Glycan characterizations with application in metal ion binding, the 1918 influenza virus neuraminidase formation and human receptors

Hui Zhou

University of New Hampshire, Durham

Follow this and additional works at: https://scholars.unh.edu/dissertation

Recommended Citation
Zhou, Hui, "Glycan characterizations with application in metal ion binding, the 1918 influenza virus neuraminidase formation and human receptors" (2009). Doctoral Dissertations. 510.
https://scholars.unh.edu/dissertation/510

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.
GLYCAN CHARACTERIZATIONS WITH APPLICATION IN METAL ION BINDING, THE 1918 INFLUENZA VIRUS NEURAMINIDASE FORMATION AND HUMAN RECEPTORS

BY

HUI ZHOU

Baccalaureate degree, Nanchang University, 1998
Master of Science Degree, Shanghai Institute of Material Medica, Chinese Academy of Science, 2002

DISSERTATION

Submitted to the University of New Hampshire
In Partial Fulfillment of
The Requirements for the Degree of

Doctor of Philosophy
in
Chemistry

September, 2009
This dissertation has been examined and approved.

Dissertation Director, Vern Reinhold, Professor Chemistry

Roy Planalp, Professor Chemistry

Charles Zercher, Professor Chemistry

Ed Wong, Professor Chemistry

Deena Small, Professor Biochemistry

Date
DEDICATION

To my beautiful wife, Lanfang Mei
ACKNOWLEDGEMENTS

I would like to offer my sincerest thanks and gratitude to my advisor Dr. Vernon N. Reinhold for his support and guidance. He has been a great mentor to me, and I will always admire his scientific passion, and professionalism. I would also like to thank all my colleagues working in the UNH Glycomics Center: Andy Hanneman, David Ashline, Hailong Zhang, Cristina Silvescu, Justin Prien, Yunsen Li, and Stephanie Maniatis for their helps and friendship. Special thanks go out to my colleague Chaohong Liu for his collaboration on influenza human receptor researches and Sherry Castle for her help on GC/MS experiments.

I would like to thank my all committee members Dr. Roy Planalp, Dr. Charles Zercher, Dr. Deena Small, Dr. Ed Wong and Dr. Steve Levery for all their encouragement and help throughout my time at UNH. And special thanks go to Dr. Zercher for his kind support and guidance when I was working in his lab.

I also want to thank Dr. Dennis Chasteen for his support on the EP glycoprotein project, and Dr. Leon Wu in R&D system for his discussion and collaboration on the 1918 influenza virus neuraminidase project.

I would like to thanks UNH Graduate School for providing one-year dissertation-fellowship (2008), and NIH grant for financial support.

Finally, special indebtedness extends to my parents, Yuguo Zhou and Changjun Cao, and my parents-in-law, Shiyun Mei and Meizhu Ma for their encouragement and love. I am particularly grateful to my wife, Lanfang Mei, for her faith in me and for his love and support.
TABLE OF CONTENTS

DEDICATION ........................................................................................................... iii
ACKNOWLEDGEMENTS ......................................................................................... iv
LIST OF TABLES ...................................................................................................... vii
LIST OF FIGURES ................................................................................................... viii
ABSTRACT ............................................................................................................... xi

CHAPTER PAGE

1. Introduction to Glycosylation ........................................................................... 1
   1.1 Significance of Glycosylation ................................................................. 1
   1.2 Structure Features of N- and O-Glycans ............................................... 4
   1.3 Biosynthesis of N-linked Glycans .......................................................... 8
   1.4 Glycome and Structural Analysis of Glycan by MS\textsuperscript{n} .......... 12

2. Characterization of Novel N-Glycans of the EP Glycoprotein from the Mollusc:
   \textit{Mytilus edulis} ......................................................................................... 17
   2.1 Introduction ............................................................................................. 17
   2.2 Experimental Procedures ....................................................................... 23
   2.3 Results ..................................................................................................... 27
   2.4 Discussion ............................................................................................... 47

3. \textit{N}-glycosylation Analysis of Neuraminidase from the 1918 Pandemic Influenza
   Virus H1N1 .................................................................................................. 54
   3.1 Introduction to Influenza Virus ................................................................. 54
   3.2 Research Background and Purposes ....................................................... 66
LIST OF TABLES

Table 2.1. GC/MS monosaccharide composition analyses of EP glycans........32

Table 3.1. Natural hosts of influenza A viruses........................................57

Table 3.2. The relative ratio of each composition of both dimer and tetramer NA obtained from MALDI-MS..............................................................78

Table 3.3. The characterized N-glycan structures from dimer and tetramer NA by MS^n analysis.................................................................91-94

Table 4.1. Measuring Sialα(2,3)Gal and Sialα(2,6)Gal residues relative ratio by DMTMM method from acidic N-glycans of bovine fetuin.................130

Table 4.2. Measuring Sialα(2,3)Gal and Sialα(2,6)Gal residues relative ratio by DMTMM method from acidic N-glycans of Detroit 562...............130
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Nine common monosaccharides found in vertebrates</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Common types of N-glycans and its core structure</td>
<td>6</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Mucin-type O-glycan core structures</td>
<td>7</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Synthesis of the precursor LLO in the ER</td>
<td>9</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>The N-glycosylation pathway: trimming and elongation in ER and Golgi</td>
<td>11</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Thirteen isomeric N-glycans from bovine ribonuclease B</td>
<td>15</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Negative ion ESI-MS of N-glycans of EP glycoprotein</td>
<td>30</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Accurate mass measurement of glycan 1 of EP glycoprotein</td>
<td>31</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Positive ion ESI MS$^2$ of m/z 1706.8 of deuterated permethylated glycan 1</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>MS$^3$ of deuterated permethylated antenna B ion of glycan 1</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>MS$^n$ spectra of m/z 789.5 and m/z 665.4 from deuterated permethylated antenna B ion of glycan 1</td>
<td>37</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>MS$^5$ spectra of m/z 427.2 from deuterated permethylated antenna B ion of glycan 1</td>
<td>38</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>MS$^3$ spectra of m/z 907.5 from permethylated antenna B ion of glycan 1 after conversion GlcA to Glc</td>
<td>38</td>
</tr>
<tr>
<td>Figure 2.8</td>
<td>(-) MS$^2$ spectra of m/z 1005.9 of the native glycan 1</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2.9</td>
<td>(-) MS$^2$ spectra of m/z 1006.5 of the native and reduced glycan 1</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2.10</td>
<td>Diagnostic E ion of Man-3 branching of glycan 1</td>
<td>42</td>
</tr>
<tr>
<td>Figure 2.11</td>
<td>Characterized structures of glycan 1-3</td>
<td>43</td>
</tr>
<tr>
<td>Figure 2.12</td>
<td>MS$^2$ spectra of deuterated permethylated and permethylated glycan 3</td>
<td>44</td>
</tr>
<tr>
<td>Figure 2.13</td>
<td>MS$^3$ spectra of m/z 1129.8 from deuterated permethylated and</td>
<td></td>
</tr>
</tbody>
</table>
1093.5 from permethylated antenna B ion of glycan 3..............44

Figure 2.14. MS^4 spectra of m/z 936.5 and 905.5 from permethylated antenna B ion of glycan 3..................45

Figure 3.1. Typical influenza virus structural model.....................................................58
Figure 3.2. The 1918 influenza virus morbidity and mortality rate by age.............66
Figure 3.3. The complete amino acid sequence of 1918 virus NA.........................69
Figure 3.4. Chromatography of the 1918 NA in monomer dimer and tetramer................71
Figure 3.5. MALDI-MS profiles of native N-glycans from tetramer and dimer .....................79
Figure 3.6. MS^n analysis of Man_3GlcNAc_2 (m/z 1187)...........................................84
Figure 3.7. MS^n analysis of FucMan_3GlcNAc_2 (m/z 1361)......................................85
Figure 3.8. MS^n analysis of isomeric structure of FucMan_3GlcNAc_2 (m/z 1361)...............86
Figure 3.9. MS^n analysis of FucMan_2GlcNAc_2 (m/z 1157).....................................89
Figure 3.10. MS^n analysis of isomeric structure of FucMan_2GlcNAc_2 (m/z 1157).............90
Figure 3.11. MALDI-MS profiles of native N-glycans from the stalk region of tetramer and dimer ........................................95
Figure 3.12. Fucose was added into dimer and tetramer by Fut8..............................96
Figure 3.13. Tetramer enzyme activity after Fut8 treatment.................................96
Figure 4.1. The anatomical distribution of human respiratory tract.......................104
Figure 4.2. Sial-Gal detection of human respiratory tract by lectins......................105
Figure 4.3. Determination of Sial-Gal linkage of IgG Fc by MS^n spectral comparison......112
Figure 4.4. MALDI-TOF spectra of N-glycans of bovine fetuin after conversion.............113
Figure 4.5. Two standard sialylated trisaccharides
reacted with DMTMM ........................................118

Figure 4.6. MALDI-TOF spectra of N-glycans of bovine fetuin
after DMTMM modification........................................120

Figure 4.7. MALDI-TOF spectra of N-glycans of both A549 and
Detroit 562 cell lines after permethylation........................122

Figure 4.8. MALDI-TOF spectra of N-glycans of Detroit 562 cell lines
after DMTMM modification........................................124

Figure 4.9. Determination of Sial-Gal linkage of A549 cell lines
by MS^n spectral comparison....................................125

Figure A.1. Negative ion MS^2 and MS^3 of glycan 1............150

Figure A.2. GC retention time comparison of EP glycans,
GlcA and GaLA standards..........................................151

Figure A.3. El spectral comparison of EP glycans,
GlcA and GaLA standards..........................................152

Figure A.4. GC retention time comparison of EP glycans,
Man and Fuc standards............................................153

Figure A.5. GC retention time comparison of EP glycans,
GlcNAc and GalNAc standards....................................154

Figure A.6. El spectral comparison of EP glycans and Fuc standard.........155

Figure A.7. El spectral comparison of EP glycans and Man standard.........156

Figure A.8. El spectral comparison of EP glycans and GlcNAc standard........157

Figure A.9. Negative ion MS^2 spectral comparison of
glycan 3 to glycan 1..............................................158

Figure A.10. Distinguished D ion and E ion of glycan 2.............159

Figure A.11. Distinguished D ion and E ion of glycan 3.............160
ABSTRACT

GLYCAN CHARACTERIZATIONS WITH APPLICATION IN METAL ION BINDING, THE 1918 INFLUENZA VIRUS NEURAMINIDASE FORMATION AND HUMAN RECEPTORS

By

Hui Zhou

University of New Hampshire, September 2009

The EP glycoprotein, the major protein component isolated from the EP fluid of the mollusc, *Mytilus edulis*, is known to bind many metal ions and to contain 14.3% of carbohydrate in weight. To structurally evaluate the sugar function of this glycoprotein, a new mass spectrometry-based strategy was applied with a focus on the comprehensive characterization of its *N*-glycans. This included: a) accurate mass measurement and GC/MS to determine the monosaccharide compositions; b) positive ion CID-MS\(^n\) to disassemble the deuterated permethylated derivates (CD\(_3\)) to reveal the antenna sequence, and natural methoxyl group location; c) negative ion CID-MS\(^n\) on native glycans to define general antennary topology. Three major complex-type *N*-glycans were characterized with two of them expressing an unique tetrasaccharides antenna sequence: 4-O-methyl-GlcA(1-4)[GlcNAc(1-3)] Fuc(1-4)GlcNAc. The third one is identified as a pentasaccharide antenna: mono-O-methyl-Fuc(1-2)-4-O-methyl-GlcA(1-4)[GlcNAc(1-3)]Fuc(1-4)GlcNAc. This is the first report of these novel *N*-glycans and their intriguing structures may implicate their involvement in metal ion binding.
The neuraminidase (NA) of the 1918 influenza virus (H1N1) has active (tetramer) and inactive forms (monomer/dimer) with difference residing in their N-glycosylations. A structural analysis was carried out on the released N-glycans from both tetramer and dimer forms. Significant amounts of fucosylated N-linked structures were detected in dimer compared to tetramer (relative ratio: 73.6% vs. 18.1%), while no other structural difference were observed. Thus, it appeared that fucosylated glycans can play a key role to inhibit the formation of the tetramer NA, especially those located in the stalk region. However, enzymatically incorporation of fucose onto the tetramer in vitro didn’t alter its enzyme activity, indicating that this regulation must occur at an early biosynthetic stage.

