of native MUP1 peptide (50-73); (c,d) the reporter ions of isoAsp (c3+57, z6-57) in ETD MS/MS spectra from the selected precursors of above peptide and its isoaspartyl isoforms.
Table 3-3. Identification and quantitation of isoAsp peptides in MUPs.

<table>
<thead>
<tr>
<th>Sequences and modifications</th>
<th>Δ mass (Da)</th>
<th>Relative amount (%)</th>
<th>*Fold change (KO/WT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IED<strong>N</strong>^53^GN**^55^FRLFLEQIHLEN**^68^SLVLK, MUP1(50-73), also shared with MUP6 and MUP8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn68 to isoAsp68</td>
<td>0.984</td>
<td>11.8</td>
<td>2.2</td>
</tr>
<tr>
<td>Asn53, Asn68 to isoAsp53, isoAsp68</td>
<td>1.968</td>
<td>12.0</td>
<td>2.2</td>
</tr>
<tr>
<td>Native form</td>
<td>0.000</td>
<td>4.5</td>
<td>93.0</td>
</tr>
<tr>
<td>Asn53 to Asp53</td>
<td>0.984</td>
<td>3.3</td>
<td>1.5</td>
</tr>
<tr>
<td>Asn53 to isoAsp53</td>
<td>0.984</td>
<td>8.4</td>
<td>1.1</td>
</tr>
<tr>
<td>HGILREN**^165^IID**^168^LSN**^171^AN**^173^RCLQARE, MUP1(159-180)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn173 to isoAsp173</td>
<td>0.984</td>
<td>2.2</td>
<td>1.3</td>
</tr>
<tr>
<td>Native form</td>
<td>0.000</td>
<td>92.6</td>
<td>93.4</td>
</tr>
<tr>
<td>Asn171 to Asp171</td>
<td>0.984</td>
<td>2.7</td>
<td>3.5</td>
</tr>
<tr>
<td>Asn173 to Asp173</td>
<td>0.984</td>
<td>2.6</td>
<td>1.8</td>
</tr>
<tr>
<td>IED**^56^N**^58^GEWHTIILASD**^46^K, MUP1(33-46), also shared with MUP2, MUP6, and MUP8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn34 to Asp34</td>
<td>0.984</td>
<td>6.0</td>
<td>3.1</td>
</tr>
<tr>
<td>Native form</td>
<td>0.000</td>
<td>45.3</td>
<td>43.0</td>
</tr>
<tr>
<td>Asn34 to isoAsp34</td>
<td>0.984</td>
<td>48.7</td>
<td>53.9</td>
</tr>
</tbody>
</table>

(*) indicates the estimated fold change by dividing the peak area of each peptide isoform of PIMT-deficient mice by that of the wild type mice. The positive and negative values show the elevated and decreased level, respectively, in PIMT-deficient mice.
3.4.4 IsoAsp Formation during Sample Preparation

As illustrated in Figure 3-1, isoAsp is spontaneously generated from the deamidation of Asn or isomerization of Asp. Thus, isoAsp might occur during sample preparation. To examine the potential of this artifact, we conducted Lys-C digestion in $^{18}\text{O}$-labeled water$^{25,26,49,50}$. Under such conditions, deamidation products would contain $^{18}\text{O}$, resulting in an increase of mass of 2 Da. Briefly, mice urine was treated at 37 °C and pH 8.3 overnight, stress conditions that are harsher than those for the sample analysis described above. Nonetheless, the "hot spot" NG sequence in two peptides, IEDGNFRLFLEQIHVLEK and INGEWHTIILASDK generated only less than 5% and 3% of artificial deamidation products, respectively. No isoAsp was detected from the other asparaginyl and aspartyl residues during sample preparation. Thus a negligible level of artificial isoAsp was generated during our sample analysis.

3.4.5 IsoAsp in MUPs from Wild Type and PIMT-Knockout Mice

Once the candidate peptides were characterized, the precursor mass of the native forms and isoAsp counterparts were extracted, and the XIC peak areas used for quantitation. Since various forms of a given peptide were observed in the same run, the estimated stoichiometry (ES, %) was used as a quantitative measure of the extent of deamidation and isomerization at each site, as utilized in a previous study$^{34}$. Briefly, the estimated stoichiometry was obtained from the peak area of the measured individual isoform divided by the sum of all forms of the given peptide. As an example, the percentages of native MUP1 peptide (50-73), IEDGNFRLFLEQIHVLENSLVLK, and the corresponding four deamidated isoforms (isoAsp, Asp and the combination thereof) in the wild type and PIMT-deficient mice are shown in Figure 3-2 and Table 3-3. PIMT-deficient male mice urine
showed a higher level of isoAsp than that of the wild type mice urine. As a particularly striking example, 8.4% of the isoAsp isoform, IEDisoDGNFRLFLEQIHVLENSLVLK, in the PIMT-deficient mouse was found, compared to only 1.1% in the wild type mouse, a nearly 8-fold higher isoAsp level. As another example, a doubly modified isoAsp peptide, IEDisoDGNFRLFLEQIHVEisoDSLVLK, had almost a 6-fold increase in PIMT knockout mice, relative to the wild type. As seen in Table 3-3, the relative quantitation analysis for other isoAsp-containing peptides also showed a trend of elevated level of deamidation-induced isoAsp in PIMT-deficient mice. The identification and quantitation of isoAsp sites will be important in future study examining the molecular basis of the excretion processes.

3.5 Conclusions

For the first time, both the secreted functional proteins (native MUPs) and excreted intact damaged proteins (deamidated and isoaspartyl-containing MUPs) in mouse urine have been systematically analyzed at the molecular level for their relative amount of isoAsp, protein identity, and sites of modifications. Our results strongly suggest that both excretion (intact isoAsp-proteins) and proteolysis (degraded isoAsp-peptides followed by excretion) are operative in the removal of isoAsp species from circulation in mice, in addition to the known repair process driven by the PIMT enzyme. From the analytical perspective, isoAsp posed unique technical challenges. The ETD-SRM targeted proteomics strategy has been shown to be particularly effective in providing confident identification of the sites of isoAsp/Asp and dramatically enhancing the sensitivity of isoAsp detection and quantitation. The strategy developed here allows the monitoring of isoAsp in individual proteins and specific sites within, so that the proposed excretion and proteolysis pathways for isoaspartyl containing proteins can now be investigated at the molecular and sequence level. Moreover,
our strategy is equally useful to examine isoAsp isoforms that are ubiquitous in protein pharmaceuticals and that accumulate during both storage and circulation.

3.6 Appendix

3.6.1 Protein Identification by In-Gel Digestion and LC-MS Analysis

During initial study of this project, in-gel digestion was performed to identify the proteins. The urine sample was diluted with SDS running buffer to 15 μL and then mixed with 5 μL LDS sample loading buffer (4X). The mixture was incubated at 100 °C for 5 min. The samples were loaded onto an SDS-PAGE gel (4-12% Bis-Tris polyacrylamide). The voltage was set at 160 V for 60 min, and then the gel was stained with Coomassie blue. The bands of interest (about 15 kDa to 20 kDa) were cut, destained with acetonitrile and 100 mM ammonium bicarbonate in water until no blue color was observed. 100 μL of 10 mM dithiothreitol (DTT) and 100 mM ammonium bicarbonate in water was added to the destained dried gel species for reduction of disulfide linkage. The reaction was incubated at 56 °C for 30 min. The gel was dried by acetonitrile, and then 200 μL of 20 mM iodoacetamide (IAA) in 100 mM ammonium bicarbonate in water was added for alkylation of the cysteines. The reaction was incubated in the dark, at room temperature for 60 min. The excess of reagents was removed by washing the gel with acetonitrile and 100 mM ammonium bicarbonate. The gel was rehydrated in 50 mM ammonium bicarbonate with 1 μg Lys-C and incubated at 4 °C for 15 min. Then, the digestion was performed at 28°C overnight. The digests were extracted using acetonitrile and 5% formic acid, and then dried by speed-vacuum. After LC-MS analysis (the same method as described in the experimental section), the major proteins in these bands were identified as proteins from the
MUP family, including MUP1, MUP2, MUP3, MUP6 and MUP8. These data were consistent with the data from in-solution digestion.

3.7 References


(39) Kim, E.; Lowenson, J. D.; MacLaren, D. C.; Clarke, S.; Young, S. G.  


Chapter 4: Complete Mapping of a Cystine Knot and Nested Disulfides of Recombinant Human Arylsulfatase A by Multi-Enzyme Digestion and LC-MS Analysis Using CID and ETD


Co-authors’ work in this chapter: Wenqin Ni: experimental design and perform, data analysis, manuscript writing and revision; Melanie Lin, Paul Salinas, Philip Savickas: sample providing, experimental discussion, manuscript revision; Shiaw-Lin Wu: idea contribution, experimental design, data analysis, manuscript writing and revision; Barry L. Karger: idea contribution, manuscript writing and revision and grant support.
4.1 Abstract

Cystine knots or nested disulfides are structurally difficult to characterize, despite current technological advances in peptide mapping with high-resolution liquid chromatography coupled with mass spectrometry (LC-MS). In the case of recombinant human arylsulfatase A (rhASA), there is one cystine knot at the C-terminal, a pair of nested disulfides at the middle, and two out of three unpaired cysteines in the N-terminal region. The statuses of these cysteines are critical structure attributes for rhASA function and stability that requires precise examination. We used a unique approach to determine the status and linkage of each cysteine in rhASA, which was comprised of multi-enzyme digestion strategies (from Lys-C, trypsin, Asp-N, pepsin, and PNGase F) and multi-fragmentation methods in mass spectrometry using electron transfer dissociation (ETD), collision induced dissociation (CID), and CID with MS3 (after ETD). In addition to generating desired lengths of enzymatic peptides for effective fragmentation, the digestion pH was optimized to minimize the disulfide scrambling. The disulfide linkages, including the cystine knot and a pair of nested cysteines, unpaired cysteines, and the post-translational modification of a cysteine to formylglycine, were all determined. In the assignment, the disulfide linkages were Cys138–Cys154, Cys143–Cys150, Cys282–Cys396, Cys470–Cys482, Cys471–Cys484, and Cys475–Cys481. For the unpaired cysteines, Cys20 and Cys276 were free cysteines, and Cys51 was largely converted to formylglycine (>70%). A successful methodology has been developed, which can be routinely used to determine these difficult-to-resolve disulfide linkages, ensuring drug function and stability.

4.2 Introduction
Cystine knots are a structural motif with three disulfides (six cysteine residues in close proximity in a protein backbone), with one of the disulfides passing through a ring, formed by the other two disulfide bonds \(^1\). Cystine knots are known to enhance protein structural stability, and they can be found in many proteins with a wide range of biological functions, such as inhibition, growth stimulation, and cyclization \(^2,3\). However, the cystine knot family shows low sequence homology, and it is therefore hard to predict cystine knot signatures by sequence alignment. Furthermore, there are 15 ways to form three disulfides into a cystine knot, and the determination of the correct assignment of the disulfide bonds is a challenge. There is a clear need to develop methodology that can be used routinely to determine the disulfide linkages, thus providing structural information for protein function and stability.

Arylsulfatase A (ARSA), a lysosomal enzyme, contains a cystine knot. ARSA catalyzes the hydrolysis of cerebroside sulfate to cerebroside and sulfate. Deficiency of this enzyme cumulates cerebroside sulfate and leads to the destruction of myelin in the central and peripheral nervous systems resulting in a progressive demyelination disease known as metachromatic leukodystrophy (MLD) \(^4\). MLD is an autosomal recessive disease with late infantile, juvenile, and adult forms, and is a terminal illness. Most children with the infantile form die by age 5 years. Symptoms of the juvenile form progress with death occurring 10 to 20 years following onset, and those persons affected by the adult form typically die within 6 to 14 years following onset of symptoms.

Patients with MLD have been reported to have a disruption of the cystine knot by the mutation of Cys 470 to Arg \(^5\). Recent studies have shown that a partially or fully reduced
cystine knot makes the protein susceptible to chemical or proteolytic degradation. The conformation of ARSA, forming a homo-dimer protein at neutral pH and a homo-octamer at acidic pH (i.e., in the lysosome), requires proper disulfide linkages. The stability of the enzyme seems to relate to the dimer-to-octamer transition in the lysosomal milieu, in which formation of the octamer has been shown to be disrupted by the replacement of Cys282 with phenylalanine.

Recombinant human arylsulfatase A (rhASA) with the sequence homology to ARSA has been investigated for use in enzyme replacement therapy, a potential treatment for MLD patients. Thus, characterization of disulfides in rhASA is an important structure attribute for biopharmaceutical manufacturing to maintain drug function and stability. Currently, experimental approaches for the characterization of cystine knots include X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR), which can be time-consuming (X-ray), and inconclusive because of the near proximity of six cysteines (NMR). The development of an effective and robust methodology is needed for manufacturers’ ability to routinely maintain and control the drug quality.

A typical strategy for cysteine characterization included proteolytic cleavage of the backbone using an appropriate endoprotease. Comparison of the reduced and non-reduced peptide maps leads to the identification of the bound and free cysteines. However, it is challenging to correctly assign the disulfide linkages for nested disulfides or cystine knots using this methodology. Recently, Wu et al. established a methodology using LC-MS with collision induced dissociation (CID), electron transfer dissociation (ETD), and CID of the isolated charge-reduced ions (MS3) to determine complicated and intertwined disulfides.
including scrambling\textsuperscript{13-15}. In this work, we applied a multi-enzyme digestion strategy combined with CID, ETD, and CID-MS3 to characterize the complete cysteine status in rhASA. Several digestion protocols with different pH were evaluated to determine whether the free cysteines of the protein promote disulfide scrambling under alkaline conditions typically used for enzymatic digestion. By optimizing specific protocols for each cysteine, the status of all the cysteines in rhASA, including the disulfide linkages from the cystine knot and nested disulfide, were successfully determined.

4.3 Experimental

4.3.1 Samples

rhASA, GMP lot JPT11001, manufactured by Shire Human Genetic Therapies (Lexington, MA, USA), was provided at 39.1 mg/mL. The sample was aliquoted (10 μL or 391 μg per vial) and stored at -80 °C before analysis.

4.3.2 Reagents

Sequencing-grade trypsin was purchased from Promega (Madison, WI, USA). Mass spectrometry grade lysyl endopeptidase (Lys-C) was obtained from Wako (Richmond, VA, USA). Pepsin was purchased from MP Biomedicals (Santa Ana, CA, USA). Asp-N, PNGase F, ammonium bicarbonate (NH\textsubscript{4}HCO\textsubscript{3}), and formic acid (FA) were from Sigma-Aldrich (St. Louis, MO, USA). LC-MS grade water was purchased from J.T. Baker (Phillipsburg, NJ, USA), and HPLC grade acetonitrile from ThermoFisher Scientific (Fairlawn, NJ, USA). Amicon centrifugal filter units (10 kDa MWCO) were obtained from Millipore (Bedford, MA, USA).

4.3.3 Enzymatic Digestion
The protein solution (2.5 μL of 39.1 mg/mL) was buffer exchanged with 100 mM ammonium bicarbonate (pH 8) or 50 mM Tris-HCl buffer (pH 6.8) using a 10 kDa molecular weight cutoff filter and concentrated to 2 mg/mL (49 μL). In a separate study, a slightly less than alkaline pH (pH 6.8) was used to examine the effect of pH on the formation of alternative disulfide linkages during the digestion procedure. If a difference was observed, pepsin digestion at pH 2 was used to eliminate scrambling that can occur under basic pH conditions. For pepsin digestion, the protein solution was buffer exchanged with 10 mM HCl (pH 2). Pepsin (1:10, wt/wt) was added to the protein solution and incubated at 37 °C for 30 min. The reaction was quenched by adjusting the pH to 6 with 10 mM sodium hydroxide. For Lys-C plus trypsin digestion, the protein solution (pH 6.8 or 8) was incubated with endoproteinase Lys-C (1:50 wt/wt) and trypsin (1:50 wt/wt) for 8 h room temperature. The enzyme was added a second time (1:50 wt/wt for each enzyme), and the digestion was allowed to occur for an additional 12 h at room temperature. For Lys-C plus trypsin, Asp-N, and PNGase F digestion, the protein solution (pH 6.8 or 8) was treated with the combination of endoproteinase Lys-C (1:50 wt/wt), trypsin (1:50 wt/wt), Asp-N (1:50 wt/wt), and PNGase F (1 units/10 μg) for 8 h at room temperature; the mixture of enzymes was then added a second time (the same ratio for each enzyme), and the digestion was allowed to occur for an additional 12 h room temperature. After checking the digestion efficiency, no differences could be observed for PNGase F added either to the same mixture or prior to the mixture. For simplicity, PNGase F is added in the same mixture. In all cases except pepsin digestion, digestion was terminated by the addition of 1 % formic acid. An aliquot of 2 μg of the enzyme digest was analyzed per LC-MS run.

4.3.4 LC-MS
An Ultimate 3000 nano-LC pump (Dionex, Mountain View, CA, USA) and a self-packed C18 column (Magic C18, 200 Å pore, and 5 μm particle size, 75 μm i.d. × 15 cm) (Magic C18 particle from Michrom Bioneresources, Auburn, CA, USA) was coupled to an LTQ-Orbitrap ETD XL mass spectrometer (ThermoFisher Scientific, San Jose, CA, USA) equipped with a nanospray ion source (New Objective, Woburn, MA, USA). Mobile phase A was 0.1 % formic acid in water, and mobile phase B was 0.1 % formic acid in acetonitrile. The peptides were eluted at 200 nL/min using a linear gradient from 2 % to 60 % B in 90 min, followed by 60 % to 80 % B over 10 min. The LTQ-Orbitrap-ETD XL mass spectrometer was operated in the data-dependent mode to switch automatically between MS (scan 1 in the Orbitrap), CID-MS2 (scan 2 in the LTQ), and ETD-MS2 (scan 3 in the LTQ). Briefly, after a survey MS spectrum from m/z 300 to 2000, subsequent CID-MS2, and ETD-MS2 steps were performed on the same precursor ion with ±2.5 m/z isolation width, with ion/ion reaction duration time being maintained constant throughout the experiment at 100 ms. CID-MS2 and ETD-MS2 spectra were repeated by targeting specific ions, in order to gain linkage information not obtained in the initial run. These targeted approaches, using the Orbitrap in scans 2 and 3 (if needed), were repeated (e.g., targeting multiple charges of a precursor ion or the same disulfide-linked peptide but with different enzymatic cleavage sites or missed cleavages) until the linkage information was complete. If necessary, the ions of interest obtained with ETD-MS2 were targeted for CID-MS3.

4.3.5 Disulfide Assignment

The expected disulfide-linked tryptic or multi-enzyme digested peptide masses with different charges were first calculated and then matched to the observed masses in the LC-MS chromatogram. The matched masses (with <5 ppm mass accuracy for highly abundant
ions and <20 ppm for low abundant ions, as determined by 10 % above or below the main peak) were further verified by analysis of the corresponding CID-MS2 and ETD-MS2 fragmentation spectra, as well as the CID-MS3 fragmentation spectra, as needed. Any internal cleavages (e.g., simultaneous cleavages at both the P1 and P2 polypeptides, or two cleavages within an intra-linked disulfide, were assigned manually. For these assignments, the other cleavages (i.e., the portion cleaved from the association with internal cleavages) should be found to confirm the assignment of the internal cleavages. In our experience, these internal cleavages for disulfide-linked peptides seem to occur more often than typical peptides (without disulfide linkages) and, thus, need to be paid more attention in assignment.

4.4 Results and Discussion

The primary structure of recombinant human arylsulfatase A (rhASA) with six disulfides linkages and three unpaired cysteines is shown in Figure 4-1. The crystal structure of human ARSA (without glycosylation) deduced these six disulfide linkages and three free cysteines, including one that is post-translationally modified to formylglycine. The cystine knot is formed at the C-terminal end of the molecule from the six cysteine residues indicated in the figure. It is important to note that ARSA is a glycoprotein with glycosylation sites indicated in Figure 4-1. It is clear that the complexity of the protein structure makes the accurate determination of the status of all cysteines very challenging. Multiple strategies are necessary for the elucidation of cysteines, as described below.
Figure 4-1. Primary structure of rhASA with disulfide linkages and unpaired cysteines.

Note: the sites of disulfide linkages and unpaired cysteines were colored in red and the N-glycosylation sites were underlined.

4.4.1 Digestion Strategy

A multi-enzyme digestion strategy is clearly needed for the complicated disulfide and unpaired cysteine structure for rhASA. In principle, identification of a single disulfide linkage is straightforward because there is usually only one possibility for connection. Consequently, proteases that can cut proteins into peptides containing only a single disulfide are desired. However, intertwined disulfides or a cysteine rich region in a protein such as in the case of rhASA may prevent enzyme digestion to the desired peptide size. Preferred peptide sizes are 1 to 5 kDa since peptide recovery and electrospray ionization efficiency can be a problem for larger peptides, while peptides less than 1 kDa may not retain well on a reversed-phase column. In some cases, the disulfide assignment will require further adjustment of peptide sizes to generate peptide lengths with sufficiently high-charge states
for effective ETD fragmentation\textsuperscript{14}. It should be noted that the enzymatic cleavages of the protein are the same using either trypsin or Lys-C plus trypsin. Nevertheless, the use of Lys-C plus trypsin seems to yield slightly higher digestion efficiency than trypsin alone. The reason could be that the protein size was reduced by Lys-C, leading to a more effective trypsin digestion. Thus, the selection of specific enzymes needs to be carefully considered. Also, for disulfide linked peptides containing N-linked glycosylation, an additional PNGase F treatment should be considered to reduce the complexity of the mass spectra. For peptides containing free cysteines, the digestion pH for the selected enzymes needs also to be optimized to maintain sufficient enzyme activity while avoiding scrambling. In this study, after surveying several enzyme combinations (Lys-C, trypsin, Asp-N, pepsin, and PNGase F), several protocols were developed for the full cysteine status of rhARSA. Table 4-1 lists the various digestion protocols including the fragmentation methods for the specific assignments. A detailed description of these steps is in the following sections, beginning with the unpaired cysteines, followed by the single disulfide, and then the nested disulfide, with the final section dealing with the complicated cystine knot.
Table 4-1. Unpaired cysteine and disulfide linkages in rhASA

<table>
<thead>
<tr>
<th>#</th>
<th>Site</th>
<th>Status</th>
<th>Enzyme</th>
<th>Corresponding peptide sequence</th>
<th>I.D. (method)</th>
<th>Theoretical mass (2+)</th>
<th>Observed mass (2+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C20</td>
<td>Free Cys</td>
<td>Pepsin</td>
<td>GCYGHPSSTTPNL (19-31)</td>
<td>CID</td>
<td>667.2957</td>
<td>667.2965</td>
</tr>
<tr>
<td>2</td>
<td>C51</td>
<td>Cys as fgly</td>
<td>Pepsin</td>
<td>YYPVSLC(fgly)TPSRAAL (45-58)</td>
<td>CID</td>
<td>729.9012</td>
<td>729.9020</td>
</tr>
<tr>
<td>3</td>
<td>C51</td>
<td>Free Cys</td>
<td>Pepsin</td>
<td>YYPVSLCTPSRAAL (45-58)</td>
<td>CID</td>
<td>738.8976</td>
<td>738.9059</td>
</tr>
<tr>
<td>4</td>
<td>C276</td>
<td>Free Cys</td>
<td>Pepsin</td>
<td>RMSRGGCGL (270-279)</td>
<td>CID</td>
<td>512.2448</td>
<td>512.2453</td>
</tr>
<tr>
<td>5</td>
<td>C282-C396</td>
<td>One disulfide linkage</td>
<td>Pepsin</td>
<td>LRCGKGETTETYEYGGVRE (280-294) FTQGSAHSDTADPACCASSSL (381-402)</td>
<td>CID/ETD</td>
<td>1271.9117 (3+)</td>
<td>1271.9153 (3+)</td>
</tr>
<tr>
<td>6</td>
<td>C138-C154, C143-C150</td>
<td>Two disulfide linkages</td>
<td>Lys-C+ Trypsin+ Asp-N+ PNGase F</td>
<td>DQGPCQ (134-139) DLTCPFPATPCOGC (140-154)</td>
<td>CID/ETD 1069.8997 (2+)</td>
<td>1069.9006 (2+)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>C470-C482, C471-C484, C475-C481</td>
<td>Three disulfide linkages</td>
<td>Pepsin</td>
<td>PALQICCHPCCFPRAPACCHCPCPD PHA (465-489)</td>
<td>CID/ETD/CIDMS3 874.3564 (3+)</td>
<td>874.3596 (3+)</td>
<td></td>
</tr>
</tbody>
</table>

Note: fgly means formylglycine, in which the side chain of cysteine, -CH2SH, is converted to -CHO.

4.4.2 Unpaired Cysteines: Cys20, Cys51, Cys276

When trypsin or Lys-C plus trypsin digestion was used to assign the unpaired cysteines, disulfide scrambling, which formed various disulfides mainly among the free cysteines, was observed using a standard digestion buffer at pH 8 (~40 %), and to a lesser extent at pH 6.8 (~5 %). As expected, we did not observe scrambled disulfides with pepsin.
digestion at pH 2. It should be noted that the scrambled disulfides obtained at higher pH (i.e., pH 8) provided us the types and linkage information. Thus, we could target these scrambled disulfides at the lower pH analysis. Although the amount of scrambled disulfides could be lower at the lower pH, the targeting approach (extraction of specific ions for targeted MS^n) should be more sensitive than in the initial discovery mode. Although pepsin digestion is non-specific, major cleavages were found after leucine residues (C-terminal side), followed by aromatic amino acids, proline, and glutamic acid residues. Thus, we could focus on these cleavages for potential scrambled disulfides as well. The major pepsin fragment containing the unpaired Cys20 was identified. The peptide with the corresponding mass and charge is shown in Figure 4-2a. As shown, the precursor ion scan was performed in the Orbitrap, and its monoisotopic mass accurately matched the theoretical peptide mass with an unmodified free cysteine, as m/z 667.2965 (observed) matched to m/z 667.2957 (theoretical) for the 2+ charge state. The site of the free cysteine was determined by CID-MS2 of the precursor ion, as shown in Figure 4-2b. The remaining unpaired cysteines were identified in a similar manner, as shown in Figure 4-3a and b (Cys51 converted to formylglycine), Figure 4-4a and b (Cys51, a free cysteine), and Figure 4-5a and b (Cys276, a free cysteine). Table 4-1 (#1, #2, #3, and #4) summarizes the assignments for all the unpaired cysteines. At Cys51, it contains more than 70 % of the formylglycine form. Since the ionization efficiency of the peptide containing the free cysteine could be different from that containing the formylglycine, the ratio of the two is a rough estimation of the formylglycine conversion.
Figure 4-2. (a) Mass and charge of the pepsin-digested peptide with an unmodified (free) Cys20, and (b) CID-MS2 spectrum of the precursor from (a). The sequence and theoretical mass of the peptide are indicated in the insert (a).
Figure 4-3. (a) Mass and charge of the pepsin-digested peptide with a modified Cys51, which is converted to formylglycine, and (b) CID-MS2 spectrum of the precursor from (a). The sequence and theoretical mass of the peptide are indicated in the insert (a).
Figure 4-4. (a) Mass and charge of the pepsin-digested peptide with an unmodified (free) Cys51, and (b) CID-MS2 spectrum of the precursor from (a). The sequence and theoretical mass of the peptide are indicated in the insert (a).
Figure 4-5. (a) Mass and charge of the pepsin-digested peptide with an unmodified (free) Cys276, and (b) CID-MS2 spectrum of the precursor from (a). The sequence and theoretical mass of the peptide are indicated in the insert (a).