Common lectin methods failed to detect any Sialα(2,3)Gal residues, the proposed avian influenza virus receptors, in the human upper respiratory tract, contrasting current reports that avian viruses can grow efficiently in the human upper respiratory tract tissues. To solve this disparity, two mass spectrometry-based strategies were employed to identify the Sial-Gal linkages on two human cell lines: Detroit 562 and A549, which represent the human upper and lower respiratory tract, respectively. One strategy involved MS^n fragment ion spectral comparison with standard oligosaccharides. A second approach was to differentiate sial-Gal with a specific cyclization reagent DMTMM. Using this reagent, Sialα(2,3)Gal epitope formed a lactone while Sialα(2,6)Gal analog provided a methyl ester, easily distinguished by a 32 Da mass difference. Significant amounts of Sialα(2,3)Gal were found in Detroit 562 cell lines, confirming the presence of avian virus receptors in the human upper respiratory
tract. Thus, the poor human-to-human transmission of avian influenza virus must be attributed to other unknown reasons, rather than the paucity of SA(\(\alpha 2,3\))Gal. This is the first approach to identify influenza virus human receptors by mass spectrometry.
CHAPTER 1

Introduction to Glycosylation

1.1 Significance of Glycosylation

Glycosylation is one of the most important post-translation modifications on proteins. However, due to their unique structural complexities, and the absence of appropriate protocols to characterize their structures, they were less understood [Dell 2001]. Moreover, proteins are synthesized by the accuracy and constraints imposed by a template, while their conjugated glycans are assembled through a series of separate enzymatic steps each of which introduces variability defined by Golgi trafficking and the dynamic aspect of resident cofactors [Taylor 2006; Varki 1999]. Thus, the products represent a series of closely-related structures, usually referred as glycomers [Rademacher 1988]. A well-known example is the glycoprotein obtained from hen’s egg and identified as chicken ovalbumin. This glycoprotein has a single *N*-linked glycosylation site, with over 40 unique glycomers associated at that site, including high-mannose, hybrid, and complex types [Harvey 2000; Zhang, Unpublished data].

Regardless of the challenging problems in structural characterization, it becomes more evident that glycosylation is involved in numerous biological activities, and plays an indispensible role in tissue and cell development.
[Lairson 2008; Varki 1999]. Glycosylation, as well contributes to the conformational structure of proteins, and an excellent example is the human CD2 glycoprotein located on the surface of killer T lymphocytes that mediates binding to target cells [Recny, 1992]. Either mutation of the N-glycosylation site (Asn-65 to Glu-65) or complete removal of N-glycans by PNGase F will eliminate the binding ability to its receptor CD58 on target cells. Since this glycosylation site lies almost on the opposite side of the binding domain, the attached glycan residue does not participate in the binding directly, but it seems to function as a receptor of positive charge of several neighboring lysine residues (Lys61, 69, and 71). No N-glycosylation was needed on this site for rat CD2 since its 61 position is a negative charge amino acid glutamic acid, being able to interact with other positive charge amino acids as well [Taylor 2006]. In addition, the flexible nature of glycosylation appears to act as a cover to protect against proteolysis. This is somewhat analogous to the human influenza virus HA, which appears to gain more N-glycosylation sites after being circulated in the human population, and is proposed to protect against host antibody recognition [Kaverin 2002]. Carbohydrate epitopes are also the recognizing residues in cell adhesion and signaling. The serum mannose-binding protein (MBP), also called mannose-binding lectin (MBL), and mannose receptor on endothelial cells and macrophages, binds to mannose and GlcNAc residues that are frequently found on the surface of microorganisms, rather than human cell surface. Binding initiates a series of
reactions for macrophages to internalize pathogens [Taylor 2006]. The influenza virus HA recognizes the sialic acid-galactose residue on the host cell and upon binding initiates virus entry [Stevens 2006b].

Another fundamental role exhibited by glycosylation is a control of protein folding of newly synthesized glycoproteins. Calnexin and calreticulin are two such chaperone molecules involved in this process, located in the endoplasmic reticulum. This is initiated by the chaperone protein binding to a glucose residue that terminates the 3-branched antenna of an N-linked high mannose glycan. The overall process guides properly-folded proteins to the Golgi for glycan processing or shifts the misfolded proteins to a degradation pathway [Taylor 2006]. Glycosylation is also involved in tissue development, and altered glycan structures are associated with many disease states [Gornik 2008; Zhao 2008]. Thus, the study of glycobiology includes the characterization of structure and their functional relationships [Taylor 2006].

Molecular glycosylation occurs mostly on lipids and proteins where the carbohydrates are referred to as glycans and the products glycoconjugates. More specific terms are becoming popular with glycosylation of proteins and lipids referred to as the glycoproteome and glycolipidome. There are two major classes of glycans within the glycoproteome, the N-, and O-linked glycans, according to the attachment amino acid residue of the sugar moiety.
1.2 Structure Features of *N* - and *O*-glycans

Glycans are assembled with monosaccharide building blocks. The most common of which is a hexapyranoside. Because four of six carbon atoms in a hexose have chiral centers, and each can exist in two configurations, there are a total of sixteen possible hexoses. It would be an extremely difficult task to determine their identity if all isomers existed in nature, fortunately, only nine common monosaccharides have been found in vertebrates [Varki 1999]. These include three hexoses (D-Glucose, Glc; D-Galactose, Gal; D-Mannose, Man), one pentose (D-Xylose, Xyl), one deoxyhexose (L-Fucose, Fuc), one acidic hexose (D-Glucuronic acid, GlcA), two N-acetamidohexoses (*N*-Acetyl-D-glucosamine, GlcNAc; *N*-Acetyl-D-galactosamine, GalNAc), and one unique nine carbon monosaccharide (*N*-Acetylimuraminic acid, NeuAc). Representative monosaccharides and their common abbreviations are shown in Figure 1.1.

*N*-glycans of proteins are attached to the amide nitrogen of asparagine (Asn) residue by an *N*-glycosidic bond. Five different *N*-glycan linkages have been reported [Spiro 2002] of which *N*-acetylg glucosamine to asparagine (GlcNAc/ β1-Asn) is the most common (Figure 1.2) and is the only linkage found in animal cells [Varki 1999]. *N*-glycosylation occurs on many secreted and membrane-bound glycoproteins at the amino acid sequence (sequon) Asn-XXX-Ser/Thr in which XXX can be any amino acid except proline. It is important to note that this sequon is required for the receipt of a glycan.
However, the presence of this sequon doesn't always indicate glycosylation. It was estimated that two thirds of those sequons published in databases are most likely to be glycosylated [Ben-Dor 2004]. Therefore, these sites are commonly referred to as potential N-glycosylation sites.

![Diagram of monosaccharides]

Figure 1.1. Nine common monosaccharides found in vertebrates. A commonly used abbreviation (three letter code) and unique geometric shape are used to represent the different structures throughout this thesis [Varki 1999].
Figure 1.2. Common types of N-glycans and its core structure.
Attachment of N-glycan (GlcNAc) to an asparagine (Asn) residue of proteins via β-linkage. Three different types of N-glycans have been identified, high mannose, complex and hybrid. Each contains the pentasaccharide core Man₃GlcNAc₂ structure [Varki 1999].

All N-glycans share a common pentasaccharide core structure,
Man[α(1-6)Manα(1-3)]Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ(1-1) (Figure 1.2) and differing extensions to this core gives rise to three classification types; (1) high mannose, in which only mannose residues are attached; (2) complex, in which extending residues are initiated with GlcNAc; and (3) hybrid, in which only mannose residues extend from the Man-6 arm, and GlcNAc extensions from the Man-3 arm (Figure 1.2). Another common modification is the addition of an β(1-4) GlcNAc to the central core Man (the one with β(1-4) linkage to GlcNAc) of the complex glycans. Such structures are frequently referred to as “bisecting”. It is common to see other residues (fucose and
xylose) added to the core and they can be further extended into complicating arrays (Hanneman, 2006).

\( \text{O-glycosylation} \) is the second major modification in which the sugar moiety is linked through the hydroxyl group of serine (Ser) or threonine (Thr). Unlike \( N \)-glycosylation, \( O \)-glycosylation does not have an established sequon. Thus, the task to define the exact location of \( O \)-glycosylation sites is challenging due to the presence of numerous such residues in proteins. The most common protein family exhibiting \( O \)-glycosylations is the mucins. \( O \)-glycans are attached through an \( \alpha \)-linked \( N \)-acetylglactosamine (GalNAc) moiety and such glycoproteins are ubiquitous secretions on cell surfaces and in body fluids [Varki 1999]. Mucin type \( O \)-glycans are extended with residues of galactose, \( N \)-acetylglucosamine, fucose, and sialic acid, but not mannose, glucose, or xylose. Eight common core structures of mucin type \( O \)-glycans have been identified (Figure 1.3).

![Diagram of eight common core structures of mucin type O-glycans.](image)

**Figure 1.3. Eight common core structures of mucin type O-glycans.** [Varki 1999]
1.3 Biosynthesis of N-linked Glycans

The structural diversity of N-glycans is a consequence of the biosynthetic pathway that begins on the surface of the endoplasmic reticulum (ER). This pathway can be divided into four distinct stages: (1) formation of a lipid-linked precursor oligosaccharide; (2) *en bloc* transfer of the oligosaccharides to a nascent polypeptide; (3) trimming of oligosaccharides in the ER and Golgi; (4) addition of new sugars in the medial and trans-Golgi.

All eukaryotes share the first three steps of N-glycosylation, but greater glycan complexity is observed in more highly evolved species due to extensive processing that occurs in the Golgi [Prien 2009].

N-glycosylation begins not with a polypeptide substrate, but rather with the polyisoprenoid lipid, dolichol phosphate (Dol-P). The formation of the dolichol-linked oligosaccharide precursor, Glc₃Man₅GlcNAc₂-P-P-Dol, is constructed on both sides of the ER (Figure 1.4). The assembly of this complex molecule is the consequence of 14 enzymes which may lack exacting specificity [Wiswall, unpublished results].
Figure 1.4. Synthesis of the lipid-linked oligosaccharide precursor Dol-P-Glc₃Man₉GlcNAc₂ in the ER.

The dolichol-linked oligosaccharide precursor is transferred to the protein in the lumen of ER, which is catalyzed by the oligosaccharyltransferase enzyme complex. The nascent glycoprotein undergoes immediate quality control processing before moving to the Golgi. Glucosidase I and II in the ER remove the terminal and penultimate glucose residues, which is followed by binding to the quality control lectins, calnexin/calreticulin, to insure proper folding. Glucosidase II acts again on those correctly folded by removing the third glucose and prepares the glycoprotein for processing. The ER \( \alpha(1,2) \) mannosidase I hydrolyzes the middle terminal mannose from Man-6 arm to generate Protein-Man₅GlcNAc₂ complex (B, Figure 1.5) before the glycoprotein exits the ER. Once the glycopeptide enters into the Golgi, another three mannose residues can be removed by a series of Golgi mannosidases to generate Protein-Man₅GlcNAc₂ complex (C, Figure 1.5). High-mannose type glycans
will remain on the glycoprotein if no processing occurs. Recent results in our
laboratory have provided structural evidence that processing is not as
exacting as often described with multiple isomeric intermediate products
[Prien 2009].

As for complex and hybrid type structures, the processed glycan
linked protein is rebuilt according to the specificity of the tissue. This
rebuilding is initiated by the GlcNAc-transferase I; an enzyme that adds one
GlcNAc residue to Man-3 arm (D, Figure 1.5). Following the addition, other
mannosidases in the Golgi continue to remove the remaining two mannoses
(outside the core) on the Man-6 arm. Further addition of GlcNAc can be
obtained by different GlcNAc-transferases in Golgi, providing cell specificity
and branching pattern to the product glycans. Elongation on the GlcNAc
antennae is provided by the addition of a Gal and sialic acid residue (G, H,
Figure 1.5). The resulting sialic acid-galactose moiety is often a capping
structure (no more elongation), which is also the receptor of many pathogens,
such as influenza virus.
Figure 1.5. The N-glycosylation pathway: trimming and elongation in ER and Golgi [Lehrman 2004].
1.4 Glycome and Structural Analysis of Glycans by MS

The total complement of glycans in a cell or organism is defined as the glycome. Genomic DNA sequences dictate the primary structure of proteins. In contrast, glycan structures are not encoded directly by DNA sequences. In fact, the glycan structure is determined by the resident glycosyltransferases in the Golgi varies between cell types and at different stages in development, thus, the glycome is a characteristic of a particular cell type at a specific state. Importantly, glycosylations is known to be altered during many disease conditions, such as cancer [Zhao 2008; Prien 2008; Dennis 1999]. Thus, analysis of glycome provides a basis for understanding the functions of glycans in cell differentiation and diseases.

Since glycans have different physical and chemical properties than peptides and exhibit glycomer variations, both N- and O-glycans must be released and characterized differently. N-glycans can be released from asparagine using a bacterial enzyme known as peptide-N-glycosidase (N-glycanase F, PNGase F). PNGase F is able to remove all types of N-glycans from the peptides, except for some modification (such as 3-Fuc) on the core innermost GlcNAc, which is usually observed in plants, insects, and parasites [Hanneman 2006]. At this time there has been no comparable O-glycanases reported, and O-glycans must be released by chemically. The classical method employs strong base (NaOH) and a reducing agent NaBH₄, which converts the released glycan to its corresponding alditol in situ to
prevent any “peeling” reactions [Karlsson 2002]. Alternative release procedures have been established for both N- and O-linked glycans using hydrazine at different temperature, but yields are compromised [Merry 2002].

Although many strategies and methods have been presented to characterize glycan structure [Zaia 2009; Harvey 2009a; Harvey 2006; Zaia 2004], structural analysis remains a daunting task. One difficulty lies in the structural complexity of each individual glycan. A complete glycan structure should include the following information; composition, monosaccharide identity, anomeric configuration, intervening linkage and, branching information (generally referred as topology). Another problem is the heterogeneous of glycans (referred as glycomers). Since glycans share similar chemical characteristics, it is challenging to resolve the glycans by chromatographically, especially for structural isomers. Thus, due to the above practical problems, the reporting of a partial sequence, assumed motifs, arrays absent of linkage and branching information, multiple structures enclosed in brackets are currently acceptable in scientific reports.

Numerous reports have been published introducing alternatively strategies for structural analysis, most frequently using combinations of chromatography [Merchref 2008; Novotny 2008; Volpi 2008], lectins [Gemeiner 2009; Hirabayashi 2008], and enzymes [Park 2008; Domann 2007]. Moreover, the lack of exacting specificity offered with many biological tools (lectins) and their inability to define positions of antennal release
(exoglycosidases) coupled with the ultimate need for product characterization introduce irresolvable complications. As we will discuss in chapter 4, a lectin approach is neither reliable nor accurate to determine sialic acid galactose linkage. Our group has established a unique strategy to characterize glycan structure by multiple steps of disassembly on permethylated glycans in ion trap mass spectrometry (MS^n) [Prien 2009; Prien 2008; Ashline 2007; Hanneman 2006].

Glycan methylation has long been a reliable derivative for carbohydrate analysts. Methylated products exhibit improve carbohydrate signals in mass spectrometry, and provide extensive structural detail upon sequential mass spectrometry (MS^n) of product ions [Ashline 2005]. This strategy has been successfully demonstrated on identification of novel N-glycans from therapeutical IgG [Ashline 2007], unusual core-modification of N-glycans from C. elegans [Hanneman 2006], and cancer biomarkers [Prien 2008]. An example that has further demonstrated the power of sequential MS^n was an application characterizing the isomeric glycan structures detected in bovine ribonuclease B (RNase B) [Prien 2009]. The glycomers Man$_5$GlcNAc$_2$, Man$_7$GlcNAc$_2$ and Man$_8$GlcNAc$_2$ from this glycoprotein were long established to posses seven isomers, one, and three each for the latter two, respectively. There were multiple reports supporting this observation, most recently by a combination of HPLC coupled with FTICR-MS, [Zhao 2008; Costello 2007]. Recent studies in our laboratory
have identified thirteen isomers from the same three glycomers, six previously unreported (Star-highlighted, Figure 1.6) [Prien 2009]. Additionally, incomplete α-glucosidase activity on the total glycans of RNase B appears to account for these isomeric structures [Prien 2009]. The detail of this strategy will be discussed in Chapter 2 and Chapter 3 to identify N-glycan structures from EP glycoprotein and neuraminidase of influenza virus.

Figure 1.6. Thirteen isomeric glycans were identified from bovine ribonuclease B. Six of them were previously unreported (designed as red star) [Prien 2009].

Key features that make MS\textsuperscript{n} a valuable strategy for understanding carbohydrate structures can be summarized as follows: (1) Samples prepared by methylation and reduction places monomers in a glycan array, thus, termini, extending, and branched residues can be defined. (2) Pathways of disassembly can be directed which defines domains of structure.
providing antennal sequence. (3) Fragments exposed during disassembly that fail to define a single topology are an indicator of structural isomers, which may be isolated and selectively characterized. (4) A searchable library of small oligomer fragments provides the details of linkage, monomer ID, and stereochemistry.
CHAPTER 2

Characterization of Novel \( N \)-Glycans of the EP Glycoprotein from the Mollusc: *Mytilus Edulis*

### 2.1 Introduction

**Biomineralization**

Biomineralization is the matrix-mediated processes by which living organisms regulate and organize mineral deposition [Lowenstam 1989; Seed 1980]. These processes involve the selective extraction and uptake of inorganic elements from the local environment and their incorporation into functional structures under strict biological control such as seashells, bone, teeth, ivory, corals, etc. These structures fulfill a diversity of biological functions [Mann 2001] including protection, exoskeleton (eg. seashell), endoskeleton (eg. invertebrate bone), and gravity receptor (eg. inner ear bones in mammals), optical imaging (eg. calcite lens in the compound eyes of trilobites), and temporary storage (eg. amorphous calcium carbonate in plant leaves to store calcium). The mineral crystals formed by the organisms are usually referred to as biominerals. Biominerals has two unique features distinguishable to other inorganic deposit minerals. First, the biominerals have unusual, diverse, complex, external morphologies, and sometimes even
chirality. Second, many biominerals are actually composites or agglomerations of crystals separated by organic material/matrix. Generally, living organisms regulate mineralization via the use of organic macromolecules (proteins, polysaccharides). These macromolecules are responsible for nucleation, growth regulation, and growth cessation of the mineral crystals [Belcher 1996; Addadi 1992; Mann 1988].

**The EP Glycoprotein from *Mytilus Edulis***

*Mytilus edulis*, commonly known as blue mussel, is one of the most extensively studied marine organisms and found worldwide in most polar waters. In the shell of the mollusc *Mytilus Edulis*, calcium carbonate is the dominant composition accounting for 95 – 99% by weight [Yin 2005]. The exterior of the shell is covered by a thin layer of sclerotized protein, termed the periostracum. A series of compartments are involved in the shell formation. The most important compartment is the inner shell surface, the extrapallial cavity and the outer mantle epithelium [Crenshaw 1980]. The extrapallial (EP) fluid is secreted by the outer mantle epithelium cells and fills the extrapallial cavity, the space between the shell and the outmost visceral organ of the mantle. The EP fluid contains proteins, carbohydrates, and amino acids, and is also believed to be supersaturated with respect to the shell minerals. Its anatomical location and biomolecular components of the EP fluid imply its key role in the formation of the shell *in vivo* [Misogaines 1979; Weiner 1977].
The EP glycoprotein, the major protein component of EP fluid, was isolated and characterized as a 213 amino acid protein with a single N-glycosylation site [Yin 2005; Hattan 2001]. This glycoprotein comprises 56% of the total protein in the EP fluid and the carbohydrate accounts for 14.3% in the EP glycoprotein by weight. The EP protein is acidic (PI = 4.43) and rich in histidine (11.1%) as well as in Asp and Glu residues (25.2 % total) [Hattan 2001]. The intriguing property of this glycoprotein is its calcium-binding-induced self-assembling into a series of higher order protomers [Hattan 2001]. Circular dichroism results show that the protein-calcium binding/protomer formation is coupled to a significant rearrangement of the protein’s secondary structure [Hattan 2001]. Besides Ca\(^{2+}\) ion, the EP glycoprotein also binds to other divalent ions, such as Cd\(^{2+}\), Cu\(^{2+}\), Mn\(^{2+}\), and Mg\(^{2+}\) [Yin 2005]. Although the presence of several carboxylic acid residues (eg. Asp) may serve as metal ion binding sites; the binding domain of the metal ions can not be fully defined only by the protein backbone sequence [Yin 2005].

The EP glycoprotein sequence has indicated that it may be the same protein as a heavy metal binding protein (HIP) from the hemolymph of *M. edulis* [Yin 2005] and as a histidine-rich glycoprotein (HRG) from the plasma of *M. edulis* [Nair 1999]. It seems that these proteins are widely present in the various tissues of *M. edulis*, and also in other marine animals [Nair 1998], suggesting its various biological functions besides shell
formation such as: metal ion transportation, de-toxification, etc. [Yin 2005; Hattan 2001; Nair 2001a; Nair 2001b; Nair 2000]. The key feature of the EP glycoprotein is its strong affinity with many metal ions such as: Cu$^{2+}$, Cd$^{2+}$, Mg$^{2+}$, Mn$^{2+}$, Hg$^{2+}$, Ni$^{2+}$, Pd$^{2+}$, Zn$^{2+}$, and Ca$^{2+}$ [Devoid 2007; Hattan 2001; Nair 2000]. The metal ion binding ability of this protein is remaining even after heat denaturation and polyacrylamide gel electrophoresis [Nair 1999], suggesting the chelation between protein and metal ions are not completely dependent on the proximity of acidic amino acids in the native conformation. Given the carbohydrate components having acidic residues and multiple antennae [Naggar 2000], it is reasonable to suspect that N-glycosylation may take part in the metal ion binding directly. Therefore, the complete structural information of N-glycans is essential to understand the EP glycoprotein’s biological functions. The structures of N-glycans from EP glycans have proven to be quite unique and have eluded past attempts at characterization [Yin 2005; Naggar 2000]. Herein we resolve the structural ambiguity by combining a series of techniques such as: accurate mass measurement, GC/MS monosaccharides composition analyses, chemical modification, positive ion and negative ion MS$^n$ disassembly.

**New MS-based Strategy to Characterize Complex-type N-glycans**

The complete structural information of glycans should include following information such as: monosaccharides identities, the linkage positions of inter-residues, anomericity, and array topology, where array
topologies describe linear and branching sequence information as discussed in Chapter 1. As for non-mammalian system, glycan structural characterization may be particularly challenging because some unusual monosaccharide residues may be encountered; or the monosaccharides may be modified by natural methylation or carboxylation, etc [Voisin 2005; Srikrishna 2005; Puanglarp 1995]. In these cases, the primary task will be identifying the possible monosaccharides identities and compositions. Therefore, reliable techniques have to be sought to determine the monosaccharide compositions, for example: combining GC/MS and accurate mass measurement. GC/MS is a classical method on carbohydrate composition analysis and linkage analysis [Levery 1987].

Negative ion MS/MS on small oligosaccharides has been demonstrated to provide some complementary structural information such as: chain location, branching pattern and sialic acid linkage [Chen 2008; Yu 2006; Chai 2006; Chai 2005; Chai 2002; Chai 2001]. Recently this technique gained extensive attention and has been reported to provide unique fragment ions of multi-antennary N-glycans [Harvey 2005a; Harvey 2005b; Harvey 2005c], difficult to obtain through positive ion fragmentation. A good example is the formation of ion D [Harvey 2009b; Harvey 2008a; Harvey 2005b; Harvey 2005c] containing the moiety of the core mannose and 6-branching mannose residues and associated antennae, which is diagnostic to determine the Man-6 branching composition. This information is usually
difficult to obtain under positive ion CID fragmentation pathway due to the labile linkage between 6-mannose and the neighboring antennary GlcNAc. Due to the acidic property of the glycans from EP protein, it may be easy to seek the branching information via negative MS$^n$.

So far many mass spectrometric strategies have been proposed to characterize glycan structures [Harvey 2009a; Zaia 2009; Merchref 2009]. The most widely-used strategies are combining both exoglycosidase digestion and mass spectrometry or HPLC before and after digestion. The monosaccharides and their linkages are revealed by the specificity of the employed enzymes, and the general topology is deduced by the change in the property being measured (chromatographic retention time, mass, etc.). Although popular, these strategies are highly dependent on a readily available source of the exoglycosidase in a pure state. Furthermore, the terminal residues are removed by the specific enzyme step by step, some structural information may be lost during the process, especially for those cases with multiple positions available to attach the terminal residues removed by enzymes. Most importantly, specific enzymes may not be available for less common monosaccharide residues such as galacturonic acid and N-acetylgalactosamine, or extra modification on the common monosaccharides, such as methylation, acetylation, etc. Sometimes biosynthetic pathways have to be fully considered to propose the most plausible structure [Dell 2001]. Therefore, a general strategy is indeed
needed to identify the complete glycan structures with unusual monosaccharides residues and other non-carbohydrate modifications, especially for those less-well studied species. A new strategy is proposed herein which combines various modern mass spectrometry techniques and chemical modifications to characterize the N-glycans of the EP glycoprotein from the mollusk: *Mytilus edulis*. This strategy includes following techniques: GC/MS and accurate mass measurement to define the monosaccharide identities and compositions; positive MS\textsuperscript{n} on deuterated permethylated derivates of glycans to characterize the antenna sequence and inter-residue linkage information; negative MS\textsuperscript{n} on native glycans to determine the general branching mode (topology).

### 2.2 Experimental Procedures

#### Materials and Reagents

Blue mussels, *M. edulis*, were collected from the waters off the northeast coast of the United States (Dover Point, NH), and the EP protein was extracted and purified following procedures as described previously [Hattan 2001]. The enzyme glycerol-free PNGase F and associated reagents (10 × denaturing buffer, 10X NP-40, 10X G7 buffer) were purchased from New England Biolabs (Beverly, MA); Dowex 50W X 8 hydrogenform cation-exchange resin, dimethyl sulfoxide, sodium hydroxide, iodomethane,
iodomethane-d₃, iodine; sodium borohydride, sodium borodeuteride, tetrahydrofuran, toluene and ethanol were obtained from Sigma-Aldrich (St. Louis, MO); chloroform, acetonitrile, and methanol were from EMD chemicals (Gibbstown, NJ); water was from J. T. Baker (Phillipsburg NJ). Chemicals and solvents were of ACS or HPLC grade.