4.4.3 Single Disulfide: Cys282–Cys396

For the peptide with a single disulfide (Cys282–Cys396, #5 in Table 4-1), the linkage assignment was straightforward. Although alkali pH (i.e., pH 8) should not cause the disulfide linked cysteines to scramble, the other free cysteines in the protein could potentially cross react with the disulfide at alkali pH. Indeed, a minute amount of cross-reacted disulfides was observed using a digestion buffer at pH 8. As expected, no cross-reacted disulfides could be observed when trypsin at pH 6.8 or pepsin (pH 2) was used for
the digestion. The assignment of the disulfide-linked peptide is illustrated by the pepsin digestion protocol (Figure 4-6). In Figure 4-6a, the observed accurate mass matched the theoretical peptide mass with one disulfide (a loss of 2H from the backbone sequence). The corresponding CID-MS2 spectrum, b and y ions in Figure 4-6b, verified the correct sequence. For the corresponding ETD-MS2 spectrum, the disulfide bond was preferentially dissociated as expected [14, 15], resulting in two dissociated peptides designated as P1 and P2 (Figure 4-6c), which confirms that the two peptides are linked together. In this disulfide linkage assignment, we used the low pH approach instead of alkylating the free cysteine to prevent scrambling since it is difficult to control the alkylation properly. Alkylation under denaturing conditions, the scrambling (particularly from free cysteine) can occur quickly before the alkylation step. If the protein is at native conditions, the alkylation step is often not optimized. Incomplete alkylation could cause the remaining free cysteine to scramble as well. Therefore, we used low pH to protonate (inactivate) the free cysteine to assign the disulfide linkages. In addition, with the use of low pH (not to alkylate the free cysteine), the free cysteine in the sequence can be assigned as well, as shown in Figure 4-2 to Figure 4-5.
Figure 4-6. (a) Mass and charge of the pepsin-digested peptide with one single disulfide linkage (Cys282–Cys396); (b) CID-MS2 spectrum of the precursor from (a); and ETD-MS2 spectrum of the precursor from (a). The sequence and theoretical mass of the peptide are indicated in the insert (a).
4.4.4 Nested Disulfide: Cys138–Cys154 and Cys143–Cys150

As shown in Figure 4-1, the cysteines for the nested disulfides are located in Cys138–Cys154 and Cys143–Cys150. Since there are four cysteines, other potential linkages could be either as two separate disulfides (Cys138–Cys143 and Cys150–Cys154) as well as two crossed disulfides (Cys138–Cys150 and Cys143–Cys154) (see Figure 4-7). Furthermore, the complexity is increased by two N-linked glycosylation sites, one within, and the other next to the two disulfides (N underlined in Figure 4-1). To reduce the complexity, the two N-linked glycans were removed with PNGase F, converting Asn (N) to Asp (D). This conversion provided a target for Asp-N digestion. Thus, in addition to Lys-C plus trypsin (to reduce the protein size), the addition of PNGase F and Asp-N enzymes effectively cut the disulfide-linked peptide to a suitable size for mass spectrometric analysis (see Figure 4-8). These nested disulfide bonds form a ring, which significantly reduces CID fragmentation efficiency for the amino acids inside the ring [13], thereby complicating the assignment for disulfide linkages inside the ring. Although ETD is effective to break the disulfides, the peptide length obtained by trypsin or pepsin alone is too large for effective fragmentation (m/z >1000) \(^{14}\). The digestion protocol required four enzymes to obtain the proper size for effective fragmentation by mass spectrometry (see Figure 4-8 and 4-9). The assignment of the disulfide-linked peptide based on the mass spectra is shown in Figure 4-10. In Figure 4-10a, the observed accurate mass matched the theoretical peptide mass with two disulfides (a loss of 4H from the backbone sequence). Since the ring structure formed by nested disulfides was broken by the additional Asp-N digestion, the disulfide linkages could be conclusively assigned as long as cleavages can be observed in the backbone between the CDGGC amino acid residues. As shown in Figure 4-10b, the y1, y3, b11, and
b12 fragments in the CID-MS2 spectrum provide strong evidence for the linkages Cys138 with Cys154, and Cys143 with Cys150. In addition to the CID-MS2 spectrum, the corresponding ETD-MS2 spectrum (Figure 4-10c) confirms that the two linked peptides (P1 and P2) are connected.

It should be noted that although Asp-N should cleave aspartic acid in the protein backbone, the aspartic acid residue adjacent to a cysteine (a disulfide) inside the ring was not cleaved (see Figure 4-8). For digestion at pH 8, a significantly scrambled disulfide was observed at a different LC retention time (see Figure 4-11), as the structure resembled to scramble 1 in Figure 4-7 (note: no scramble 2 could be observed). Without chromatographic separation, it would be difficult to determine the disulfide isomers since they are often isobaric. Nevertheless, the scrambled disulfides often have different configurations from the correct one. Thus, a different LC retention time or a shoulder peak with identical mass as the correct one is often the potential area for evaluation of scrambling. At pH 6.8, the scrambled disulfide was reduced to a trace amount and could not be observed at pH 2 with pepsin. Although the pepsin-digested disulfide could not be effectively fragmented by CID, the fragmentation did indirectly confirm the nested disulfide linkage (see Figure 4-12). ETD was also tested to fragment the pepsin-digested disulfide but was not successful, due to minimal fragmentation and mainly charge-reduced species in the ETD spectrum. Although CID-MS3 and even MS4 have been attempted to fragment the charge-reduced species, the fragmentation efficiency was still poor for the peptide of this size. As described above, the use of an additional enzyme (i.e., Asp-N) to obtain a proper size and configuration of the disulfides was critical.
**Figure 4-7.** The sequence of nested disulfide in trypsin digestion.