**Glycan Sample Preparation**

\(N\)-glycans were enzymatically released from the purified EP glycoprotein by PNGase F following the procedure according to the manufacture instruction. In briefly, 1 mg dried EP protein was dissolved in 50 \(\mu\text{L}\) 1X protein denature buffer and incubated at 100 °C for 45 minutes. Subsequently, 5 \(\mu\text{L}\) of 10X NP-40 and 5 \(\mu\text{L}\) of 10X G7 buffer were added into the solution, along with 1-2 \(\mu\text{L}\) PNGase F. The enzyme reaction was kept at 37 °C for at least 10 hours. The glycans were separated from protein, enzymes and detergents by passage through a C-18 Sep-Pak cartridge (Waters, Milford, MA). In selected cases the \(N\)-glycans were reduced with sodium borohydride (10 mg/mL NaBH₄ dissolved in 0.01 M NaOH) at room temperature overnight and desalted with a column of Dowex 50W cation-exchange resin. Further purification step was carried out on a hand-packed cellulose column (Medium fibrous, Sigma, St. Louis, MO) in an organic solvent of 1-butanol: EtOH: H₂O (4:1:1, v/v/v) [Shimizu 2001]. Briefly, the glycan samples were taken up in 200 \(\mu\text{L}\) of EtOH: H₂O (1:1, v/v), and mixed with 2 mL of the organic solvent, and then applied onto the cellulose
column. The column was washed with 4 mL of the organic solvent before the glycans were eluted with 4 mL of EtOH: H₂O (1:1, v/v). The reduced glycans were dissolved in DMSO and permethylated by treating with powdered sodium hydroxide with iodomethane using the method previously reported [Ciucanu 2003]. Deuterium permethylation was performed employing the same procedure except that iodomethane was replaced by iodomethane-d₃.

**Conversion of Glucuronic Acid to Glucose**

Direct reduction of the carboxylic acid group to the primary hydroxyl group was carried out using the method reported by Kanth and Periasamy [Kanth 1991]. Briefly, the native glycans were dissolved in fresh distilled tetrahydrofuran, and sodium borodeuteride was slowly introduced to form a suspension. After gas evolution ceases, a small amount of iodine was added and the reaction was allowed to stir overnight at 50 °C. Acetic acid was added to stop the reaction in the ice bath, and excess borate was removed by co-evaporation with 2 mL ethanol, 3 × 2 mL of 1% acetic acid in methanol, and 2 × 2 mL toluene. The product was desalted using Dowex 50W cation-exchange resin and permethylated and stored for MSⁿ analysis.

**Mass Spectrometry**

Mass spectrometry was performed on a Thermo linear ion trap LTQ instrument, (Thermo, Waltham, MA) equipped with a nanospray ESI source. Native (without permethylation) glycans were dissolved in methanol: H₂O (1:1)
containing 0.5% acetic acid and permethylated glycans in 100% methanol for mass spectrometry. Samples were infused at flow rates ranging between 0.30 and 0.60 μL/min, and spectra were collected using Xcalibur 1.4 software (ThermoFinnigan). Accurate mass determination of the native glycans was obtained under negative ion mode using the ultra-zoom scan function and off-line internal mass calibration against the disialylated glycan (Hex)_2(HexNAc)_2(NeuAc)_2+(Man)_3(GlcNAc)_2 from IgG (Sigma-Aldrich, St. Louis, MO). All positive ion and negative ion MS^n spectra were acquired at 25-40% normalized collision energy with capillary temperature at 230 °C, and the activation Q and activation time were set at their default values, 0.25 and 30 ms, respectively.

**GC/MS Monosaccharide Composition Analysis.**

GC/MS composition analysis was performed on a Polaris Q TraceGC Ultra (ThermoFinnigan, Waltham, MA) equipped with a DB-5, 30 m × 0.25 mm × 25 μm column (Restek, Bellefonte, PA). The overall experiment condition was following the partially methylated alditol acetates (PMAAs) linkage analysis procedure reported by Levery [Levery 1987]. Instead of permethylated glycans, the native glycans was hydrolyzed, reduced, and acetated for GC/MS analysis. In summary, the dried glycans was dissolved in 100 μL of 0.5 N H₂SO₄ in 90% acetic acid, and incubate at 80 °C for 6 hrs. Then the reaction solution was neutralized by 130 μL of 0.5 N NaOH and repeatedly dried with several additions of ethanol. The hydrolyzed
monosaccharides were reduced with NaBD$_4$ (10 mg/mL in 0.01N NaOH) overnight and the excess borate was removed by several addition of methanol and dried. Following the acetylation was carried out with acetic anhydride at 100 °C for 2 hrs. Excess acetic anhydride was removed by several addition of toluene and dried under nitrogen stream. The residues were partitioned between chloroform and water layers. The chloroform layer was collected, and dried, and reconstituted in 20 μL hexane. Take 1 μL for GC/MS. Because the hexauronic acid moiety was modified naturally by the methoxyl group, two uronic acid standards, GlcA and GaIA, were methylated using the same procedure as permethylation [Ciucanu 2003], rather than acetylated after hydrolysis and reduction. The temperature program for gas chromatography was set as below: hold at 40 °C for 1 minute, and then elevate to 130 °C with the ramp 30°C/min. After reach to 130 °C, the ramp was decreased to 4 °C/min until 250 °C.

### 2.3 Results

**Mass Measurement of Glycan 1 and GC/MS Composition Analysis**

The N-glycan pool released from the EP glycoprotein showed prominent three glycans signals (1-3) with the major component **glycan 1** (Figure 2.1). All the glycans showed good signals under negative ion ESI-MS with multiple negative charge states ranging from 3 to 5 (Figure 2.1),
indicating the presence of multiple acidic groups: such as carboxylic acid, phosphate groups, or sulfate groups, etc. Deducing the molecular weight of these N-glycans (Figure 2.1) from the MS profile, it suggested that these N-glycans may contain unusual modification or residues because common nine monosaccharides as discussed in Chapter 1 did not fit their molecular weights. Therefore, the primary task was to determine the possible substituent compositions of these glycans. The negative charge groups were proposed to be mono-methoxyl-phosphate groups (Figure 2.2) previously [Naggar 2004] according to the nominal mass of glycan 1 (M.W. 4027).

Whereas, disassembling of glycan 1 by negative ion MS^n show a series of prominent neutral loss of 190 Da [MS^2 on 1006.2 (4-)→1278.3 (3-), Δ = 190; MS^3 on 1006.2 (4-)→1162.0 (3-)→1648.5 (2-), Δ = 190] and 32 Da [MS^2 on 1006.2 (4-)→998.2 (4-), Δ = 32; MS^3 on 1006.2 (4-)→1162.0 (3-)→1151.4 (3-), Δ = 32] (see appendix for spectra), suggesting the presence of terminal mono-methoxyl-hexauronic acid moiety (C_7H_{10}O_6 M.W. = 190). In order to determine the exact molecular weight of glycan 1 to clarify this uncertainty, accurate mass measurement was performed with mass calibration against the internal standard glycan: Man_3GlcNAc_4Gal_2Sial_2 from human IgG using LTQ nano-ESI ultra-zoom scan technique (Figure 2.2). The adjusted molecular mass of glycan 1 was 4027.5441 Da, fitting to a composition (mono-MeO-HexA)_4(HexNAc)_10(Hex)_3(DeoxyHex)_5 (4027.4589, dif. 21 ppm) rather than the previously proposed composition:
(mono-MeO-phosphate)$_6$(HexNAc)$_{10}$(Hex)$_7$(DeoxyHex)$_1$ (4027.0409, dif. 125 ppm) (Figure 2.2), both of them having the identical nominal mass: 4027 Da. To further confirm the presence of the mono-MeO-HexA residues and other monosaccharides, the GC/MS monosaccharide composition analyses were carried out. Due to the natural methoxyl modification on HexA moiety, the total EP glycan pool were hydrolyzed to monosaccharide residues by strong acidic solution, then reduced with sodium borohydride and per-methylated and compared with two HexA standards (glucuronic acid and galacturonic acid) following the exactly same treatment. According to this result, the HexA moiety was assigned as glucuronic acid unambiguously (Table 2.1). No isomeric GalA residue was detected. Because of the absence of GalNAc, Gal, Glc, and Rha residues (Table 2.1), it was evident to assign HexNAc, Hex, and DeoxyHex as following: GlcNAc, Man, and Fuc, respectively, according to the GC/MS results using the acetylated derivatives (See Appendix for additional Spectra). Thus, the monosaccharides composition of glycan 1 was defined as (mono-MeO-GlcA)$_4$(GlcNAc)$_{10}$(Man)$_3$(Fuc)$_5$ unambiguously. The presence of ten GlcNAc residues and only three mannose residues strongly suggested that glycan 1 was a complex type N-link glycan with multiple antennae. Thus, next step will be identifying each antenna sequence.
Figure 2.1. Negative ion LTQ ESI-MS of total enzymatically released N-linked glycans from the EP glycoprotein. All the peaks are multiple charges and the charge state of each individual peak is annotated in the parenthesis. Three major glycans (1-3) are observed and each glycan may have multiple peaks due to the presence of various charge states. The molecular weight (M.W.) of each glycan is calculated and showed in the right-top panel.

**(+) MSⁿ analysis of Deuterated Permethylated Glycan 1**

As demonstrated in our previous report, permethylated glycan show better signals compared to native glycan [Reinhold 1995; Ashline 2005].

More important, blocking the free hydroxyl group with a methoxyl group, it provides the opportunity to obtain the inter-residue linkage information with MSⁿ disassembling pathway [Ashline 2005]. The monosaccharide composition of glycan 1 \([(\text{mono-MeO-GlcA})₄(\text{GlcNAc})₁₀(\text{Man})₃(\text{Fuc})₅\] clearly indicated that glycan 1 was a complex type N-linked glycan. In general under positive ion collision-induced-dissociation (CID) fragmentation pathway, the antenna B ion is abundant and distinguished in the low mass range in MS/MS.
Figure 2.2. Negative ion ESI accurate mass measurement of glycan 1 against the internal standard: Man$_3$GlcNAc$_2$Gal$_2$Sia$_2$ from IgG by LTQ ultra-zoom scan. The accurate molecular weight of glycan 1 is calculated as 4027.5441 Da, which is closer to the structure including mono-methoxyl-GlcA unit rather than the previously proposed structure with MeO-phosphate residue [Naggar 2004].
Table 2.1: GC/MS monosaccharide composition analyses of EP glycans. n/a: not applicable. Monosaccharide identities were assigned confidently by comparison of both retention time and El mass spectra (See appendix for spectra)

<table>
<thead>
<tr>
<th>STD</th>
<th>Acetylated derivative Retention time (Min)</th>
<th>Methylated derivative Retention time (Min)</th>
<th>EP Glycan (Min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rha</td>
<td>18.98</td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td>Fuc</td>
<td>19.23</td>
<td></td>
<td>19.23</td>
</tr>
<tr>
<td>Man</td>
<td>24.98</td>
<td></td>
<td>24.98</td>
</tr>
<tr>
<td>Glc</td>
<td>25.27</td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td>Gal</td>
<td>25.49</td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td>GalNAc</td>
<td>29.17</td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>28.61</td>
<td></td>
<td>28.61</td>
</tr>
<tr>
<td>GlcA</td>
<td></td>
<td>12.13</td>
<td>12.13</td>
</tr>
<tr>
<td>GalA</td>
<td></td>
<td>13.16</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Figure 2.3. Positive ion ESI MS/MS (→ 1706.8) spectrum of deuterium methylated glycan 1. Distinguished antenna B ion (m/z 952.6, monoisotopic) is abundant in the low mass range (<1200 Da) because the glycosidic bond between chitobiose core and complex glycan antenna’s GlcNAc is easily broken down under positive ion CID fragmentation mode. All peaks larger than 1200 Da are multiple charge ions, corresponding to large fragments. All fragment ions contain a sodium atom.
of permethylated N-glycan spectra due to the labile nature of the glycosidic bond between the core Man-3 or Man-6 and the neighboring GlcNAc in antenna. In order to locate the naturally occurring methoxyl group position, the deuterium methylated derivative was prepared in parallel and subjected to MS/MS fragmentation, showing a distinguishing deuterated antenna B ion (m/z 952.6, Figure 2.3), which accounts for a composition \([((\text{mono-methoxyl-GlcA})(\text{GlcNAc})_2(\text{Fuc}))\) (Figure 2.3).