A (correct)

```
C138  C143  C150  C154
FLGIPYSHDQGPCQNLTCFPPATPCDGCDQGLVIPILLANLSDVEAQPPWLPGEAR
```

B (scramble 1)

```
C138  C143  C150  C154
FLGIPYSHDQGPCQNLTCFPPATPCDGCDQGLVIPILLANLSDVEAQPPWLPGEAR
```

C (scramble 2)

```
C138  C143  C150  C154
FLGIPYSHDQGPCQNLTCFPPATPCDGCDQGLVIPILLANLSDVEAQPPWLPGEAR
```

**Figure 4-8.** The digestion strategy for nested disulfides with multi enzyme digestion.

A (Lys-C + trypsin)

```
C138  C143  C150  C154
FLGIPYSHDQGPCQNLTCFPPATPCDGCDQGLVIPILLANLSDVEAQPPWLPGEAR
```

B (Lys-C + trypsin + PNGase F)

```
C138  C143  C150  C154
FLGIPYSHDQGPCQDLTCFPPATPCDGCDQGLVIPILLADLSVEAQPPWLPGEAR
```

C (Lys-C + trypsin + PNGase F + Asp-N)

```
FLGIPYSHDQPCOQLTCFPPATPCDGCDQGLVIPILLADLSVEAQPPWLPGEAR
```

```
Figure 4-9. The sequence of nested disulfides after Lys-C, trypsin, PNGase F and Asp-N digestion.
Figure 4-10. (a) Mass and charge of the Lys-C plus trypsin plus Asp-N plus PNGaseF-digested peptide with two disulfides (Cys138 with Cys154, and Cys143 with Cys150), and (b) CID-MS2 spectrum of the 2+ charged precursor from (a); (c) ETD-MS2 spectrum of the 3+ charged precursor.
Figure 4-11. (a) Chromatographic retention for correct and scrambled disulfides; (b) Precursor ion of scrambled disulfide (Cys138 with Cys143, and Cys150 with Cys154), and CID-MS2 spectrum of the 2+ charged precursor.
Figure 4-12. Precursor ion of nested disulfide (Cys138 with Cys154, and Cys143 with Cys150) peptide without glycosylation in pepsin digestion, and CID-MS2 spectrum of the 4+ charged precursor.

4.4.5 Cystine Knot: Cys470–Cys482, Cys471–Cys484, and Cys475–Cys481

The cysteine knot cannot be broken by any of the enzymes or combination of enzymes employed. In addition, CID fragmentation cannot produce backbone cleavages within the cystine knot. Thus, ETD was examined. For the amino acid sequence in this region, pepsin digestion was selected in order to obtain the proper peptide length with fewer acidic residues for effective fragmentation by ETD (i.e., eliminated additional glutamic and aspartic acid residues as compared to the corresponding tryptic fragment). The corresponding mass and charge of the pepsin-digested peptide is shown in Figure 4-13a.
The monoisotopic mass matched the expected peptide mass with three disulfides (a loss of 6H from the backbone sequence). Limited sequence information, as expected, was obtained by CID-MS2 (Figure 4-13b). Nevertheless, and significantly, ETD-MS2 dissociated the disulfides, which allowed cleavage of the peptide backbone, as shown in Figure 4-14a and -14b. The fragmentation of this disulfide linked peptide for two different charge states is shown in Figure 4-14a (m/z 656.30, 4+) and Figure 4-14b (m/z 525.20, 5+). The fragmentation data from the two different charge states demonstrates consistency with respect to cleavage sites and verifies that the linkage assignments are correct. Since the peptide was linked through three intertwined disulfides, a partial reduction of a particular disulfide (with mass shift by only 1 Da), the high resolution-accurate mass instrument (Orbitrap) with ETD provided even more convincing evidence for the disulfide bond assignments. As seen in both ETD spectra of Figure 4-14a and 14b, z7 and c18, along with the internal cleavages from the dissociated disulfide, confirm the connection between Cys471 and Cys484. In addition, one of the charge-reduced species (m/z 1312.6, [M + 4H]2+) in the ETD spectrum was further fragmented (CID-MS3 using the Orbitrap) as shown in Figure 4-14c. The MS3 spectra contain additional disulfide and backbone cleavages, such as y17 and b8, confirming the connection between Cys470 and Cys482. The fragmentation pattern and assignments were also observed with the same CID-MS3 spectrum generated in the LTQ ion trap (Figure 4-15), which makes the method applicable even with low resolution MS instruments. After assigning the two disulfide linkages, the non-dissociated (the third) disulfide was left with the only possible remaining connection, which was a linkage between Cys475 and Cys481. In summary, the combination of ETD-
MS2 and CID-MS3 mass spectral analysis confirms the linkage sites as Cys470 with Cys482, Cys471 with Cys484, and Cys475 with Cys481.

Figure 4-13. (a) Mass and charge of the pepsin-digested peptide with cystine knot (Cys470 with Cys482, Cys471 with Cys484 and Cys475 with Cys481), and (b) CID-MS2 spectrum of the 3+ charged precursor from (a).
Figure 4-14. (a) ETD-MS2 spectrum (using the Orbitrap) of the cysteine knot precursor (m/z 656.30, 4+), (b) ETD-MS2 spectrum (using the Orbitrap) of the same peptide from (a) but with different charged precursor (m/z 525.20, 5+), and (c) CID-MS3 spectrum (using the Orbitrap) of m/z 1312.6 from (a).
Figure 4-15. CID-MS3 spectrum (using ion trap) of m/z 1312.6 from ETD-MS2 spectrum (using ion trap) of the cysteine knot precursor (m/z 656.30, 4+).

4.5 Conclusions

In this study, in-depth LC-MS protocols have been developed to assign the status of all 15 cysteine residues in rhASA, including the disulfide linkages from the nested disulfide and cystine knot. Although both cystine knot and nested disulfides are difficult to resolve, strategies with a combination of different enzymes and MS fragmentation methods could successfully determine the assignments. The successful assignment of the disulfide linkages in the cystine knot demonstrates the power of the approach, which should be generally useful for other cystine knots. Using the described methods, it becomes feasible to monitor the disulfide linkages of recombinant rhASA.
4.6 Appendix

4.6.1 Data Annotation after Pepsin Digestion

Pepsin is a non-specific enzyme; thus data analysis of pepsin digestion sample is challenging. The data were searched against a database with the primary sequence of rhASA in Bioworks or Proteome Discoverer using a non-enzymatic search. After the database search, each peptide was manually confirmed by their mass accuracy, CID-MS/MS spectrum and intensity. The peptides with free cysteines were identified, as shown in Figure 4-2, 4-4 and 4-5. The peptide with the cysteine modified to formylglycine was manually checked with the mass shift of -17.9928 compared to the corresponding unmodified cysteine containing peptide, as shown in Figure 4-3. The unidentified or very low intensity peptides (usually 20 fold less than the major peaks) were listed. The disulfide linked peptides were in the list of unidentified or very low intensity peptides. The possible disulfide linked peptides were manually extracted and confirmed by CID-MS2, ETD-MS2 or CID-MS3. The inter-chain disulfide linkage between two peptides such as p1 and p2 is shown in Figure 4-6. The cysteine in the p1 peptide was assumed to be modified with the molecular weight of the peptide p2 with a loss of two hydrogen atoms. Similarly, the cysteine in the p2 peptide was assumed to be modified with the molecular weight of the peptide p1 with a loss of two hydrogen atoms. The theoretical precursor ion or product ions in CID-MS2 was calculated based on the cysteine modification. The intra-chain disulfide linkage is shown in Figure 4-13. Each cysteine in the disulfide linkage was assumed to lose one hydrogen atom. The theoretical mass of relative ions was calculated based on the cysteine modification.
4.7 References


Chapter 5: In-Depth Characterization of N-Linked Oligosaccharides Using Fluoride-Mediated Negative Ion Microfluidic Chip LC-MS


Co-authors’ work in this chapter: Wenqin Ni: experimental design and perform, data analysis, manuscript writing and revision; Jonathan Bones: idea contribution, experimental design and perform, manuscript writing and revision; Barry L. Karger: idea contribution, manuscript writing and revision and grant support.
5.1 Abstract

Characterization of N-glycans by liquid chromatography-positive electrospray ionization (ESI) tandem mass spectrometry (LC-MS/MS) using a microfluidic chip packed with porous graphitized carbon (PGC) represents a rapidly developing area in oligosaccharide analysis. Positive ion ESI-MS generates B/Y-type glycosidic fragment ions under collisional induced dissociation (CID). Although these ions facilitate glycan sequencing, they provide little information on linkage and positional isomers. Isomer identification in these cases is by retention on the PGC stationary phase where the specific structural isomers can, in principle, be separated. In this paper, we broaden the applicability of the PGC microfluidic chip/MS platform by implementing fluoride-mediated negative ESI-MS. Ammonium fluoride, added to the mobile phase, aids in the formation of pseudomolecular oligosaccharide anions due to the ability of fluoride to abstract a proton from the glycan structure. The negative charge results in the generation of C-type glycosidic fragments, highly informative A-type cross ring fragment ions and additional gas phase ion reaction products (e.g., D- and E-type ions), which, when combined, lead to in-depth oligosaccharide characterization, including linkage and positional isomers. Due to the separation of anomers by the PGC phase, comparison of oligosaccharides with an intact reducing terminus to their corresponding alditols was performed, revealing a more sensitive MS and, especially, MS/MS response from the glycans with a free reducing end. Fluoride also ensured recovery of charged oligosaccharides from the PGC stationary
phase. Application to the characterization of N-glycans released from polyclonal human and murine serum IgG is presented to demonstrate the effectiveness of the chip/negative ESI approach.

5.2 Introduction

N-glycosylation is a diverse but physiologically important post-translational modification that modulates the physical, chemical and biological properties of proteins \(^1\). Structural elucidation of complex carbohydrates is necessary to gain a deeper understanding of the potential role of these post-translational modifications as functional modulators of therapeutic proteins \(^2,3\) and also as contributors to disease, e.g. cancer and inflammation \(^4,5\). Liquid chromatographic and electrophoretic separation techniques, often coupled to mass spectrometry, have found widespread application for the characterization of protein N-glycosylation \(^6\). In particular, porous graphitized carbon (PGC) has become a popular LC stationary phase for hydrophilic oligosaccharides due to its ability to separate anomers and linkage and positional isomers \(^7\). Recently, a microfluidic chip containing an integrated nano-scale LC column packed with PGC has been introduced for nano-flow LC-MS (nLC-MS) analysis of oligosaccharides \(^8\). The chip offers the advantages of simplified operation and high reproducibility. Applications have been published demonstrating the characterization of oligosaccharides in the positive ion mode either as glycosylamines when coupled with an online PNGase F reactor or as alditols following off-line sample preparation \(^8-12\).
Positive mode ionization of oligosaccharides results in the formation of 
\([\text{M+H}^+]\) and \([\text{M+Na}^+]\) pseudomolecular ions and adducts \(^{13,14}\). Fragmentation of these ions by CID results in predominantly B/Y-type glycosidic cleavages \(^{15}\). These B/Y ions aid in the determination of the composition of the oligosaccharide; however, such ions generally do not allow elucidation of linkage and positional isomer information. Identification of these isomers is typically accomplished by retention differences on the PGC phase. Mass spectral structural analysis would clearly be useful to confirm the chromatographic identification, assuming an appropriate database was established. As an alternative approach, derivatization, such as permethylation, and multiple stages of tandem mass spectrometry have been used for isomer structure elucidation \(^{16-18}\). However, both the required sample amounts and associated method complexities are increased using such strategies. Furthermore, monosaccharide rearrangements during CID of protonated oligosaccharide cations, even following permethylation, have been reported which can complicate spectral annotation and structure assignment \(^{19}\).

Negative ion ESI-MS of oligosaccharides has been developed using either elevated pH to promote glycan deprotonation or adduction with anions \(^{20-22}\). Anion adduction represents an attractive approach as it can be performed at neutral or acidic pH, an important feature since the basic pH limit of the microfluidic commercial chip used in this work was 8.0 \(^{23}\). Anions can form hydrogen bonds with hydroxyl groups present on the sugar rings, and, depending on the gas phase basicity of the specific
anion, either stable negatively charged adducts can be formed [M + X] where X is the anion, or a proton can be extracted by X from the sugar ring generating [M - H] with the neutral loss of HX$^{24-27}$.

Relative to positive ESI, analysis of glycans in the negative ion mode has several potential advantages. First, and most importantly, CID MS/MS spectra of oligosaccharide anions contain, besides B/Y ions, C-type glycosidic and A-type cross ring fragments as well as highly informative D- and E-type ions, which, when combined, facilitate annotation of linkage and positional isomers$^{13,20-22,28}$. Secondly, monosaccharide rearrangements during CID fragmentation observed in positive ESI have generally not been seen for negative ESI$^{19}$. Migration of sulfate was recently reported using a linear ion trap; however, the migration appeared to be overcome when converting from CID to HCD fragmentation$^{29}$. On the other hand, the overall sensitivity in the negative ion mode is found to be lower than in the positive ion mode for neutral glycans$^{30}$; however, higher sensitivity in the negative ion mode can be observed for negatively charged glycans, e.g., those with sialic acids.

In this paper, we have broadened the applicability of the PGC chip coupled to MS for the analysis of glycans using negative ESI. The performance of different ammonium anions added to the mobile phase was first examined, from which ammonium fluoride at neutral pH in a water/acetonitrile mobile phase was selected. The high gas phase basicity of the fluoride ion was utilized to promote anion attachment and proton extraction from the oligosaccharides$^{26}$. Application of the
method for the analysis of glycans with a free reducing end and their corresponding alditol was also investigated, revealing more sensitive MS and, especially, MS/MS response for the non-reduced oligosaccharides. The combination of PGC separation and MS analysis resulted in high confidence structural elucidation of linkage and positional isomers without the need for additional strategies such as exoglycosidase digestion. The GlycoWorkBench platform was successfully used for spectral interpretation and structural annotation. To demonstrate the applicability of the chip-based nLC-MS negative ESI platform, N-linked oligosaccharides presented on polyclonal antibodies extracted from human and murine sera have been characterized, demonstrating the ability of the method to determine both linkage and positional isomers in complex oligosaccharide pools, including not only the traditional G0, G1/G1’ and G2 series but also the presence of galactose α-1-3 linked galactose residues and N-glycolyl neuraminic acid on the N-glycans liberated from the murine antibodies.

5.3 Experimental

5.3.1 Chemicals and Reagents

1,3-β-Laminarihexaose and 1,4-β-D-cellohexaose were purchased from Megazyme (Bray, Ireland). Galactosylated triantennary (NA3) was purchased from Prozyme (Hayward, CA). The dextran ladder and malto-oligosaccharides, Dp3-7, and all other reagents, including the murine polyclonal antibody, were from Sigma-Aldrich (St. Louis, MO). Standard N-glycans were purchased from Prozyme
(Hayward, CA), and peptide-\(N\)-glycosidase F (PNGase F) was from New England Biolabs (Ipswich, MA). LC-MS grade water was obtained from JT Baker (Phillipsburg, NJ) and LC-MS grade acetonitrile from Thermo Fisher Scientific (Fairlawn, NJ). Phytips, packed with 5 \(\mu\)L PGC, were a generous gift from PhyNexus (San Jose, CA). Microfluidic chips packed with PGC or for direct infusion were purchased from Agilent Technologies (Waldbronn, Germany).

5.3.2 Reduction of Oligasaccharides

Maltoheptaose (Dp6) or galactosylated triantennary (NA3) was dissolved in 0.1 M sodium borohydride and incubated at 65\(^\circ\)C for 1 hour, followed by quenching of the reaction \(\text{via}\) gradual addition of acetic acid. The sample was then purified using PGC packed micro-extraction Phytips, washed extensively with water, and subsequently eluted with 40\% v/v aqueous acetonitrile containing 0.1\% trifluoroacetic acid.

5.3.3 Purification of Polyclonal Human and Mouse IgG and N-glycan Release

Polyclonal IgG from human and murine serum (Sigma) were purified by Protein G enrichment in a microplate format (Pierce, Rockford, IL). Following elution and buffer exchange into sodium bicarbonate, pH 7.0, the glycans were enzymatically liberated by PNGase F, using an enzyme to protein ratio of 1:10 (v/v) at 37\(^\circ\)C overnight. Following incubation, the \(N\)-glycans were collected \(\text{via}\) centrifugation through a 10 kDa molecular weight cut-off filter, reduced to dryness
via vacuum centrifugation and treated with 1% v/v formic acid to promote conversion of the reducing terminal glycosylamine to the corresponding reducing sugar.

5.3.4 Data Analysis and Spectral Interpretation

Data analysis was performed on the Agilent MassHunter software B.02.00. Peaks were obtained using extracted ion chromatograms (EIC) generated with a 50 ppm mass accuracy window. The theoretical fragments from each oligosaccharide were automatically calculated by GlycoWorkBench Version 2.1 (Build 132) \(^3\) \(^2\). For automated annotation, MS/MS spectra were exported from MassHunter as Mascot generic format (.mgf). Mgf files were then loaded into Glycoworkbench as a peak list. Annotation of MS/MS spectra was performed using a combination of automatic searching with GlycoWorkBench in conjunction with manual verification. Annotation of fragment ions was as previously described by Domon and Costello \(^3\) \(^3\).

5.3.5 Negative Ion Electrospray Ionization Mass Spectrometry (nESI-MS)

Direct infusion, negative ionization electrospray mass spectrometry (nESI-MS) was performed on an Agilent 6520 Series Q-TOF instrument (Agilent Technologies, Palo Alto, CA) under the control of MassHunter Data Acquisition Software. Samples were prepared in 50% v/v water:acetonitrile containing 10 mM ammonium fluoride, ammonium chloride or ammonium bicarbonate. Samples were infused at a constant flow rate of 1 µL/min using a nano-spray voltage of 1950 V with a drying gas of 4 L/min nitrogen at 325°C. The fragmentor voltage was set at 175 V, and the skimmer voltage was at 65 V. For tandem mass spectrometry, the instrument
was operated using the automatic MS/MS switching mode, with the m/z range for MS from 400 to 2000 and 100 to 3000 for MS/MS spectral acquisition. The collision energy was set at 35 V. Data analysis details can be found in the supporting information.

5.3.6 LC-MS Analysis

Oligosaccharides were separated on an Agilent Technologies HPLC Chip II (Waldbonn, Germany), consisting of a 40 nL enrichment column and a 75 μm i.d. x 43 mm separation column, both packed with porous graphitized carbon. Mobile phase A was 10 mM ammonium fluoride in 97% water and 3% acetonitrile, and mobile phase B was 10 mM ammonium fluoride in 10% water and 90% acetonitrile. Samples were transferred to the microfluidic chip in 100% mobile phase A using the capillary pump at a flow rate of 4 μL/min. The nano-flow pump was employed to generate the analytical gradient with a flow rate of 500 nL/min. The gradient for standard oligosaccharides separation was 0% B, 0–2.5 min; 0–16% B, 2.5–20 min; 16–44% B, 20–30 min; 44–100% B, 30–35 min; and 100% B, 35–45 min with a return to initial starting conditions in 1 minute, followed by an isocratic hold for 20 minutes to ensure complete column re-equilibration.

5.4 Results and Discussion

Utilization of a microfluidic chip packed with porous graphitized carbon (PGC) for LC-MS of glycans has become a popular approach because of the ease of operation and reproducibility. To date, the platform has operated in the positive
ionization mode, which, as noted earlier, has limitations in the annotation of linkage and positional isomers. To broaden the applicability of the PGC chip/MS platform, we describe negative ESI LC-MS (Q-TOF) for the characterization of the \(N\)-glycans, demonstrating the effectiveness of the approach in the analysis of the \(N\)-glycans present on human and murine polyclonal serum antibodies. The PGC packing separates reducing end anomers, positional and linkage isomers, while negative ESI-MS yields in depth \(N\)-glycan analysis providing specific structural information not available by positive ESI-MS. Although the separated anomers can increase the complexity of the analysis, reduction can be performed for their removal. However, comparison of reducing sugars with their corresponding alditols revealed superior MS and MS/MS sensitivity for the reducing sugars (see below). The presence of fluoride in the mobile phase is not only demonstrated to successfully generate informative negative ESI but also to aid in the recovery of charged oligosaccharides from the retentive PGC phase.

5.4.1 Preliminary Investigation of Anion-Mediated Negative Ion Using Direct Infusion Q-TOF MS

Governed by an upper operating pH limit of pH 8.0 on the current commercial microfluidic chip, we first identified a series of organic and inorganic anions, present in the anionic form at pH 7.0, for evaluation for anion-mediated negative ionization of oligosaccharides on the Q-TOF MS. Previous reports have recommended, among others, the use of nitrate and phosphate for static infusion anion mediated MS of
oligosaccharides. However, considering the addition of the anion to the LC mobile phase and the flow rates used (0.5 μL/min), these anions were not considered acceptable, given the potential of ion source and MS fouling. From the initial list of anions identified, we selected fluoride, chloride and bicarbonate for initial study by direct infusion, based on their high gas phase basicity and suitability for LC-MS. Although fluoride was previously reported in one case to yield high levels of in-source fragmentation, other studies found the anion to be effective in improving sensitivity for small molecules in negative ion LC-MS.

Mass spectrometric parameters affecting ion transmission were first optimized to ensure optimal transport of the [M-H] and [M + anion adducts] through the mass analyzer for each of the three anions investigated. The fragmentor voltage, (voltage governing ion transport through the instrument optics) was varied from 90 to 250 V in 10 V intervals. Negligible in-source fragmentation occurred for all three anions when transmission voltages less than 180 V were employed. Above 180 V, the ion signal decreased, presumably due to in-source fragmentation. Separately, the collision energy (CE) was optimized to maximize CID-MS/MS spectral sensitivity. At low CE values, minimal glycan fragmentation was observed while at high CE values, glycosidic ions and cross-ring fragments ions were found to undergo additional cleavage leading to loss of spectral information. An optimum collision energy of 35 V was selected for use in all subsequent experiments for all three anions.
In the case of ammonium chloride, it was found that infusion of 10 mM ammonium chloride prepared in 50/50 acetonitrile/H$_2$O at 1 µL/min resulted in an intense series of ammonium chloride ion clusters which dominated the resulting MS spectra, in agreement with that reported by others when using FT-ICR-MS $^{35}$. These intense adduct clusters reduced overall sensitivity, with the resulting MS/MS spectra containing product ions of multiple chloride adducts that complicated annotation (data not shown). Based on these results, chloride was eliminated from further study.

We next examined fluoride and bicarbonate, again by infusion of the oligosaccharides prepared in 10 mM solutions of anion dissolved in 50/50 acetonitrile/H$_2$O. Although the use of ammonium bicarbonate yielded a buffer solution pH above the upper limit of the microfluidic chip, it was included as a benchmark based upon its use in previous studies $^{36}$. The infusion results revealed that the levels of adduction present in the mass spectra with bicarbonate were significantly higher than when fluoride was employed, see Figure 5-1 (A) vs. Figure 5-1 (B). This trend was not surprising given the higher gas phase basicity of fluoride compared to bicarbonate $^{26,27}$, resulting in fluoride more readily able to extract a proton from the oligosaccharide than bicarbonate. Given the higher levels of adduction in multiple forms for bicarbonate, the overall precursor MS sensitivity was 1.5 fold higher for fluoride as a dominant [M-H]$^-$ pseudomolecular oligosaccharide anion was produced, Figure 5-1 (B). Based on these results, the focus of our study became fluoride-mediated negative mode ionization. The successful use of fluoride in
the present study as opposed to other studies where fluoride was observed to yield significant in-source fragmentation\textsuperscript{20} appears to be due to the different source configuration and associated energetics of the instrument used in this study\textsuperscript{20,37}.

Infusing 10 mM ammonium fluoride in 50/50 (v/v) acetonitrile/H\textsubscript{2}O, the \([\text{M-H}]^-\) pseudomolecular anion of maltohexaose (glucose \(\alpha\)-1-4 glucose) was fragmented by CID leading to a highly informative spectrum containing an intense C-type glycosidic series for determination of the oligosaccharide sequence, Figure 5-1 (C). Importantly, and in contrast to the positive ion mode, intense \(^{2,4}A\), \(^{0,2}A\) and \(^{0,2}A\)-H\textsubscript{2}O ion series, arising from cross-ring fragmentation, were observed, leading to the determination of the specific linkage. For comparison purposes, positive ion CID-MS/MS was also performed on the \([\text{M+H}]^+\) pseudomolecular ion for maltohexaose. As expected, only B/Y-type glycosidic fragments were identified for positive ESI with no information rich cross ring fragments observed, see Figure 5-1 (D). Sensitivity was approximately five times higher in the positive, relative to the negative ion mode, as expected for neutral glycans\textsuperscript{8}; however, the overall level of information generated was higher in the negative mode, allowing for a more in-depth oligosaccharide characterization. Sensitivity would be expected to be higher in the negative ion mode for negatively charged glycans, e.g. sialylated oligosaccharides\textsuperscript{13}.  


Figure 5-1: Positive and negative ESI ion CID analysis of infused maltohexaose glucose-α-1-4-Glucose (Dp6). Precursor ions of [M-H]⁻ and adducts of Dp6 in negative ion analysis with (A) 10 mM ammonium bicarbonate and (B) 10 mM ammonium fluoride in 50% water and 50% acetonitrile; (C) annotated negative ion CID-MS/MS of the pseudomolecular anion of Dp6 [M-H]⁻ in 10 mM ammonium fluoride in 50% of water and 50% of acetonitrile; (D) annotated positive ion CID-MS/MS of the protonated ion of Dp6 [M+H]⁺ in 0.1% formic acid in 50% of water and 50% of acetonitrile. The infusion flow rate was set at 1 µL/min and the fragmentor voltage at 175 V. Additional details can be found in the Experimental Section.

As further examples of the success in linkage determination in fluoride-mediated negative ESI, cellohexaose (glucose-β-1-4-glucose), glucose-α-1,6-glucose
hexamer and laminarihexaose (glucose- β-1-3-glucose) were infused in the negative ion mode into the Q-TOF-MS. A strong C-type glycosidic ion series was observed for all three solutes, see Figure 5-2. A-type cross ring fragments were obtained for the cellohexaose (2,4 A, 0,2 A ions) and α-1-6-glucose hexamer (3,5 A, 0,3 A ions), suggesting the ability to decipher linkage orientation of the glycosidic bond from the MS/MS spectra, see Figures 5-2 (A) and (C), respectively. For the β-1-3 linked laminarihexaose, only C-type glycosidic ions were found in the CID-MS/MS spectrum of the [M-H]− pseudomolecular anion, see Figure 5-2 (B). The presence of a free hydroxyl group at the C-3 position was previously noted to be required for initiation of cross-ring fragmentation38; thus, the absence of an A-ion series in the MS/MS spectrum of laminarihexose is consistent with earlier reports20-22,38.
Figure 5-2: Annotated ESI negative ion CID-MS/MS analysis of [M-H]⁻ ions of infused (A) cellohexaose (glucose-β-1-4-glucose), (B) laminarihexaose (glucose-β-1-3-glucose) and (C) glucose-α-1,6-glucose hexamer. The collision energy was set at 35 V. Additional details in the Experimental Section.

When performing fluoride-mediated negative ionization for all the linear oligosaccharides, three precursor ions were observed: a dominant pseudomolecular [M-H]⁻ anion and two lower level fluoride adducts [M+F]⁻ and [2M-H+F]²⁻ of lower intensity. The CID MS/MS spectra of each adduct was found to be identical to that observed for the pseudomolecular [M-H]⁻ anion, see Figure 5-3 and Figure 5-1 (C), respectively. In contrast to other anions, no product ions were found that retained the anionic adducts in the MS/MS spectra. This lack of adduction of MS/MS product
ions also simplified data analysis using GlycoWorkBench, wherein spectral annotation could be automatically performed without the need for custom modification to account for the incorporation of any potential fluoride adduction. The dominant production of [M-H]\(^-\) compared to the fluoride adducts suggests that the gas phase basicity of the linear oligosaccharides is considerably lower than that of the highly basic fluoride anion, such that the extraction of a proton by fluoride with subsequent neutral loss as HF is facilitated \(^{24-27,39}\).

**Figure 5-3**: Annotated ESI negative ion CID-MS/MS analysis of infused Dp6: (A) \([2M-H+F]^{2-}\) ion, and (B) \([M+F]^{-}\) ion. The collision energy was set at 35 V.

Additional details in the Experimental Section.

### 5.4.2 Fluoride-Mediated Negative Ion nLC-MS/MS Using a Microfluidic Chip
Given the success of the infusion results with fluoride in negative ESI, we next turned to utilize the fluoride containing solvent system for LC separation on the PGC chip. The application of PGC-based stationary phases for oligosaccharide LC-MS was recently reviewed\(^7\). A concern when using PGC phases is incomplete analyte elution, particularly for sialylated oligosaccharides\(^{40,41}\). Therefore, the fluoride concentration in the nLC-MS mobile phase was investigated, in the range of 10 μM to 100 mM, using the neutral core fucosylated bi-antennary digalactosyl glycan (FA2G2) and the acidic bi-antennary digalactosyl disialylated glycan (A2G2S2) as model compounds to analyze recovery, ionization efficiency and spectral content. At low fluoride concentrations, A2G2S2 was retained on the PGC column (data not shown), while recovery was complete at fluoride concentrations greater than 10 mM. Thus, fluoride not only aids in negative analyte ionization but also in recovery from the PGC chip.

Under optimized conditions, both glycans eluted as a split peaks, due to the anomeric separation (α or β reducing-end anomers) by the PGC stationary phase, Fig. 5-4 (A) and (B) for FA2G2 and A2G2S2, respectively. Detection limits were estimated as 92 fmol and 12 fmol for FA2G2 and A2G2S2, respectively, determined as the sum of the two anomeric peaks with a S/N level of 3 for the lower level anomer. As expected for negative ESI, the sialylated species had a roughly 8 fold lower detection limit than the neutral species. Additionally, it is important to note that the detection limits can be further lowered using next generation Q-TOF