Antenna sequence was obtained by further MS\(^n\) disassembly of the deuterated methylated B ion. First of all, the evident ion at m/z 920.5 (\(\Delta = 32\)) from MS\(^3\) fragmentation pathway (1078.6\(\rightarrow\)952.6, Figure 2.4) clearly confirmed the presence of one natural methoxyl group in the antenna. Neutral loss of a CD\(_3\)OH moiety was also observed at m/z 917.5 (\(\Delta = 35\)), but relatively low in abundance compared to the loss of a CH\(_3\)OH molecule, indicating the loss of CH\(_3\)OH was favored for some reason. Further disassembly of ion m/z 920.5 showed a dominant ion at m/z 770.5 in MS\(^4\) spectrum (1078.6\(\rightarrow\)952.6\(\rightarrow\)920.5, Figure 2.4, inset), indicating the GlcA moiety was the terminal monosaccharide and the methoxyl group located at C-4 of GlcA (Scheme 2.1). Therefore it can be explained rationally that the neutral loss of 32 Da is favored due to the formation of a conjugated double bond (C-4 and C-5) with the carboxyl ester group at C-6 in GlcA. Furthermore, the secondly abundant ion at m/z 681.4 (\(-\text{C}_{12}\text{D}_{12}\text{H}_9\text{O}_5\text{N}, \Delta = 271\)) from MS\(^3\) fragmentation (1078.6\(\rightarrow\)952.6, Figure 2.4) indicated the presence of a
non-reducing end terminal GlcNAc due to the labile nature of the GlcNAc glycosidic bond. In addition, the most abundant ion at \( m/z \) 789.5, which comes from the retro Diels-Alder fragmentation of internal GlcNAc moiety, indicated either C-4 or C-6 of internal GlcNAc was linked by other residues. Further fragmentation of this ion showed a dominant fragmentation ion at \( m/z \) 665.4 (1078.6→952.6→789.5, Figure 2.5a), which derived from a [1,7] sigmatropic hydrogen rearrangement [Lee 2004] (Scheme 2.2), suggesting the other three residues attaching to C-4 of the internal GlcNAc. Given the presence of both terminal GlcNAc and GlcA residues, the Fuc residue was the only choice for the core branching residue connected by other three monosaccharides in the antenna. Further MS\(^5\) disassembly of the ion at \( m/z \) 665.4 (1078.6→952.6→789.5→665.4, Figure 2.5b) showed a distinguishing fragment ion at \( m/z \) 322.2, indicating the GlcA moiety connected to C-4 of the branching Fuc residue. While the abundant ion at \( m/z \) 378.2, most likely came from a [1,5] sigmatropic hydrogen rearrangement (Scheme 2.3), showing that the terminal GlcNAc was attached to C-3 rather than C-2 of the branching Fuc. Another MS\(^n\) fragmentation pathway containing the GlcNAc and Fuc disaccharides moiety also supported the notion that both terminal GlcNAc and GlcA attaching to C-3 and C-4 of the branching Fuc in the antenna, respectively (Figure 2.6).
Scheme 2.1. Proposed mechanism to form ion m/z 770.5. The naturally occurring methoxyl group located at C-4 of GlcA facilitate the natural MeOH loss to form α,β double bond conjugated with the C-6 ester intermediate. Further retro Diels-Alder fragmentation leads to open the six-member ring of GlcA to form $^0_2X$ ion at m/z 770.5. All ions mass include one sodium atom.

Figure 2.4. (+) ESI MS$^3$ Fragmentation of the deuterated antenna B ion (→ 1706.8 → 952.6), and (inset), MS$^4$ spectrum of m/z 920.5 (1706.8 → 952.6 → 920.5) of glycan 1. Neutral loss of a MeOH from antenna B ion as designed confirms the presence of naturally occurring methoxyl group in glycan 1. Further fragmentation of m/z 920.5 forming a retro Diels-Alder fragment $^0_2X$ ion 770.5, demonstrate that the natural methoxyl group is locating at C-4 of GlcA and easily lost to form a conjugated double bond with the C-6 carboxyl group in GlcA. All fragments are single charge and contain one sodium atom.
Scheme 2.2. Proposed deuterium [1.7] sigmatropic shift mechanism to form ion m/z 665.4. The formation of ion m/z 665.4 indicates the terminal GlcNAc and 4-O-methyl-GlcA residues attach to the branching fucose either C-3 or C-4 position. All ions mass include one sodium atom.

Scheme 2.3. Proposed proton [1.5] sigmatropic shift mechanism to form ion m/z 378.2 and 275.2. The formation of ion m/z 378.2 indicates the terminal GlcNAc connect to C-3 of the internal branching fucose residue. All ions mass include one sodium atom.

In order to further confirm the presence of GlcA residue, the carboxyl group (-COOH) in the GlcA was reduced to the primary deuterated hydroxyl group (CD₂OH) with NaBD₄ and I₂ [Kanth 1991]. Reduced products were then permethylated and subjected to positive ion MSⁿ analysis. An evident antenna B ion was observed at m/z 907.5 in the MS/MS spectrum (→1630.6). Further fragmentation of this ion at m/z 907.5 confirmed the presence of
terminal Glc (Figure 2.7). Therefore, the antenna sequence was characterized unambiguously as 4-O-methyl-GlcA(1-4)[GlcNAc(1-3)]Fuc(1-4)GlcNAc. This unique antennary sequence suggested that glycan 1 contains four identical antennae with a single fucosylated core structure [4-O-methyl-GlcA(1-4) [GlcNAc(1-3)]Fuc(1-4) GlcNAc]4+Man3GlcNAc2Fuc.

Figure 2.5. ESI MS^n analyses of two ions \( m/z \) 789.5 (A, \( \rightarrow 1076.8 \rightarrow 952.6 \rightarrow 789.5 \)) and 665.4 (B, \( \rightarrow 1076.8 \rightarrow 952.6 \rightarrow 789.5 \rightarrow 665.4 \)) from antenna B ion of glycan 1. Monoisotopic peaks in A and B are annotated. All fragments are single charge and contain one sodium atom. Both ions \( m/z \) 665.4 in A and \( m/z \) 378.2 in B are the interesting fragmentations, which come from hydrogen sigmatropic rearrangement in gas phase. The proposed mechanisms were discussed in Scheme 2.2 and 2.3, respectively.
Figure 2.6. ESI MS$^5$ (→1707.1→952.5→789.5→427.2) spectrum of deuterium methylated glycan 1. All fragments are single charge and contain one sodium atom.

Figure 2.7. ESI MS$^3$ (→1630.6→907.5) spectrum of the antenna B ion after converting the carboxyl group to the corresponding deuterated hydroxyl group. All fragments are single charge and contain one sodium atom.
Since glycans were released by PNGase F, it can be deduced that core fucose was in C-6 position of the innermost GlcNAc. But the position of each antenna remains unknown.

**Negative Ion MS\textsuperscript{n} Analysis of Glycan 1 to Define the Branching Mode**

Negative ion MS/MS techniques have proved to be most useful to determine core fucose linkage and branching patterns [Harvey 2008; Harvey 2005b; Harvey 2005c]. Although the exact fragmentation mechanism was obscure, repeated and reliable diagnostic ions such as ion E, ion D [Harvey 2005b; Harvey 2005c] have been observed consistently in many large N-glycans and been recently summarized by Harvey *et al.* [Harvey 2009b; Harvey 2008]. The distinguished $^{2,4}$A$_7$ ion at $m/z$ 928.8 from MS/MS spectrum ($\rightarrow$1005.9) of the native nonreduced glycan (Figure 2.8) indicated the core fucose was attaching to C-6 of ultimate GlcNAc [Harvey 2008; Harvey 2005b; Harvey 2005c], which was in agreement with PNGase F enzyme activity. To obtain individual antenna’s specific location to the Man-3 or Man-6 arms, negative ion MS\textsuperscript{n} on the reduced native glycan was carried out. In the MS/MS spectrum (Figure 2.9), a distinguishing D ion at $m/z$ 903.3 and $^{0,3}$A$_5$ ion at $m/z$ 867.3 showed that two antennae were attached to Man-6 arm [Harvey 2008], suggesting that the other two antennae were attached to Man-3 arm. Further MS\textsuperscript{3} fragmentation on C$_4$ ion at $m/z$ 831.3 ($\rightarrow$1006.5$\rightarrow$831.3) show a relatively abundant E ion at $m/z$ 792.2 [Harvey 2005b; Harvey 2005c]; while MS\textsuperscript{4} disassembling on C$_4$ ion $m/z$ 831.3 ($\rightarrow$1006.5$\rightarrow$903.3$\rightarrow$831.3), the
specific pathway from Man-6 arm, lack that diagnostic E ion (Figure 2.10a
and 2.10b). The presence of E ion from →1006.5→831.3 pathway [Harvey
2005b; Harvey 2005c], mainly from Man-3 arm, showed the two antennae
connected to C-2 and C-4 of Man-3 arm, respectively. While the absence of E
ion from Man-6 arm pathway may suggest two antennae most likely attach to
C-2 and C-6 of core Man-6, respectively, not same as the Man-3 arm.
Therefore, the complete structure of glycan 1 was characterized as shown in
figure 2.11.

**Structural Characterization of Minor Glycan 2**

The monosaccharides composition of glycan 2 was defined as
(mono-MeO-GlcA)₅(GlcNAc)₁₂(Man)₃(Fuc)₆ according to its molecular weight
(4769.4 Da, Figure 2.1). Positive ion MS/MS fragmentation on deuterated
permethylated peak shows that glycan 2 has the same antenna B ion as
glycan 1. Further MSⁿ disassembly of the antenna B ion show the exact
fragmentation pattern as 1, indicating that the antenna sequence was the
same as 1: 4-O-methyl-GlcA(1-4)[GlcNAc(1-3)] Fuc(1-4)GlcNAc. Then it was
evident to deduce that glycan 2 was a penta-antennary complex N-glycan in
which three of them were assigned at Man-6 arm and the other two antenna
at Man-3 arm according to the specific D and E ion from (-) MSⁿ analyses
(See appendix for spectra).
Figure 2.8. (-) ESI MS$^2$ (→1005.9) spectrum of the native glycan 1. The reducing end GlcNAc is easily fragmented under negative MS$^0$ [Harvey 2005b]. The $^{2+4}A_7$ ion confirmed the core fucose position on C-6 rather than C-3 of the innermost GlcNAc. All the ions are negative four charges.

Figure 2.9. (-) ESI MS$^2$ (→1006.5) spectrum of the reduced and native glycan 1. Two diagnostic ions D at $m/z$ 903.3 (2') and $^{0+3}A_6$ 867.3 (2') clearly indicate four antennae are equally distributed into 3-Man and 6-Man arms. All charge state of each peak is annotated in parenthesis.
Figure 2.10. ESI MS$^2$ spectrum of C$_4$ ion ($\rightarrow$1006.5$\rightarrow$831.4, A) and MS$^3$ spectrum of C$_4$ ion from Man-6 pathway specifically ($\rightarrow$1006.5$\rightarrow$903.4$\rightarrow$831.4, B). The absence of distinguished E ion showed that two antennae attached to Man-6 arm on C-2 and C-6 rather than C-2 and C-4 as Man-3 arm. All peaks are negative single charge except ion E, which is a double charge fragment ion.
Figure 2.11. Characterized structures of Glycan 1-3
Figure 2.12. (+) ESI-MS² spectrum of permethylated glycan 3 (A) and deuterium methylated glycan 3 (B). Distinguishing antenna B ion was annotated in the figures, which include one sodium atom.

Figure 2.13. (+) ESI MS³ spectra of antenna B ion of glycan 3 (PM, 1879.2→1093.5, A) and (DM, 1944.0→1129.8, B). All ions contain one sodium atom.
Structural Characterization of Minor Glycan 3

Negative ion MS/MS spectrum of glycan 3 (M.W. 4667.6, Figure 2.1) revealed that glycan 3 has the same fragmentation pattern as glycan 1 except for 160 Da increasing in glycan 3 (See Appendix for spectra), suggesting an extra mono-methyl-O-Fucose moiety present in the antenna composition. This assumption was further supported by the presence of the proposed B ions from both permethylated and deuterium methylated derivates (Figure 2.12a and 2.12b). Further fragmentations on the B ion indicated the antenna sequence (Figure 2.13a and 2.13b). Ions m/z 834.5 and 905.5 in the permethylated MS³ spectrum suggested the presence of one terminal GlcNAc and one terminal Fuc residue (Fig 2.13a). The abundant ion m/z 936.5 came from the retro-Diels-Alder fragmentation of internal
GlcNAc, suggesting the internal Fuc1→4GlcNAc linkages. Similar fragment ions were observed in the deuterium methylated glycan (Fig 2.13b); in which distinguishing ions at $m/z$ 1097.8 and 947.6 indicated one naturally occurring methoxyl group locating at C-4 of GlcA same as glycan 1. These data suggested the mono-methyl-O-Fuc residue should be located at one of the following positions: C-2 of GlcA; or C-2 of core Fuc; or C-6 of internal GlcNAc. The abundant ion $m/z$ 818.5 derived from sigmatropic [1.7] hydrogen shift rearrangement in MS$^4$ spectrum of ion $m/z$ 936.6 (MS$^4$ pathway: 1872.9→1094.0→936.5, Fig 2.14a) showed that the extra mono-methyl-O-Fuc residue was attached to C-2 of GlcA. Moreover two small ions $m/z$ 429.2 and 447.3 further indicated that mono-methyl-O-Fuc group was attached to GlcA residue. This notion was also supported by another MS$^4$ fragment ion at $m/z$ 687.5 (1872.9→1094.0→905.5, Fig 2.14b), in which the terminal GlcA moiety was lost including the C-2 hydroxyl group from loss of mono-MeO-Fuc residue. It was noteworthy that the specific position of this mono-methoxyl group in the terminal Fucose was not defined due to the lack of diagnostic ions. Thus, the antenna sequence of glycan 3 was defined as mono-O-methyl-Fuc(1-2)-4-O-methyl-GlcA(1-4) [GlcNAc(1-3)]Fuc(1-4)GlcNAc. The branching information of glycan 3 was also obtained by exploiting (-) MS$^n$ spectra on native glycan (See Appendix for Spectra), which showed the same branching pattern as glycan 1. Therefore, glycan 3 was characterized as shown in Figure 2.11.
2.4 Discussion

New Overall MS-based Strategy

As demonstrated above, this new mass spectrometry-based strategy provided the opportunity to characterize the detail structure of $N$-glycans from the EP glycoprotein, except for the anomericity. Theoretically, the anomericity may be obtained through MS$^n$ FragLib comparisons [Hanneman 2006; Ashline 2005], such as identification of Gal|$\beta$ (1-4)Fuc modification on $C$. elegan core GlcNAc via synthesized samples [Hanneman 2006], but the availability of these pure state standard carbohydrate moieties were always limited. Accurate mass measurement was obtained on a common LTQ instrument using ultrazoom scan rather than the expensive LTQ-Orbitrap or LTQ-FTICR. The methylated derivates of monosaccharides were used to perform GC/MS experiments, which were allowed to identify GlcA residue with natural methoxyl modification.

Although (-) MS/MS potential use on $N$-glycan structural process has been summarized recently [Harvey 2008; Harvey 2005c], and has been employed to identify many biological glycans [Harvey 2009b]. But this is the first time to employ negative ion MS$^n$ (not MS/MS) to define the exact antenna location using distinguishing E ion [Harvey 2005c], which retains both C-2 and C-4 linkages of the branching mannose (Man-3 or Man-6). Since the D ion consisting of the Man-6 arm portion could be obtained specifically, further disassembly of the D ion may reveal the antennae
locations of Man-6 arm. Because the E ion was not found from the Man-6 arm of D ion pathway in glycan 1, the two antennae most probably connected to C-2 and C-6 of Man-6 rather than C-2 and C-4 as seen in Man-3 arm. The biological significance of the specific antenna location information has never been appreciated due to the lack of a general method to reveal this exact structural information. Since (-) MS^n could be used on either acidic glycans or neutral glycans with anionic ions adduct [Harvey 2008], this strategy appeared providing a general method to define the branching pattern of N-glycans. Furthermore it was believed that Fuc-3 modification on the innermost core GlcNAc will inhibit PNGase F to release N-glycan from the protein backbone, negative ion MS^n on native N-glycan provide a convenient and direct analytical method to differentiate Fuc-3 and Fuc-6 attachments (Figure 2.8). This technique may be particularly valuable when both of Fuc-3 and Fuc-6 modifications are present on the N-glycan core structure [Varki 1999] or even unusual modification on core structure [Hanneman 2006].

**Proposed Mechanisms to Interpret the Fragment Ions**

Several mechanisms were proposed to interpret the prominent fragmentation ions in the spectra. In scheme 2.1, a distinguishing daughter ion m/z 920.5 came from the CD$_3$ labeled antenna B ion m/z 952.6 (figure 2.4). This 32 Da neutral loss not only confirmed the presence of a natural methoxyl group in the antenna but also indicated its labile property under
positive ion CID fragmentation pathway due to the presences of many
deuterated methoxyl groups. Because the neutral loss of terminal mono-GlcA
moiety [M-190] has been observed in negative ion MS^n spectra, and the
presence of GlcA was confirmed by GC/MS, it will be readily deduced that the
natural methoxyl group must be located on the C-4 of the GlcA, thus it can
easily be explained that the elimination of C-4 methoxyl group will form a
conjugated double bond with the carboxylic group on C-6 in GlcA, which is
extraordinarily favored due to the low energy of the product comparing to
CD$_3$OH loss ($m/z$ 917.5, figure 2.4). In addition, the presence of this
conjugated intermediate is further confirmed by following disassembly of that
ion $m/z$ 920.5 (figure 2.4 inset), an abundant fragment ion $m/z$ 770.5 was
obtained after a neutral loss of 150 Da. That fragment ion $m/z$ 770.5 can be
rationalized via a retro Diels-Alder mechanism (Scheme 2.1). It will be worthy
to point out that retro Diels-Alder fragmentation are the most abundant and
common cross-ring cleavage ions in CID MS^n spectra on carbohydrate
[Ashline 2005], which is diagnostic for most inter-residue linkage. During
this case, deuterated methylated derivatives can provide particular
information to identify natural methoxyl group position under positive ion MS^n
disassembly.
Scheme 2.4. Commonly neutral loss of 42 Da from HexNAc 1,2-ene parent ion due to sigmatropic [1.5] hydrogen rearrangement fragmentation

In schemes 2.2 and 2.3, sigmatropic [1.7] and [1.5] hydrogen rearrangement are proposed to explain the formation of ions m/z 665.4 (figure 2.5a) and m/z 378.2 (figure 2.5b), respectively. Although little was known how the sodium cations are involved in these fragmentation pathways, considering the abundance of these daughter ions, it may be speculated that sodium cation will decrease the activation energy compared to the normal thermal sigmatropic rearrangements. Moreover, [1.5] sigmatropic hydrogen rearrangement have been widely observed in many standard oligosaccharides fragmentation, even if systematic investigation on these mechanisms has never been reported. A better known example is the common 42 Da neutral loss from HexNAc moiety (Scheme 2.4). [1.7] sigmatropic hydrogen rearrangement is less common but also observed on pyronaridine’s fragmentation in positive mode low-energy CID ion trap mass spectrometry [Lee 2004].
Structural Features of these Novel N-glycans

The most obvious structural features of the EP N-glycans are both novel antenna sequences: 4-O-methyl-GlcA(1-4)[GlcNAc(1-3)]
Fuc(1-4)GlcNAc and mono-O-methyl-Fuc(1-2)-4-O-methyl-GlcA
(1-4)[GlcNAc(1-3)]Fuc(1-4)GlcNAc, which have never been reported previously. A similar tetrasaccharides moiety was found in a minor
glycosphingolipid 4-O-methyl-GlcA(β1-4)[3-O-methyl-GalNAc(α1-3)]
Fuc(α1-4)GlcNAc(β1-2)Man(α1-3)[Xyl(β1-2)]Man(β1-4)Glcβ1-ceramide
isolated from the spermatozoa of fresh water bivalve, *Hyriopsis schlegelli*
[Scarsdale 1986; Hori 1983], suggesting the presence of a unique glycan biosynthetic pathway leading to the branching fucosylation in the bivalve class. This notion was also supported by a recent report in which internal fucose branching in N-glycan antenna was found in the hemocyanin of
*Rapana thomasiana* [Gielens 2005; Idakieva 2004]. The antenna sequence of *Rapana thomasiana* was also a tetrasaccharides moiety starting from the non-reducing end: 3-O-Methyl-Gal-[GalNAc]-Fuc-GlcNAc, which was similar to the antenna sequence of the major glycan 1. Fucosylation is commonly regarded as a terminal modification like sialic acid, which will prevent any further extension to the sugar chain, like a "Cap" [Varki 1999]. Since no other internal fucosylation has been reported so far according to our knowledge, it is not clear whether this is a unique glycosylation pattern in the bivalve class, or not. The key point is that these cases demonstrate the complexity of
natural glycosylation, which are not always synthesized under strict regulation according to the known glycosylation pathway. Therefore, a mass spectrometry-based de novo strategy is indeed needed to define glycan structures, especially for those less-studied organisms.

Another structural feature of EP glycans is its naturally occurred methoxyl modification. Natural methoxyl modification has been found widely present in many low-level animals [Gutternigg 2004; Puanglarp 1995; Khoo 1991; Hall 1977]. Unfortunately, the biological functions of these natural methoxyl modifications have not been fully understood so far. Since methoxyl group is modified usually on the non-reducing end terminal residues (GlcA and Fuc), it may suggest this modification may be well-controlled for its biological functions, such as signal transformation, receptor recognition, etc.

**Implications for Its Metal Ion binding Ability**

The fascinating property of the EP protein is its strong affinity with many metal ions [Hattan 2001]. It is evident to deduce these multiple antennary N-glycans have metal ions binding ability due to its flexible arms and terminal glucuronic acid residues. Since the majority of N-glycans contains 4- or 5 antennae in EP glycoprotein, it also suggests that the binding of the metal ions may need cooperation of two or three antennae, and functions as a chelation agent. According to these identified N-glycan structures, the EP glycoprotein can keep the metal ion binding ability even if it
was denatured since its $\text{N}$-glycans may take part in metal ion binding directly [Nair 1999].

**Summary**

In a word, we have characterized novel $\text{N}$-glycosylation structures from EP glycoproteins isolated from mollusk: *Mytilus edulis* via a new mass spectrometry strategy in which combined accurate mass measurement, GC/MS, chemical modification, positive ion MS$^n$, and negative ion MS$^n$ techniques. The unique antenna sequence, 4-$\text{O}$-methyl-$\text{GlcA}$(-4)[GlcNAc(-3)] $\text{Fuc}$(-4)GlcNAc, and multiple antennae may provide the structural basis of its binding ability to various metal ions.
CHAPTER 3

N-Glycosylation Analysis of Neuraminidase from the 1918 Pandemic Influenza Virus H1N1

3.1 Introduction

Introduction to Influenza Virus

Influenza viruses are highly contagious and may cause severe respiratory illness and death in humans. There are three viral genera within the influenza family: A, B and C, distinguished on the basis of their antigenic differences between their matrix proteins and nucleoproteins [Lamb 2001; Wright 2001]. Influenza A, B and C viruses also differ with respect to host range, variability of surface glycoproteins, genome organization, and morphology [Lamb 2001]. Of the three types of influenza virus, type A infects a wide range of avian and mammalian species and is responsible for all major pandemic outbreaks of influenza, and for most of the well-known annual “flu” epidemics [Wright 2001]. Both influenza B and C also infect humans but only lead to mild common-cold-like symptoms [Wright 2001] and recovery may be expected in a few days without any medical treatment. Therefore, influenza A virus is the primary research focus, and gains the most clinical concern.
Influenza A virus can be further classified into subtypes according to the serological reactivity of its surface glycoprotein antigens: hemagglutinin (HA) and neuraminidase (NA) [Lamb 2001]. Currently, sixteen subtypes of HA (H1 to H16) and nine of NA (N1 to N9) have been identified in avian and mammals species (WHO, Table 3.1). All subtypes have been found in avian species, routine tests of wild birds always find some influenza A viruses present, especially in migrating aquatic wild birds. It is believed that all influenza viruses originated from avian species, which are usually referred as “avian influenza”. The aquatic birds serve as a reservoir for all influenza A viruses, providing the likelihood of trans-infection to domestic birds [Hinshaw 1980], and mammals. Surprisingly, the vast majority of these viruses are harmless to their wild hosts. Only limited subtypes have been identified in human and other mammals (Table 3.1). As for human influenza viruses, the subtypes that account for the three pandemics of the last century were the 1918 H1N1 “Spanish flu”; the 1957 H2N2 “Asian flu”; and the 1968 H3N2 “Hong Kong flu” [Wilschut 2005]. Several other influenza subtypes: H5N1, H7N7 and H9N2, have also been reported to infect human sporadically since 1997 and cause high death rate [World Health Organization Website]. Since these viruses have not been transmitted in the human population and do not carry the human viruses’ characteristics, they are generally regarding as avian viruses rather than human viruses. Besides human, the influenza virus
also infects many other mammals such as: pigs, horses, seals, whales, etc. (Table 3.1) [Wilschut 2005; Lamb 2001].

The current nomenclature system for the influenza A virus include the host of origin, geographic locations of the first isolation, strain number and year of isolation [WHO memorandum 1980]. The HA and NA subtypes are usually specified in parentheses, such as: A/Swine/Iowa/15/30 (H1N1). By convention, the host origin of human strains is omitted, such as: A/Vietnam/1203/04 (H5N1), and A/Puerto Rico/8/34 (H1N1).