instrumentation containing ion funnels for more efficient ion transmission into the mass spectrometer \( ^{42} \).

CID-MS/MS fragmentation produced glycosidic ions and cross-ring fragments, as expected. The composition of the glycans was derived from the precursor ion for FA2G2 and A2G2S2, respectively, and fragment ions present in the resulting fluoride-mediated MS/MS spectra, Fig. 5-4 (A) and (B) for FA2G2 and A2G2S2, respectively. The MS/MS spectrum of the neutral FA2G2 N-glycan, Fig. 5-4 (A), was dominated by the \( ^{2,4}{A}_6 \) ion arising from cleavage across the reducing terminal GlcNAc residue, (m/z 738.7799, \( z=2 \)). The composition of the antennae was confirmed based on the presence of the \( ^{1,3}{A}_3 \) (m/z 424.1634, \( z=1 \)) and \( ^{0,2}{A}_2 - \text{H}_2\text{O} \) (m/z 263.0893, \( z=1 \)) fragments, the \( C_2 \) glycosidic fragment (m/z 382.1526, \( z=1 \)) and the \( D \) (m/z 688.2495, \( z=1 \)), D-18 (m/z 670.2422, \( z=1 \)) and E ions (m/z 466.1780, \( z=1 \)). The presence of the fucose residue was confirmed by the \( C_{FUC} \) (m/z 163.0701, \( z=1 \)) and \( ^{1,5}X_{FUC} \) ions (m/z 833.3229, \( z=2 \)).

The MS/MS spectrum of the acidic A2G2S2 oligosaccharide, Fig. 5-4 (B), showed ions of m/z 958.3197 (\( z=2 \)) and 1059.8643 (\( z=2 \)) with relative high intensity arising from \( ^{0,2}{A} \) cross-ring cleavage of the GlcNAc residues chitobiose core. \( ^{2,4}{A} \) cross-ring fragments (m/z 618.5431, \( z=3 \) and 686.2323, \( z=3 \)) of the chitobiose core were also observed. Ions present at m/z 979.3205 (\( z=1 \)) and m/z 961.3167 (\( z=1 \)) corresponded to the D and D-18 ion series, previously shown to be characteristic of
the composition of the antenna extending from the α-1-6 linked antennary mannose residue\textsuperscript{21,22}.

Given the separation of the anomers on the PGC packing, some researchers have reduced the oligosaccharides to generate single peak alditols\textsuperscript{36}; therefore, we decided to compare the two approaches. Figs. 5-5 (A) and (B) show MS/MS spectra with a free reducing terminal GlcNAc (A) and the corresponding alditol (B) for the triantennary trigalatosyl glycan, A3G3. As can be seen in the figure, the aldose form is at least 10 fold higher in MS/MS sensitivity. The fragmentation spectrum for the aldose form is also seen to be considerably more informative due to the presence of more intense D- and D-18 ions. In addition, we found the precursor signal to be 4 or more fold higher for the aldose form. These differences in MS and MS/MS signals were found to be generally true for a variety of glycan structures. Furthermore, a much higher collision energy was required for the fragmentation of the alditols due to the increased stability of the [M-H]\textsuperscript{-} pseudomolecular ion of this form. This lower sensitivity in both MS and MS/MS signal in alditol is possible due to increased sample preparation steps (reaction and clean up) for generation of alditol and the potential lower response factor. Indeed, a recent study has recommended the use of released \textit{N}-glycans with the native reducing aldose form because of its greater simplicity\textsuperscript{43}. As a consequence of the above, we have used the aldose form in our studies, in spite of the anomeric peak splitting.
Figure 5-4: Extracted ion chromatogram and negative CID-MS/MS spectra of (A) FA2G2 and (B) A2G2S2. Monosaccharide symbol key - blue square: GlcNAc, green circle: mannose, red triangle: fucose, yellow circle: galactose, pink diamond: N-acetylneuraminic acid (Neu5Ac), white diamond: N-glycolylnuraminic acid (Neu5Gc). The buffer contained 10 mM ammonium fluoride. Additional details in the Experimental Section.
Figure 5-5: Negative CID-MS/MS spectra of (A) the N-glycan with a free reducing terminal GlcNAc residue and (B) the corresponding alditol for the triantennary trigalactosyl N-glycan (A3G3). Monosaccharide symbols as described in Figure 5-4. Further experimental details in Figure 5-4.

5.4.3 Human and Murine Polyclonal Serum IgG N-glycan Analysis Using Fluoride-Mediated Negative Ion LC-MS/MS

To demonstrate the power of the methodology, we applied our fluoride-mediated LC-MS platform for the characterization of the N-linked oligosaccharides present on polyclonal IgG extracted from human and murine serum. N-glycans present at asparagine 297 in the C_H2 domain of an antibody play a crucial role in stabilizing the structure of the C_H2 domain and also in modulating the interaction of
the Fc region of the antibody with Igγ receptors present on the surface of cells of the innate immune system\textsuperscript{44-46}. The ability to characterize these N-glycans in detail is important both from the regulatory point of view when considering monoclonal antibody therapeutics \textsuperscript{3} and also as these N-glycans have been shown to be altered in disease \textsuperscript{47}. Following Protein G enrichment, N-glycans from polyclonal human and murine serum IgGs were enzymatically liberated and analyzed using the fluoride-mediated negative ionization microchip LC-MS/MS platform. Previously, we performed an exhaustive analysis of both samples using ultra high-performance hydrophilic interaction liquid chromatography and capillary electrophoresis with laser induced fluorescence detection, with exoglycosidase digestion for structural annotation\textsuperscript{31,48}. Table 5-1 lists N-glycans, including low abundance afucosylated glycans, detected on polyclonal human serum IgG using fluoride-mediated accurate mass Q-TOF LC-MS/MS using the information rich glycosidic and cross ring fragments. Shown in Figure 5-6 (A) is the separation of F(6)A2G[6]1 and F(6)A2G[3]1, more commonly referred to as G1 and G1’, respectively, demonstrating the ability to resolve these positional isomers on the short 43 mm PGC column on the microfluidic chip. Figure 5-6 (B) depicts an annotated spectrum for F(6)A2G[6]1 wherein the position of the galactose residue is confirmed by the presence of the D and D-18 ion couple. The significance of these two ions, which reveal the composition of the antenna extending from the α-1-6 branching mannose, was previously discussed\textsuperscript{21,22}. In addition, linkage specificity could also be determined
from the fluoride-mediated negative ion MS/MS spectra, as shown in Figure 5-6 (C). Here, the oligosaccharide structure was annotated as F(6)A2G[3]1S(6)1, based on the presence of the D and D-18 ion couple, the C$_{2\alpha}$ glycosidic fragment and the $^{0,4}$A$_{2\alpha}$ - CO$_2$, $^{2,5}$A$_{3\alpha}$ and $^{0,3}$A$_{5}$ cross ring fragments$^{22,49}$. The annotation of this structure is in agreement with that determined in our previous studies$^{31}$, without the need for exoglycosidase digestion.

The analysis of the murine serum polyclonal N-glycans using fluoride-mediated negative ion LC-MS/MS revealed the presence of antigenic epitopes, including galactose-$\alpha$-1-3-galactose (see Table 5-2) in agreement with our previous report$^{48}$ and also the presence of N-glycolylneuraminic acid (Neu5Gc) in a significantly higher quantity than N-acetylneuraminic acid (Neu5Ac) as the terminal monosaccharide present on the antennary chains of the oligosaccharides. Figure 5-7 (A) depicts an example base peak chromatogram for the separation of the murine IgG N-glycan pool. Both glycan epitopes have an immunogenic potential in humans$^{50-53}$. Figure 5-7 depicts MS/MS spectra for two oligosaccharides containing galactose-$\alpha$-1-3-galactose epitopes (B) and Neu5Gc residues (C), as determined in a position specific manner, based on the presence of D and D-18 ions and A ion cross ring fragments. The annotated structures are in agreement with those previously annotated using capillary electrophoresis with laser induced fluorescence detection and exoglycosidase digestion$^{48}$. Importantly, the N-glycan structure, F(6)A2G2Gal[6]1S(Neu5Gc)[3]1, as annotated from the MS/MS spectrum in Figure
5-7 (C), was determined to contain both α-1-3-galactose and Neu5Gc epitopes. The position of attachment was also revealed. Based on the MS/MS data, the α-1-3-galactose epitope was determined to be attached to the galactose residue on the antenna extending from the α-1-6-core mannose, whereas the Neu5Gc residue was found to be on the galactose residue on the antenna extending from the α-1-3-core mannose. This result suggests that the presence of acidic amino acids in the primary protein sequence of the C\textsubscript{H2} domain play an important role in governing the processing of the glycans antenna\textsuperscript{45,47}. Table 5-2 contains details of all oligosaccharide structures annotated in the murine polyclonal IgG N-glycan pool by fluoride-mediated negative ion ESI. The results of the IgG glycan analysis of human and mouse demonstrate the potential of negative ESI glycan characterization including linkage and positional isomer determination.
Table 5-1. N-linked oligosaccharides identified by fluoride-mediated negative ion chip-LC-MS/MS in the human polyclonal IgG oligosaccharide pool.

<table>
<thead>
<tr>
<th>Glycans</th>
<th>Theoretical m/z&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Observed m/z&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Retention time (min)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA2</td>
<td>730.2649</td>
<td>730.2677</td>
<td>17.6, 18.9</td>
</tr>
<tr>
<td>FA2G[3]1</td>
<td></td>
<td>811.2956</td>
<td></td>
</tr>
<tr>
<td>A2G2</td>
<td>819.2888</td>
<td>819.2951</td>
<td>17.2&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>FA2G2</td>
<td>892.3178</td>
<td>892.3225</td>
<td>19.0, 20.2</td>
</tr>
<tr>
<td>FA2BG2</td>
<td>993.8575</td>
<td>993.8601</td>
<td>16.4, 17.0</td>
</tr>
<tr>
<td>FA2G2S1</td>
<td>1037.8655</td>
<td>1037.8697</td>
<td>20.1, 21.2</td>
</tr>
<tr>
<td>FA2BG2S1</td>
<td>1139.4052</td>
<td>1139.4047</td>
<td>18.2, 18.4</td>
</tr>
<tr>
<td>FA2G1S1</td>
<td>956.8391</td>
<td>956.8428</td>
<td>19.6, 20.8</td>
</tr>
<tr>
<td>FA2BG1S1</td>
<td>1058.3787</td>
<td>1058.3720</td>
<td>17.8&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>FA2G2S2</td>
<td>1183.4132</td>
<td>1183.4143</td>
<td>20.8, 21.6</td>
</tr>
<tr>
<td>FA2BG2S2</td>
<td>1284.9529</td>
<td>1284.9534</td>
<td>18.8, 19.2</td>
</tr>
</tbody>
</table>

<sup>a</sup>: The charge in each case was -2; <sup>b</sup>: The two retention times represent the anomers; <sup>c</sup>: Anomer separation was not observed.
Figure 5-6: (A) Extracted ion chromatogram of F(6)A2G[6]1, more commonly known as G1 and F(6)A2G[3]1, more commonly known as G1’. (B) Annotated ESI negative ion CID MS/MS spectrum for F(6)A2G[6]1 at the retention time of 19.6 min. (C) Negative ion CID-MS/MS spectra of F(6)A2G[3]1S(6)1 at the retention time of 20.7 min, indicating diagnostic fragments for comprehensive structural annotation. Monosaccharide symbols as described in Figure 5-4. Further experimental details in Figure 5-4.
Table 5-2. N-linked oligosaccharides identified by fluoride-mediated negative ion chip-LC-MS/MS in the murine polyclonal IgG oligosaccharide pool.

<table>
<thead>
<tr>
<th>Glycans</th>
<th>Theoretical m/z&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Observed m/z&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Retention time (min)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA2G[3]1S(Neu5Gc)</td>
<td>964.8635</td>
<td>964.8633</td>
<td>19.1, 19.6</td>
</tr>
<tr>
<td>FA2G2S[3]1 (Neu5Gc)</td>
<td>1045.8629</td>
<td>1055.8898</td>
<td>23.8&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>FA1G[3]1S1 (Neu5Gc)</td>
<td>863.2968</td>
<td>863.3217</td>
<td>19.7, 21.1</td>
</tr>
<tr>
<td>FA2G2S1</td>
<td>1037.8655</td>
<td>1037.8898</td>
<td>20.1, 21.4</td>
</tr>
<tr>
<td>FA2G2S2 (Neu5Gc)</td>
<td>1199.4081</td>
<td>1199.4412</td>
<td>20.7, 22.0</td>
</tr>
<tr>
<td>FA2G2S2</td>
<td>1183.4132</td>
<td>1183.4475</td>
<td>21.1, 22.2</td>
</tr>
<tr>
<td>FA2</td>
<td>730.2649</td>
<td>730.2851</td>
<td>17.5, 18.8</td>
</tr>
<tr>
<td>FA2G2</td>
<td>892.3178</td>
<td>892.3423</td>
<td>18.9, 20.2</td>
</tr>
<tr>
<td>FA2G2Gal2</td>
<td>1054.3706</td>
<td>1054.3965</td>
<td>22.8&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

a: The charge in each case was -2.

b: The two retention times represent the anomers.

c: Anomer separation was not observed.
Figure 5-7: (A) Base peak chromatogram of murine IgG; annotated ESI negative ion CID-MS/MS spectra of (B) FA2G2Gal[6]1 at the retention time of 21.3 min and (C) F(6)A2G2Gal[6]1S(Neu5Gc)[3]1 at the retention time of 22.4 min. Monosaccharide symbols as described in Figure 5-4. Further experimental details in Figure 5-4.

5.5 Conclusion

We describe the use of fluoride-mediated negative ion LC-MS/MS for the in-depth characterization of N-linked oligosaccharides. Fluoride added to the mobile phase facilitates the generation of oligosaccharide pseudomolecular anions in the negative ion mode due to the high gas phase basicity of the fluoride anion. Ionization proceeds via a two-step process, the fluoride ion initially hydrogen bonds with the proton on C3 of the reducing terminal N-acetyl glucosamine ring, followed by proton
abstraction and neutral loss of HF. The efficiency of proton abstraction by fluoride facilitates high sensitivity analysis due to the dominance of [M-H] ions in the resulting MS spectrum. CID MS/MS analysis of these pseudomolecular oligosaccharide anions resulted in the generation of C-type glycosidic fragments, highly informative A-type cross ring fragment ions and D- and E-type ion reaction fragments, which, when combined, facilitated oligosaccharide compositional and linkage characterization. Furthermore, a comparison of the behavior of reducing sugars and their reduced alditol counterparts revealed significantly higher MS and MS/MS sensitivity for the oligosaccharides with the aldose compared to the alditol. Besides the higher LC-MS/MS sensitivity and thus higher information content of the aldose, omitting the reducing step simplifies the sample preparation. Additionally, use of the aldoses creates the possibility to ultimately use the PNGase F microfluidic chip thereby generating an automated glycan analysis platform.

The presence of fluoride in the mobile phase also ensured recovery of sialylated oligosaccharides from the PGC column. Overall method sensitivity was evaluated and found to be similar to levels attainable when using fluorescence detection of derivatized glycans. The combination of the orthogonal information provided through isomeric separation on the PGC phase, accurate mass Q-TOF MS analysis and information rich MS/MS spectra, all generated in a single analysis, facilitates detailed oligosaccharide structural elucidation. Indeed, the platform is complementary to traditionally used positive ion LC-MS.
The method has been applied to the characterization of polyclonal human and murine serum IgG N-glycans. The combination of PGC separation, with accurate mass fluoride-mediated negative ion Q-TOF-MS and CID-MS/MS analysis, allowed deep characterization of the N-glycans present, including the identification of glycans displaying the potentially immunogenic galactose-α-1-3-galactose epitope and Neu5Gc residues in both a linkage and positionally specific manner. We are currently investigating the application of the platform for the characterization of glycoproteins bearing larger N-glycans with higher degrees of sialylation.

5.6 Appendix

5.6.1 Introduction