**Major Structural Components of Influenza A Virus**

The influenza virus has a roughly spherical shape, and its size ranges between 80-120nm in diameter. The capsid is surrounded by a lipid envelope, from which projects two major surface glycoproteins, HA and NA. Within the capsid resides the genome which is divided into eight ribonucleoprotein (RNP) segments. The genetic material is a single strain (-) sense RNA (complementary to mRNA) and associated with three polymerase polypeptides [Duesberg 1969]. Figure 3.1 is a model of the overall structure of influenza virus. The virus has approximate 500 spike-like projections on the surface. These represent the distinguished HA (cone shape) and NA (mushroom shape) glycoproteins [Wright 2001; Steinhauer 1998; Colman 1998].
Table 3.1. Natural hosts of influenza A viruses (adapter from Wilschut 2005)

<table>
<thead>
<tr>
<th>Haemagglutinin</th>
<th>Predominant Hosts</th>
<th>Neuraminidase</th>
<th>Predominant Hosts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designation</td>
<td></td>
<td>Designation</td>
<td></td>
</tr>
<tr>
<td>H1</td>
<td>Human, pig, birds</td>
<td>N1</td>
<td>Human, pig, birds</td>
</tr>
<tr>
<td>H2</td>
<td>Human, pig, birds</td>
<td>N2</td>
<td>Human, pig, birds</td>
</tr>
<tr>
<td>H3</td>
<td>Human, pig, birds, horse</td>
<td>N3</td>
<td>Birds</td>
</tr>
<tr>
<td>H4</td>
<td>Birds</td>
<td>N4</td>
<td>Birds</td>
</tr>
<tr>
<td>H5</td>
<td>Birds, (human)</td>
<td>N5</td>
<td>Birds</td>
</tr>
<tr>
<td>H6</td>
<td>Birds</td>
<td>N6</td>
<td>Birds</td>
</tr>
<tr>
<td>H7</td>
<td>Birds, horse, (human)</td>
<td>N7</td>
<td>Birds, horse</td>
</tr>
<tr>
<td>H8</td>
<td>Birds</td>
<td>N8</td>
<td>Birds, horse</td>
</tr>
<tr>
<td>H9</td>
<td>Birds, (human)</td>
<td>N9</td>
<td>Birds</td>
</tr>
<tr>
<td>H10</td>
<td>Birds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H11</td>
<td>Birds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H12</td>
<td>Birds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H13</td>
<td>Birds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H14</td>
<td>Birds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H15</td>
<td>Birds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H16</td>
<td>Birds</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Haemagglutinin (HA) is the major envelope glycoprotein, usually four times abundant than neuraminidase (NA). The primary function of HA is to recognize and attach onto the glycoconjugates that terminate with a sialic acid residue. Such structures are components of glycoproteins and glycolipids embedded in the host cell surface, and appear to be the essential step for the viral entry and infection [Gamblin 2004; Colman 1998]. It is believed that the major difference between avian virus and human virus is their HA binding specificity. Generally human virus recognizes sialic acid $\alpha$ (2,6) galactose residue [SA$\alpha$(2,6)Gal], which are believed to be dominant in the human respiratory track epithelium cells; whereas avian influenza virus prefer SA $\alpha$(2,3) Gal residue, an abundant epitope on the surface of the
avian intestinal tract [Yamada 2006; Stevens 2006a; Stevens 2006b; Weis 1988].

HA is synthesized in the endoplasmic reticulum as a single polypeptide chain (HA0, approximately 560 amino acids). The mature HA is assembled as a trimer consisting of three identical monomers before being exported to the host cell surface. On the cell surface, HA0 is cleaved by trypsin-like proteases secreted by the host cell into two disulphide bond connected subunits, HA1 and HA2. This host specific digestion is essential for HA to mediate membrane fusion to the host cell [Skehel 2000; Bullough 1994; Kim Carr 1993; Wilson 1981]. For this reason, most human influenza A viruses are limited to infect the human respiratory tract where there is a distribution of relevant proteolytic enzymes [Steinhauer 1999; Klenk 1994]. As for the high pathogenic avian influenza viruses, such as H5N1 and H7N2, their HAs can be cleaved by intercellular proteases, in a process that occurs just before HA0 reaches the cell surface of the infect host cell, due to the presence of multiple basic amino acids located at the cleavage site in these HAs [Kawaoka 1988; Webster 1987]. Because these proteases are ubiquitous in all cell types, viruses with HA containing these multiple basic amino acids in the cleavage site could easily spread throughout the body, and usually lead to a fatal systemic infection.

Another surface glycoprotein of virus is NA, a mushroom-shape tetrameric glycoprotein (Figure 3.1). Each monomer usually contains around
450-500 amino acid residues depending on the individual virus strain, and is subjected to oligomerize before reaching the host cell surface. All nine subtypes of NA are tetrameric and share a common structure consisting of a globular head, a thin stalk region, and a small hydrophobic region that anchors the proteins in the virus membrane (Figure 3.1) [Colman 1998]. The principal biological function of NA is to catalyze the cleavage of the terminal sialic acid residue from various glycocojugates such as glycoproteins and glycolipids on the host cell surface, which are virus HA binding receptors. NA is important during the final stages of influenza virus infection because it removes the sialic acid from infected cell surfaces which are attached by newly formed virions, thus facilitating progeny viruses release and spread of the infection to the neighboring cells [Bucher 1975; Palese 1974]. The active site of NA consists of a pocket on the surface of each subunit formed by fifteen charged amino acids. These amino acids are highly conserved in all viruses [Colman 1983; Varghese 1983], which makes NA the perfect target for the antiviral drug [Itzstein 1993]. Two current anti-flu drugs oseltamivir (Tamiflu) and zanamivir (Relenza) are transitional state analogues of sialic acid-galactose, which efficiently inhibit the enzymatic activity of NA, thus preventing the spread of infection. Furthermore NA may also contribute to the removal of the decoy sialic acid receptors on mucins, cilia, and the cellular glycocalyx in the human airway, which will promote the penetration of the virus through the ciliated epithelium cells [Matrosovich 2004a].
The virus envelop also contains a small number of copies of integral membrane protein, M2, which forms a tetramer with ion channel activity [Zebedee 1988; Lamb 1985]. M2 is believed to modulate the pH within virions, weakening the interaction between the viral ribonucleoproteins (RNPs) and M1 protein. M2 is the target for the anti-influenza drugs amantadine and rimantadine [Hay 1992].

The viral core contains eight strands of negative-sense single stranded RNA [Lamb 2001]. Each RNA segment is connected with multiple copies of nucleoproteins (NP), and with the viral transcriptase consisting of RNA polymerase components PB1, PB2 and PA, thus forming the ribonucleoproteins (RNP) complex. The RNPs are surrounded by a layer of the matrix protein: M1. M1 is the most abundant structural protein within the virus, approximately 3000 copies per virion [Lamb 2001].

Because of its simple biological structure, the influenza virus is constantly evolving by two mechanisms: antigenic drift and antigenic shift [Bouvier 2008; Shoham 2006]. Antigenic drift is induced by the selection pressure of the host immune system which accounts for the most annual epidemics. Antigenic drift occurs as a result of continuous mutation of the RNA genome of the virus because of the errors made during the RNA replication process. Therefore, amino acid changes in the viral proteins arise, including substitutions in the antigenic area of HA and NA. Virus variants with a selective advantage are remained by the pre-existing immunity of the host.
Another pathway to form new viruses is "antigenic shift" also called as "reassortment", which occurs in the same host infected by two or more different types of influenza viruses at the same time. New viruses may emerge by exchanging some portion of their RNA fragments with others, which accounted for the pandemic influenza outbreaks in 1957 and 1968 [Kawaoka 1989; Scholtissek 1978]. Since pigs can be infected by both avian and human viruses, they have been considered as "transmission vessels" between avian and human species for a long time [Ludwig 1995; Scholtissek 1994]. Currently high pathogenic avian H5N1 and H7N7 viruses have been reported to infect humans directly without any intermediate vessel. This emphasizes the possibility that new pandemic virus strain could emerge because the reassortment can take place directly in human. Fortunately, these high pathogenic avian viruses have not obtained the capability for efficient human-to-human transmission. It was believed that avian virus could not infect the human upper respiratory tract due to the lack of appropriate SA-Gal epitope, which will be discussed in detail in Chapter 4.

The Unique 1918 Influenza Virus (H1N1) “Spanish Flu”

The 1918 influenza "Spanish Flu" was the most devastating pandemic of any infectious disease in the recorded history. It spread globally, and infected almost 25 to 30% of the world's population. It was estimated that "Spanish Flu" killed approximately 40 million people worldwide, including more than 600,000 deaths in the United States. In contrast, both 1957 "Asian
flu (H2N2) and 1968 “Hong Kong flu” (H3N2) were less severe and caused less than 80,000 deaths for each pandemic in the United States [Wilschut 2005]. Unlike the 1957 and 1968 viruses, the causative agent of the 1918 pandemic was not known at that time, and no original 1918 virus samples were kept intact. Our understanding to this virus has relied on the historical accounts and the clues from the current viruses. Fortunately Dr. Taubenberger and his co-workers have been able to obtain all eight segment RNA sequences of the 1918 virus from an Alaskan Inuit woman who was interred in permafrost in Brevig Mission, Alaska. Thus the strain of this virus was designed as A/Brevig Mission/1/18 (H1N1) [Reid 2002; Basler 2001; Reid 2000; Reid 1999; Taubenberger 1997], and has provided the opportunity to research this notorious virus for the first time and further improve our understanding to this pandemic. Such studies may help us to prepare for any potential influenza pandemics in the future, such as current Mexico H1N1 flu.

Besides its severity, the 1918 influenza flu had several unique features, distinguishing it from other pandemic and epidemic viruses. The first was its mysterious origin. Prior to 1918, no highly virulent viruses were recorded to circulate in humans. It appeared that this virus suddenly emerged from several locations in the world mysteriously [Taubenberger 2006]. According to phylogenetic analyses of its HA and NA sequence, this virus most likely originated from an avian species, but may have circulated in humans for a
short time before the pandemic [Reid 1999; Reid 2000]. Thus, it kept most of
the avian characteristics but acquired enough changes to spread and
replicate efficiently in humans. This notion was indeed supported by recent
HA glycoarray binding experiments: its HA can bind both SA-Gal $\alpha(2,3)$ and
$\alpha(2,6)$ epitopes [Stevens 2006b; Gamblin 2004].

The second feature of the 1918 virus is its exceptional virulence.
Unfortunately neither 1918 HA nor 1918 NA genes has obvious genetic
characteristic that can account for such virulence [Taubenberger 2006].
Recent highly-pathologic H5 and H7 viruses have been found able to cause
systemic infection due to the presence of multiple basic amino acid sites for
HA1 and HA2 cleavage, which could be activated by ubiquitous proteases in
cell [Kawaoka 1988; Webster 1987]. However, this kind of feature hasn't
been observed in the 1918 HA sequence [Reid 1999]. Another genetic
mutation has been reported in the NA sequence (N146R or N146Y, leading
to the loss of a glycosylation site), appearing to allow the virus to replicate in
many tissues beyond the respiratory tract in a mouse model [Li 1993].
However, this mutation was not found in the 1918 NA sequence [Reid 2000].
This information is in agreement with the historical medical records that
reported the 1918 virus didn't replicate outside of the respiratory system
[Taubenberger 2006]. Thus, the factors relating to its far-reaching virulence
of the 1918 virus has caught the attention of many virologists.
Another fascinating feature of 1918 virus is its high mortality rate among young adult populations, especially for the age group (25-34), who are usually not affected seriously by common influenza viruses. As shown in Figure 3.2, the age group (25-34) showed a dramatic high death rate comparing to other age groups: (5-14) and (44-54). An explanation for this has been proposed that there may have some sort of precursor of the 1918 H1N1 viruses circulating before 1889 providing the limited immunity. Therefore, people born before 1889 may have had partial immunologic protection [Taubenberger 2006]. This theory may partially explain the high morbidity rate in young age groups, but certainly it does not address why the age group (25-34) has a higher death rate than the age group (5-14) because both groups were immunologic naïve to the virus. It could be speculated that there must have some unique biological features associated with age group (25-34) to facilitate the virus infection.
3.2 Research Background and Purpose

The Neuraminidase of the 1918 Virus and its Glycosylation

Influenza virus NA is the second most abundant surface glycoprotein and essential for the virus infection. As discussed above, the primary function of NA is to hydrolyze the terminal sialic acid residue from the host cell surface, thus enabling the progeny viruses to spread and infect other cells and tissues [Colman 1998]. Furthermore, NA is also the major antigenic component besides HA. Host immunologic selection drives the constant change of the NA sequence [Johansson 2007]. Therefore, an investigation of NA structure...
is one of the key steps to understand influenza virus infection. As for 1918 influenza virus, it has been demonstrated that HA, NA and PB1 are related to its exceptional virulence by single gene reassortants experiments [Pappas 2008].