Sialic acids are a family of neuraminic acids with a nine-carbon backbone. To date, more than 50 different sialic acids have been identified including the most abundant N-acetyleneuraminic acid (Neu5Ac) and nonhuman N-glycolyneuraminic acid (Neu5Gc)\textsuperscript{54}. Sialic acids play crucial roles in various physiological and pathological processes, such as cancer, anti-inflammatory activity, viral infection and immunogenicity\textsuperscript{55,56}. Furthermore, sialic acids can enhance the half-life for therapeutic proteins by protecting the Gal residues\textsuperscript{57,58}. In erythropoietin (EPO), desialylation is reported to decrease the half-life of EPO about 150 fold\textsuperscript{59}. Sialic acids are usually attached to the Gal residue via α-2,6- or α-2,3- linkage. The linkage of sialic acid to the Gal residue can affect protein binding and function\textsuperscript{60}. For example, α-2, 6-sialylation inhibited the binding of N-glycans to galectin 3, while unsialylation...
or α-2, 3-sialylation demonstrated much higher binding to galectin 3\(^6\). Therefore, detection and characterization of sialylated glycan is important for quality control of therapeutic proteins and discovery of disease associated biomarkers.

Currently, the characterization of sialylated glycans includes chromatographic methods (e.g. HPLC), capillary electrophoresis, lectin affinity and mass spectrometric methods. Mass spectrometry with positive ionization is also commonly used for characterization of oligosaccharides. However, sialylated glycans usually display low sensitivity due to the negative charge of sialic acid. Furthermore, the linkage information is not easily accessible due to the lack of cross-ring fragments in the MS/MS spectra. Monosaccharide migration has also been reported during mass spectrometric analysis using positive ionization\(^{19}\), thus complicating data annotation. Recently, negative ionization has been developed for glycan structure analysis. Compared to positive ionization, negative ionization has several advantages such as high sensitivity for sialylated glycans, more linkage information for glycan structure and less monosaccharide migration. A detailed in this chapter, we have developed a platform for comprehensive glycan structure characterization using a porous graphite carbon chip and negative ionization Q-TOF mass spectrometry. Porous graphite carbon provides high resolving power for separation of anomeric, linkage and positional isomers of oligosaccharides. The structure of each isomer is further annotated by MS/MS spectrum. The negative ionization also provides high sensitivity for the negative charged sialylated glycans. The chip based format has the advantage
of high sensitivity and easy operation and low consumption of sample and solvent. In this work, we utilize the platform of PGC chip and negative ESI Q-TOF to study the complex structures of sialylated N-glycans released from alpha-1-acid glycoprotein.

5.6.2 Experimental

5.6.2.1 N-glycan Release

Alpha-1-acid glycoprotein was purchased from Sigma and dissolved in 200 µL of 20 mM sodium bicarbonate, pH 7.0. The glycans were enzymatically liberated by PNGase F, using an enzyme to protein ratio of 1:10 (v/v) at 37°C overnight. Following incubation, the N-glycans were collected via centrifugation through a 10 kDa molecular weight cut-off filter, reduced to dryness via vacuum centrifugation and treated with 1% v/v formic acid to promote conversion of the reducing terminal glycosylamine to the corresponding reducing sugar.

5.6.2.2 2-Aminobenzamide (2-AB) Labeling

The oligosaccharides were labeled via reductive amination with 2-aminobenzamide (2-AB) and sodium cyanoborohydride in 70% (v/v) dimethyl sulfoxide (DMSO) and 30% acetic acid. The reaction was incubated at 65°C for two hours. Following incubation, 90 µL water was added to quench the reaction, and then 900 µL of acetonitrile was added. The sample was cleaned using PhyNexus normal phase PhyTips (San Jose, CA). Briefly, the tips were conditioned by 95% (v/v) acetonitrile. After loading the sample, the tips were washed by 95% (v/v) acetonitrile,
and then the samples were eluted by water and dried via vacuum centrifugation.

Other experimental details were described in the experimental section in this chapter.

5.6.2.3 Fraction of N-glycan by Anion Exchange Chromatography

Oligosaccharides were fractionated by a Waters UPLC with a Waters 2475 fluorescence detector under the control of Empower Chromatography Workstation. The analytical column was a Prozyme GlycoSep C polymeric Diethylaminoethyl (DEAE) anion exchange column, 75 x 7.5 mm i.d., 10 μm particle size. Mobile phase A consisted of 20% acetonitrile in water and mobile phase B of 0.1 M ammonium acetate buffer (pH 7.0) in 20% acetonitrile and 80% water. The gradient for pre-fraction was 0% B from 0-5 min, 0-100% B from 5-50 min, 100% B from 50-53 min and 100-0% B from 53-54 min with the flow rate of 0.75 mL/min, followed by an isocratic hold for 6 minutes with the flow rate of 1 mL/min to ensure complete column re-equilibration. The wavelengths were λex = 330 nm and λem = 420 nm, and the data collection rate was 20 Hz. The 2-AB labeled oligosaccharides were analyzed to identify the elution time for each fraction, and then the unlabeled oligosaccharides were pre-fractionated based on the retention time.

5.6.3 Results and Discussion

Human alpha-1-acid glycoprotein (AGP) is a glycoprotein with a molecular weight of 41-43 kDa. It is made up of 45% carbohydrate which is attached in the form of five N-linked glycans. This protein has a low pI of 2.8-3.8 due to the highly sialylated glycans. AGP is an acute phase protein, and the concentration change in
plasma and glycosylation change are used as a diagnostic biomarker\textsuperscript{62}. The glycosylation change on AGP is associated with disease, such as cancer, rheumatoid arthritis and hepatitis\textsuperscript{61,63}. However, the carbohydrates on AGP are very complicated being composed of multiple antennae. In addition, the carbohydrates are highly sialylated, roughly 10-12\% neuraminic acid\textsuperscript{61}.

In this work, two-dimensional separation was utilized to separate the glycans. First, anion exchange chromatography was employed to pre-fractionate the sialylated glycans based on their charges. Then, each fraction were further separated by the PGC chip and analyzed by fluoride mediated negative ESI Q-TOF. This method provides high sensitivity for detection of sialylated glycans, along with linkage and positional isomer information for mono- and di-sialylated glycans. However, the glycans with tri- or tetra-sialic acid usually display the instability of loss of sialic acid during MS/MS fragmentation, thus limiting the structural information. The glycans with tri- or tetra- sialic acid need further derivatization before analysis.

5.6.3.1 Pre-fraction of N-glycans Released from AGP by Anion Exchange Chromatogram

As AGP contains numerous N-glycans with complicated structures and multiple sialic acids at the terminus, a two-dimensional separation strategy for pooled N-glycans released from AGP was designed. In the first dimension, anion exchange chromatography was utilized to separate the N-glycans into four fractions based on their charges (the numbers of sialic acids) on each oligosaccharide. The released N-
glycans were labeled via 2-AB for fluorescence detection. A shallow gradient over 60 min was used for separation. Figure 5-8 shows the elution of oligosaccharides with different charges. Then, based on the elution time, each fraction of oligosaccharides without 2-AB labeling was collected for further analysis. In anion exchange, the retention mechanism is the negative charge interaction with the positive charge of DEAE. Thus, the oligosaccharides with more charges will bind stronger and elute later. The peaks in Figure 5-8 could be grouped into four sections: 11-16 min, 18-22 min, 28-32 min and 37-42 min, assuming 1, 2, 3 and 4 negative charges. The multiple peaks in each time window are due to the different glycan sizes with the same net charges. It is reported that larger glycans with the same net charge usually elute slightly earlier than the smaller glycans.

**Figure 5-8.** Anion exchange chromatogram for glycan fraction.

### 5.6.3.2 Detection of Sialylated N-glycans from AGP Based on the Accurate Mass Using PGC-nESI-Q-TOF
After the pre-fraction of N-glycans of AGP, each fraction was analyzed by the PGC chip and fluoride mediated negative ESI Q-TOF detection. This method utilizes negative ionization in mass spectrometry; thus it has the advantage of high sensitivity for sialylated glycans analysis. The low fmol detection limit was reported in this chapter. Q-TOF mass spectrometer with the high resolution and high mass accuracy can provide the composition of N-glycans. Twenty nine sialylated N-glycans with different composition were detected based on their accurate mass. Table 5-3 lists the N-glycans detected in each fraction.

5.6.3.3 Linkage Identification of N-glycans with One or Two Sialic Acids

As mentioned in this chapter, the PGC chip provides the advantage of separation of linkage and positional isomers. Collision induced dissociation with negative ionization mode can generate structure informative cross ring fragment ions, glycosidic fragments and linkage diagnostic ions such as D and D-18 ions in the MS/MS spectrum. The complete structure assignment of individual oligosaccharides can be determined by combining these fragments.

The oligosaccharide with the composition of A2G2S1 has an m/z of 964.8635 with two negative charges. Eight peaks were extracted from the LC chromatogram based on the accurate mass. All of these peaks showed the same MS spectra (same m/z and same charges) and similar MS/MS spectra. The extracted LC chromatograms are shown in Figure 5-9 (A), with the eight peaks being eluted between 17.5 min to
25.5 min. The MS spectra for these eight peaks had the same precursor m/z and the same charges. For all these eight peaks, the ion at 964.8263 (2-) was the deprotonated form while the ion at 974.8229 (2-) was the adduct form with one fluoride, as shown in Figure 5-9 (B). The MS/MS spectra were very similar to each other. For example, the ions at 884.2875 \( (^{2,4}A_7) \) and 782.7484 \( (^{2,4}A_6) \) were identified in all of the spectra. However, slight differences were also observed for these eight MS/MS spectra.

Figure 5-9 (C) shows the annotation of peak # 3. The ions of 979.3037 (1-) and 961.2865 (1-) represented the D and D-18 ions, which were diagnostic ions for the 6-arm antenna, suggesting that the sialic acid was on that 6-arm antenna. Furthermore, the ions of C\(_{2a}\) and O\(_{2a}\) suggest that the sialic acid is attached to the 6 position of the gal residue, based on reported studies \(^49\). Thus, the structure of peak # 3 including the sialic acid and linkage was assigned as shown in Figure 5-9. The MS/MS spectrum of peak # 6 is the same as that of peak # 3; thus, peak #3 and peak #6 are assumed to be anomeric isomers which are separated on the PGC chip. Similarly, we can assign all the structures of these eight peaks as positional, linkage and anomeric isomers, as shown in Figure 5-9 (D).
Table 5-3. N-linked oligosaccharides identified in α-1-acid glycoprotein by fluoride-mediated negative ion chip-LC-MS/MS.

<table>
<thead>
<tr>
<th>Fractions</th>
<th>Glycans</th>
<th>Theoretical m/z and charges</th>
<th>Observed m/z and charges</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fraction #1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A2G2S1</td>
<td>964.8635, 2-</td>
<td>964.8236, 2-</td>
</tr>
<tr>
<td></td>
<td>A2G2S1F</td>
<td>1037.8655, 2-</td>
<td>1037.8559, 2-</td>
</tr>
<tr>
<td></td>
<td>A1G1S1F</td>
<td>855.2994, 2-</td>
<td>855.2964, 2-</td>
</tr>
<tr>
<td></td>
<td>A1G1S1</td>
<td>782.2704, 2-</td>
<td>782.2599, 2-</td>
</tr>
<tr>
<td></td>
<td>A3G3S1</td>
<td>1147.4026, 2-</td>
<td>1147.3844, 2-</td>
</tr>
<tr>
<td></td>
<td>A3G3S1F</td>
<td>1220.4316, 2-</td>
<td>1220.4193, 2-</td>
</tr>
<tr>
<td></td>
<td>A4G4S1</td>
<td>1329.9687, 2-</td>
<td>1329.9466, 2-</td>
</tr>
<tr>
<td></td>
<td>A4G4S1F</td>
<td>1402.9977, 2-</td>
<td>1402.9841, 2-</td>
</tr>
<tr>
<td><strong>Fraction #2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A2G2S2F</td>
<td>1183.4132, 2-</td>
<td>1183.4187, 2-</td>
</tr>
<tr>
<td></td>
<td>A2G2S2</td>
<td>1110.3842, 2-</td>
<td>1110.3900, 2-</td>
</tr>
<tr>
<td></td>
<td>A3G3S2</td>
<td>861.6311, 3-</td>
<td>861.6373, 3-</td>
</tr>
<tr>
<td></td>
<td>A3G3S2F</td>
<td>910.3171, 3-</td>
<td>910.3240, 3-</td>
</tr>
<tr>
<td></td>
<td>A3G3S2F2</td>
<td>1439.0082, 2-</td>
<td>1439.0038, 2-</td>
</tr>
<tr>
<td></td>
<td>A4G4S2</td>
<td>983.3419, 3-</td>
<td>983.3467, 3-</td>
</tr>
<tr>
<td></td>
<td>A4G4S2F</td>
<td>1032.0278, 3-</td>
<td>1032.0304, 3-</td>
</tr>
<tr>
<td></td>
<td>A4Lac1G4S2</td>
<td>1105.0526, 3-</td>
<td>1105.0551, 3-</td>
</tr>
<tr>
<td></td>
<td>A4Lac1G4S2F</td>
<td>1153.7386, 3-</td>
<td>1153.7447, 3-</td>
</tr>
<tr>
<td><strong>Fraction #3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A3G3S3</td>
<td>958.6629, 3-</td>
<td>958.6691, 3-</td>
</tr>
<tr>
<td></td>
<td>A3G3S3F</td>
<td>1007.3489, 3-</td>
<td>1007.3545, 3-</td>
</tr>
<tr>
<td></td>
<td>A4G4S3</td>
<td>1080.3737, 3-</td>
<td>1080.3786, 3-</td>
</tr>
<tr>
<td></td>
<td>A4G4S3F</td>
<td>1129.0596, 3-</td>
<td>1129.063, 3-</td>
</tr>
<tr>
<td></td>
<td>A4G4S3F2</td>
<td>1177.7456, 3-</td>
<td>1177.7463, 3-</td>
</tr>
<tr>
<td></td>
<td>A4Lac1G4S3</td>
<td>1202.0844, 3-</td>
<td>1202.087, 3-</td>
</tr>
<tr>
<td></td>
<td>A4Lac1G4S3F</td>
<td>1250.7704, 3-</td>
<td>1250.7500, 3-</td>
</tr>
<tr>
<td></td>
<td>A4Lac2G4S3</td>
<td>1323.7951, 3-</td>
<td>1323.7928, 3-</td>
</tr>
<tr>
<td><strong>Fraction #4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A4G4S4</td>
<td>1177.4055, 3-</td>
<td>1177.4078, 3-</td>
</tr>
<tr>
<td></td>
<td>A4G4S4F</td>
<td>1226.0914, 3-</td>
<td>1226.0972, 3-</td>
</tr>
<tr>
<td></td>
<td>A4G4S4F2</td>
<td>1274.7774, 3-</td>
<td>1274.7801, 3-</td>
</tr>
<tr>
<td></td>
<td>A4Lac1G4S4</td>
<td>1299.1162, 3-</td>
<td>1299.1198, 3-</td>
</tr>
</tbody>
</table>
**Figure 5-9.** Extracted ion chromatogram and negative CID spectra of A2G2S1; (A) LC-MS elution profile, (B) precursor mass of doubly charged A2G2S1 in negative ion Q-TOF MS analysis, (C) annotated negative ion CID-MS/MS spectrum of Peak 3. (D) peak annotation of A2G2S1. Monosaccharide symbols as described in Figure 5-4. Further experimental details in Figure 5-4.

5.6.3.4 Future Work

Glycans with three or more sialic acids usually generated major peaks of sialic acid, and the precursor ions tend to lose some of these sialic acids (data not shown here) in the MS/MS spectrum due to the negative charge repulsion. Reduction of charges by derivatization of sialic acid, such as methylation, can be performed before LC-MS/MS analysis to facilitate the glycan structure characterization by MS/MS.
5.7 References


(23) 2005.


(28) Harvey, D. J.; Royle, L.; Radcliffe, C. M.; Rudd, P. M.; Dwek, R. A. *Anal Biochem* **2008**, *376*, 44.


*Proteome Res* **2008**, *7*, 1650.


(49) Wheeler, S. F.; Harvey, D. J. *Analytical chemistry* 2000, 72, 5027.


Chapter 6: Conclusion and Future Directions
6.1 Conclusion

In this dissertation, multiple liquid chromatography-mass spectrometry strategies have been developed for in-depth characterization and quantification of protein post-translational modifications.

Isoaspartic acid, as one kind of protein damage arising from either deamidation of asparagine or isomerization of aspartic acid, is associated with a variety of biological processes, such as aging and cancer\(^1\)\(^-\)\(^3\). Accumulation of isoaspartic acid has been implicated in protein aggregation\(^4\), loss of activity\(^1\), and altered immunogenicity\(^5\). Identification of isoaspartic acid is important for understanding the mechanism of its formation and role in protein activity and diseases. However, due to the subtle difference between isoaspartic acid and aspartic acid, it remains a critical challenge to be able to differentiate them. In this dissertation, we have developed two approaches for the characterization and quantitation of isoaspartic acid. In the first approach, we utilized the specific cleavage capability of Asp-N at the N-terminus of aspartic acid but not isoaspartic acid to differentiate and enrich isoaspartic acid. Then, the specific site of isoaspartic acid was determined by ETD mass spectrometry with the diagnostic ions of isoaspartic acid (\(c+57, z-57\)). Furthermore, these diagnostic ions of isoaspartic acid can also be used for the absolute and relative quantitation of isoaspartic acid. In the second approach, we have performed proteomic screening of isoaspartic acid by
bottom-up methods. Then, a targeted ETD method has been developed for verification and quantitation of isoaspartic acid.

Disulfide linkages are well known to play a significant role in protein folding, stability and activity\(^6\). However, characterization of intertwined disulfide linkages is still challenging, despite the advances in high resolution LC-MS/MS. Here, we have developed a method based on multi-enzyme digestion in combination with LC-MS with ETD and sequential CID analysis to characterize complex disulfide structures including a cystine knot and nested disulfides.

Protein glycosylation contributes significantly to physical, chemical and biological properties of proteins in many biological processes via regulating protein folding, stability, activity, cellular location and interactions of proteins with other biomolecules\(^7\). In addition, aberrant glycosylation is associated with diseases such as cancer and inflammation\(^8,9\). Furthermore, glycosylation may affect pharmacodynamics and pharmacokinetics of therapeutic proteins by modulating the activity, circulating half-life of protein drugs as well as cause immunogenicity\(^10,11\). Here, we have developed fluoride mediated negative ionization chip based PGC-Q-TOF-MS/MS methods for the comprehensive characterization of oligosaccharides. In this approach, the PGC packing can facilitate positional and linkage isomer separation for glycans. The Q-TOF mass spectrometer with the high resolution and high mass accuracy can ease and increase the confidence of data annotation. The negative
ionization can provide the complex structure information by formation of glycosidic and cross-ring fragments.

6.2 Future Directions

6.2.1 Isoaspartic Acid

6.2.1.1 Determination of IsoAsp in Recombinant Proteins

As deamidation of asparagine or isomerization of aspartic acid is a non-enzymatic reaction, it can occur in therapeutic proteins at any stage, e.g., purification, formulation, storage. Our two developed approaches for detection and quantitation of isoaspartic acid, which are described in Chapters 2 and 3, can be useful for monitoring isoAsp formation in therapeutic proteins during production and storage, thus ensuring the quality of the drug.

6.2.1.2 Identification of IsoAsp in Clinical Samples

In Chapter 3, we developed targeted ETD in combination with shotgun proteomic methods to identify and quantitate isoAsp in a complex biological system: mouse urine. We can apply this approach to analyze clinical samples such as plasma, serum, urine and tissue to identify potential isoAsp sites to understand the isoAsp processing pathways in vivo and also to identify potential disease-associated biomarkers.

6.2.2 Disulfide Linkages
6.2.2.1 Quality Control for RhASA

We have successfully characterized the cysteine status in rhASA including free and modified cysteines, and disulfide linkages with complex structures such as nested disulfides and cystine knots using our developed method with multi-enzyme digestion and LC-MS with CID and ETD. We also have applied this method to determine the disulfide scrambling in stressed rhASA, which has provided stability information of rhASA. In the future, this developed method can be applied for quality control of manufactured products to ensure function and stability of rhASA.

6.2.2.2 Characterization of a Cystine Knot

Cystine knots are ultra-stable motifs for protein structure and have been found in a large number of proteins. Cyclotide, also called cyclic cystine knot or cystine mini-protein, shares a head-to-tail knot structure which has been developed for scaffolds for drug design or clinical usage due to its high stability and ability to cross the cell membrane\textsuperscript{12}. The ultra-stability of cyclotide is due to three knotted disulfide linkages. Thus, characterization of disulfide linkages in a cystine knot is critical for development and quality control for cyclotide-based drugs or diagnostic reagents. Our developed method in Chapter 4 provides a roadmap for characterization of the difficult-to-resolve disulfide linkages in a cystine knot.

6.2.2.3 Determination of Disulfide Linkage in Clinical Samples

Disulfide linkage plays an important role for protein stability and function. Aberrant disulfide linkages are also associated with disease. For example, the
mutation of Cys 470 to Arg in rhASA, and thus disruption of a disulfide linkage, has been reported in patients with metachromatic leukodystrophy\textsuperscript{13}. Thus, characterization of disulfide linkages in clinical samples can provide insight for understanding the mechanism for diseases. Our method can also be applied for determination of the disulfide linkages in clinical samples.

6.2.3 Glycosylation

6.2.3.1 Characterization of Anionic Glycans: Sulfated and Phosphorylated Glycans

Oligosaccharides bearing inorganic substituents, such as sulfate or phosphate play an important role in biological systems\textsuperscript{14,15}. However, characterization of these modified glycans usually presents a big challenge as these glycans usually display low sensitivity in positive ionization mass spectrometric analysis and require multiple methods of enrichment before LC-MS analysis. Furthermore, the migration of sulfate in mass spectrometric analysis has been reported\textsuperscript{16}, thus complicating data annotation. However, these anionic glycans usually show high sensitivity in negative ESI. Furthermore, the linkage information can also be achieved from the cross-ring fragments which are generated in MS/MS. Thus, our developed method in Chapter 5 can be applied for comprehensive characterization of these anionic glycans.

6.2.3.2 Determination of Sialylated Glycans
Sialic acid has been shown to increase the half-life of therapeutic
glycoproteins, such as EPO and asparaginase\(^{10}\). Thus, efforts have been made to
improve drug quality by glyco-engineering. Our method is well suited for analysis of
sialylated glycans due to its high sensitivity for anionic glycans.

Altered sialylated glycans are associated with cancers\(^8\). Sialyltransferase is
one family of enzymes which are responsible for transfer of sialic acid to an acceptor
carbohydrate. The expression level of sialyltransferases has been observed with
dramatic change during oncogenic transformation\(^{17}\). Thus, we hypothesize that the
level of sialylated glycans could be altered in these cancer tissues. Our developed
PGC negative ionization Q-TOF platform can be used for analysis of the sialylated
glycan level in disease samples and identification of possible biomarkers.

6.2.3.3 Characterization of O-linked Glycans

O-glycosylation plays a significant role in protein structure and stability, and it
is involved in many biological processes. Furthermore, aberrant O-glycosylation is
also associated with many diseases such as familial tumoral calcinosis, Tn syndrome
(a rare autoimmune disorder) and IgA nephropathy (IgAN)\(^{18}\). Thus, characterization
of O-glycosylation is important for understanding its biological functions. However,
alyses of O-glycosylation is still challenging due to its heterogeneity. Moreover,
there are no specific enzymes for O-glycan release. Currently, the most commonly
used method for O-glycan release is β-elimination. After β-elimination, the released
O-glycans have the reduced end, which is also called alditol, while N-glycans after
PNGase F treatment, have the reducing end. We have applied our method to analyze the glycan alditol, and it has been demonstrated to generate glycosidic and cross-ring fragments in MS/MS. This result suggests that our method should also be applicable for comprehensive characterization of O-glycans.

6.2.3.4 Automation by Chip Format

The automation of analysis can be achieved by the format of lab-on-a-chip. Recently, Agilent has developed a new generation of PGC chip with an added enzyme reactor (PNGase F)\textsuperscript{19}. This PNGase F reactor can complete the online digestion in a few minutes, and then the released glycan is separated by the PGC chip. We can also use this chip to analyze glycans without offline digestion. This method is fast, easy and reproducible. However, the glycan are mainly in the amine form at the reducing end after the online digestion. It usually needs about 1 to 2 hours to convert the amine form to hydroxyl form. Once the glycan with the amine end can be converted to hydroxyl form very fast (about minutes level), the PGC with PNGase F reactor will be very useful as this method is fast, simple, reproducible, and does not require sample preparation.

In Chapter 5, we have reported 2D-LC with negative ESI Q-TOF analysis of glycans. In this method, the glycans have been offline pre-fractionated by weak anion exchange chromatography. Then, each fraction has been further resolved by the PGC chip. In the future, online 2D separation can be achieved by incorporating an anion
exchange column before the PGC column on chip. This new design 2D LC chip will be of simple operation, and high reproducibility.

6.2.3.5 Characterization of Glycans on Different Mass Spectrometers.

In Chapter 5, we have compared different mass analyzers: a linear ion trap (LTQ-XL, Thermo) and Q-TOF (Agilent), for glycan analysis. Our results show that both LTQ and Q-TOF generate A-type cross-ring fragments and C-type glycosidic ions. However, the intensity of the cross-ring fragments in the LTQ is considerably lower than that in the Q-TOF, which is attributed to the more energetic fragmentation processes employed in the Q-TOF compared to the ion trap. Thus, instruments employing higher-energy collision dissociation (HCD) will generate more cross-ring fragments due to the higher collision energy. This study will broaden the application of our method in different instruments, such as Orbitrap and Q-Exactive.

6.3. References


(12) Craik, D. J.; Cemazar, M.; Daly, N. L. Curr Opin Drug Discov Devel 2006, 9, 251.


(15) Ohtsubo, K.; Marth, J. D. Cell 2006, 126, 855.


Appendix

Publications


Copyright from American Chemical Society (ACS)

American Chemical Society’s Policy on Theses and Dissertations

If your university requires you to obtain permission, you must use the RightsLink permission system. See RightsLink instructions at http://pubs.acs.org/page/copyright/permissions.html.

This is regarding request for permission to include your paper(s) or portions of text from your paper(s) in your thesis. Permission is now automatically granted; please pay special attention to the implications paragraph below. The Copyright Subcommittee of the Joint Board/Council Committees on Publications approved the following:

Copyright permission for published and submitted material from theses and dissertations

ACS extends blanket permission to students to include in their theses and dissertations their own articles, or portions thereof, that have been published in ACS journals or submitted to ACS journals for publication, provided that the ACS copyright credit line is noted on the appropriate page(s).

Publishing implications of electronic publication of theses and dissertation material

Students and their mentors should be aware that posting of theses and dissertation material on the Web prior to submission of material from that thesis or dissertation to an ACS journal may affect publication in that journal. Whether Web posting is considered prior publication may be evaluated on a case-by-case basis by the journal’s editor. If an ACS journal editor considers Web posting to be “prior publication”, the paper will not be accepted for publication in that journal. If you intend to submit your unpublished paper to ACS for publication, check with the appropriate editor prior to posting your manuscript electronically.

Reuse/Republication of the Entire Work in Theses or Collections: Authors may reuse all or part of the Submitted, Accepted or Published Work in a thesis or dissertation that the author writes and is required to submit to satisfy the criteria of degree-granting institutions. Such reuse is permitted subject to the ACS’ “Ethical Guidelines to Publication of Chemical Research” (http://pubs.acs.org/page/policy/ethics/index.html); the author should secure written confirmation (via letter or email) from the respective ACS journal editor(s) to avoid potential conflicts with journal prior publication*/embargo policies. Appropriate citation of the Published Work must be made. If the thesis or dissertation to be published is in electronic format, a direct link to the Published Work must also be included using the ACS Articles on Request author-directed link – see http://pubs.acs.org/page/policy/articlesonrequest/index.html * Prior publication policies of ACS journals are posted on the ACS website at http://pubs.acs.org/page/policy/prior/index.html

If your paper has not yet been published by ACS, please print the following credit line on the first page of your article: "Reproduced (or 'Reproduced in part') with permission from [JOURNAL NAME], in press (or 'submitted for publication'). Unpublished work copyright [CURRENT YEAR] American Chemical Society." Include appropriate information.

If your paper has already been published by ACS and you want to include the text or portions of the text in your thesis/dissertation, please print the ACS copyright credit line on the first page of your article: “Reproduced (or 'Reproduced in part') with permission from [FULL REFERENCE CITATION.] Copyright [YEAR] American Chemical Society.” Include appropriate information.
**Submission to a Dissertation Distributor:** If you plan to submit your thesis to UMI or to another dissertation distributor, you should not include the unpublished ACS paper in your thesis if the thesis will be disseminated electronically, until ACS has published your paper. After publication of the paper by ACS, you may release the entire thesis (*not the individual ACS article by itself*) for electronic dissemination through the distributor; ACS’s copyright credit line should be printed on the first page of the ACS paper.
This is a License Agreement between Wenqin Ni ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

<table>
<thead>
<tr>
<th>License Number</th>
<th>3065960538982</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Jan 11, 2013</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Springer</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Journal of The American Society for Mass Spectrometry</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Complete Mapping of a Cystine Knot and Nested Disulfides of Recombinant Human Arylsulfatase A by Multi-Enzyme Digestion and LC-MS Analysis Using CID and ETD</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Wenqin Ni</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Jan 1, 2012</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Portion</td>
<td>Full text</td>
</tr>
<tr>
<td>Number of copies</td>
<td>1</td>
</tr>
<tr>
<td>Author of this Springer article</td>
<td>Yes and you are a contributor of the new work</td>
</tr>
<tr>
<td>Order reference number</td>
<td></td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Advances in Protein Posttranslational Modifications (PTMs) Using Liquid Chromatography-Mass Spectrometry</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Feb 2013</td>
</tr>
<tr>
<td>Estimated size(pages)</td>
<td>200</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Terms and Conditions</td>
<td></td>
</tr>
</tbody>
</table>