The 1918 NA sequence has been successfully recovered from an Alaska woman (A/Brevig Mission/1/18, H1N1) [Reid 2000]. The 1918 NA contains 469 amino acid residues and 7 potential N-glycosylation sites: Asn50, Asn58, Asn63, Asn68, Asn88, Asn146, and Asn234 (Figure 3.3). The active NA protein is configured as a homotetramer containing a large box-shape globular head, a thin stalk region, and a short hydrophobic tail buried in the virus membrane [Xu 2008]. The active site is located on the membrane-distal surface of the globular head, close to the pseudo axis through the center of each monomer, which is very well conserved in all subtypes of NA [Xu 2008]. Three potential N-glycosylation sites are in the ectodomain of NA: Asn88, Asn146, and Asn234. Both Asn88 and Asn234 are located on the inner surface of the molecule, whereas Asn146 is on the membrane-distal face, close to the active site [Xu 2008]. Asn146 is highly conserved in all N1 subtype NA including both avian and human viruses. It was found that mutation of this N-glycosylation site may lead to influenza virus spreading to the other organs in the mouse model, which is associated with virus's virulence [Li 1993]. But it is not sure whether this is the case in other hosts systems.
The stalk region is usually made up of approximately 50 amino acid residues starting from the 40th amino acid counting from the N terminus of NA. The stalk region of 1918 NA has one cysteine (Cys49) and four potential N-glycosylation sites (Asn50, Asn58, Asn63, and Asn68). It was found that under selection pressure, HA may acquire new N-glycosylation sites (up to 7 extra sites) in human strains [Matrosovich 1999; Inkster 1993; Deshpande 1987], which is believed to benefit the viruses masking the antigenic area. Unlike HA, the four glycosylation sites in the stalk region of NA are highly conserved in all H1N1 viruses regardless of human or avian strains (the exact sites may vary by strains), implicating some important functions which are not clear currently.
1918 NA amino acid sequence

<table>
<thead>
<tr>
<th>1</th>
<th>mnpnqiiki</th>
<th>gisicmvqvii</th>
<th>sliqignii sivvshsiqt</th>
<th>gnqnhpetcn</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>qsiityennt</td>
<td>wvnqtyvnis</td>
<td>ntvvagqda tsviltgnss</td>
<td>lcpisgwaiy</td>
</tr>
<tr>
<td>101</td>
<td>skdngirigs</td>
<td>kydvvirep</td>
<td>fiscshlecr tffltqgall</td>
<td>ndkhsngtvk</td>
</tr>
<tr>
<td>151</td>
<td>drcpyrllms</td>
<td>cpvgeapspy</td>
<td>nserfesavs asachdgmwv</td>
<td>ltigisgpdn</td>
</tr>
<tr>
<td>201</td>
<td>gavavklyng</td>
<td>iidtikswr</td>
<td>nnilrtqese cacvagseft</td>
<td>imtdgpsngq</td>
</tr>
<tr>
<td>251</td>
<td>asykilkieq</td>
<td>gktksieln</td>
<td>apnyhyecs cypdtkgvmc vcrdnwhqsn</td>
<td></td>
</tr>
<tr>
<td>301</td>
<td>rpwwsfdqnl</td>
<td>dyqigyicsg</td>
<td>vfgdnprpad tgtgscgpvss ngangikgfs</td>
<td></td>
</tr>
<tr>
<td>351</td>
<td>Frydngowig</td>
<td>rtkstssrsge</td>
<td>femiwpdgw tetdssfsrv qdivaitdws</td>
<td></td>
</tr>
<tr>
<td>401</td>
<td>gygsfvyqhp</td>
<td>eltlgclcmrp</td>
<td>cfwvelirgg pkentiwtsg ssisfcgovns</td>
<td></td>
</tr>
<tr>
<td>451</td>
<td>dtvgwswpdg</td>
<td>aelpfsidk</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tryptic peptide stalk region:

| IITIGSICMVQVIIISLILQIGNIIISIVVSHSIQTGNQNHPETCNQSIITYENNTVWNQT |
| YVNZSNTVVAGQDATSVILTGNSSLCPISGWAIYSK |

Figure 3.3. The complete amino acid sequence of 1918 NA is listed in the top panel. It contains 469 AA residues and seven potential N-glycosylation sites highlighted as red. The stalk region of the tryptic large peptides includes five potential N-glycosylation sites and is listed in the bottom panel.

Recently 1918 tetrameric NA protein was reported to have different glycosylation pattern compared to the corresponding monomer and dimer forms [Wu 2009]. In brief, 1918 NA DNA sequence was chemically synthesized and inserted into a baculovirus genome system. Subsequently the recombinant virus was grown in the Sl21 insect cell line, and the crude NA proteins was harvested and purified. Interestingly the active tetrameric NA only accounts for 10-15% of total product NA and the rest is the inactive monomer and dimer forms (Figure 3.4). It was found that tetrameric NA has a stable structure and will not disassociate into monomer and dimer, whereas
the dimer and monomer NA can mutually convert to each other in the solution but not tetramer (Figure 3. 4). Further analysis has shown that the tetrameric NA may have a distinct N-glycosylation pattern compared to other forms. Interestingly de-N-glycosylation of tetramer will not affect NA enzyme activity; suggesting N-glycosylations is not close to the NA active site, and not essential for keeping the tetrameric three-dimensional structure. Therefore, it can be concluded that N-glycosylation is involved in the formation of the active NA tetrameric structure. And once the mature tetramer has been formed, it will be able to keep a relatively stabilized shape resistant to proteolytic digestion under native condition. Given the insect cell lines are known to produce some simple high mannose N-glycans, which may be different with the human cells glycans [Kost 2005]. A systematic analysis of N-glycans from various oligomer of NA will be the key to understand the difference between the active NA and the inactive NA. The specific purposes of this project are defined as following:

- Detail N-glycan structural analyses of dimer and tetramer NAs by mass spectrometry, especially MS^n
- Comparison of N-glycan structures between dimer and tetramer NAs and identify the difference, especially those located in the stalk region
- According to the difference, find the connection to the real virus host systems such as human cells.
Figure 3.4. The expressed 1918 NA exists in three forms: tetramer (t), dimer (d), and monomer (m). Chromatography of these NA forms was separated on a superdex-200 gel filtration column (top panel). Re-chromatography of the separated forms after 2 weeks storage in the solution (lower panel), showed no inter-conversion between tetramer with dimer/monomer [Adapted from Wu 2009].

### 3.3 Experimental Procedure

**Materials and Reagents**

Purified dimer and tetramer NA was provided by Dr. Wu (R&D System Inc. Minneapolis, MN 55413). Procedures to obtain these materials were described in the reference [Wu 2009]. The glycerol-free PNGase F and associated reagents (10 x denaturing buffer, NP-40, G7 buffer) were purchased from New England Biolabs (Beverly, MA). Proteomic grade trypsin (25 μg per vial), dimethyl sulfoxide, sodium hydroxide, iodomethane, sodium borohydride, sodium borodeuteride, dithiothreitol, Urea, iodoacetamide, trifluoroacetic acid, tetrahydrofuran, toluene and ethanol were obtained from
Sigma-Aldrich (St. Louis, MO). 7000 MWCO membrane dialysis cassette was purchased from Pierce (Prod# 66370, Rockfold, IL). 10,000 MWCO microcon centrifugal devices were obtained from Millipore (Ultrace YM-10, Bedford, MA).

**Procedure of Trypsin Digestion**

NA proteins samples (dimer and tetramer) were received in a buffer solution (20 mM Tris and 0.15 M NaCl). The salts were removed by membrane dialysis against 5 mM NH₄HCO₃/0.01% SDS overnight using 7000 MWCO (Pierce, Rockfold, IL). The protein solution was transferred to a 1.5 mL Eppendorf vial and dried under vacuum. The dried proteins were dissolved in 50 µL 8M urea / 0.4M NH₄HCO₃ and reduced by the addition of 5 µL of 45 mM DTT (dithiothreitol) and incubated at 50°C for 45 mins. The protein was subsequently alkylated with freshly prepared 2 µL of 100 mM IAA (iodoacetamide) at room temperature for 30 mins. The trypsin was activated by 1 mM HCl solution and added into the reaction solution at weight ratio (1/25: trypsin/protein). The reaction solution was adjusted to a final concentration of 2M urea/0.1M NH₄HCO₃ and incubated at 37 °C overnight (11-15 hours). The reaction was stopped by addition of several drops of trifluoroacetic acid in an ice bath and then dried under vacuum. The salts were removed by passage through a C₁₈-SPE cartridge (Strata-X 33um 30mg/ML, Phenomenex, Torrance, CA) with 2% ACN/H₂O containing 0.1% TFA. The digested peptides were eluted with 70% ACN/H₂O containing 0.1%
TFA. In selected cases the stalk region glycopeptide was separated with other low molecular weight mixture by dialysis against 10, 000 MWCO centrifugal device (Millipore, Bedford, MA). Briefly, the tryptic peptides mixture was diluted to 500 μL water solution and applied onto the 10, 000 centrifugal device. The rotor speed was set at 11, 000 r/minutes as recommended and centrifuged for 25 mins. The membrane retention fraction was collected and subjected to de-glycosylation procedure.

**N-glycan Release by PNGase F**

The total tryptic peptide mixtures, (or the stalk region peptide), were dried and re-dissolved in 5 μL stock 10X G7 buffer (500 mM sodium phosphate) and 45 μL H₂O. To this was added 1 μL of PNGase F stock solution and incubated at 37 °C overnight. The free N-glycans were separated from peptides by a hand-packed graphitized carbon cartridge (Alltech, Deerfield, IL). The graphitized cartridge was pre-washed thoroughly with water, and the reaction mixture applied. The buffer and salt were washed through with excess H₂O, and the glycans were eluted with 5-6 volumes of 25% ACN/H₂O containing 0.1% TFA.

**Reduction and Permethylation of N-glycans**

The eluted glycans were dried and reduced by the addition of 200 μL sodium borohydride solution (10 mg/mL in 0.01M NaOH) and allowed to stand at room temperature overnight. The reaction was quenched by the
addition of several drops of acetic acid in the ice bath, followed by addition of
3 mL pure ethanol, and dried under vacuum. Excess borate was removed as
its methyl ester by the repeated addition of 3 mL 1% acetic acid in methanol
and drying under vacuum until all residue salts were gone. This is followed by
the addition of 3 mL toluene and drying under N₂ gas, repeating twice. Then
the glycans were further desalted either by passage through a cation
exchange resin (Dowex 50w, Sigma, St. Louis, MO) or graphitized cartridge.

The dried glycan sample was dissolved in 300 μL DMSO in a glass
tube, following a small amount ground NaOH powder and vortexing for 5
mins. Iodomethane (50 μL) was then introduced into the tube, followed by
vortexing for 40-50 mins, to form a white cloudy suspension solution. The
reaction was stopped by the addition of 1 mL dichloromethane and 1 mL of
5% acetic acid, sequentially, in ice bath. The methylated glycans were
isolated by liquid-liquid (H₂O/CH₂Cl₂) extraction, repeated 3 times. The
combined organic layers were then back washed 5-6 times with water. The
dichloromethane was evaporated under N₂ gas and ready for mass
spectrometry analysis.

**Radioactivity of Fuc Incorporation onto Dimer and Tetramer *in Vitro***

1 μg of substrate (dimer or tetramer NA) was mixed with 2 μl of
GDP-³H fucose (1.2 μM from Perkin Elmer), 0.5 μg of recombinant FUT8,
and diluted with buffer (10 mM Tris, 10 mM MnCl₂, pH 7.0) to a total of 20 μl.
This solution was incubated at room temperature for 10 minutes. The
incorporation of $^3$H fucose onto the dimer and tetramer neuraminidase was then measured using a filter binding assay.

**Filter Binding Assay**

After $^3$H-fucose incorporation with FUT8, 8 μl reaction sample from each reaction was spotted on to glass microfiber filter (Cat# 09-874-34; Fisher Scientific, Kansas City, MO). To this a drop of 75% ethanol was added to cover and denature the protein. The filters were dried and then thoroughly washed in 200 ml of water in a shaker for 5 minutes. Washing was repeated 3 times. The filters were countered with a LS 6500 scintillation counter in Ready Protein™ cocktail (Beckman Coulter, Fullerton, CA).

**Tetramer NA Enzyme Activity Following In Vitro Fucosylation**

The NA enzyme was first diluted with a assay buffer (50 mM Tris, 5 mM CaCl$_2$ and 200 mM NaCl at pH 7.5). To start the reaction, in a 96-well fluorescent plate, 50 μL of the diluted enzyme was added to 50 μL of 400 μM substrate 2'-(4-methylumbelliferyl)-α-D-N-acetyl neuraminic acid (4-MU-NANA, Sigma-Aldrich, St. Louis, MO). The reaction was monitored with a fluorescence spectrometer at an excitation of 365 nm and emission of 445 nm in a kinetic mode. The detection was based on the fluorogenic nature of the reaction product 4-methylumbelliferone, and the measurement revealed a reaction velocity in terms of relative fluorescent units (RFU) generated per unit time. To calculate the actual reaction velocity, a
conversion factor between an RFU and the amount of product
4-methylumbelliferone (Sigma-Aldrich, St. Louis, MO) was determined in the
assay buffer.

**Mass spectrometry**

Native glycan MS profiles were performed on a MALDI-TOF
instrument (MALDI-CFR, Shimadzu, Kyoto, Japan). 1 μL matrix solution (10
mg DHB in 1 mL of 50% acetonitrile solution containing 0.1% TFA) was
applied on MALDI plate and air dried. Then 1 μL sample solution was applied
to the well and mixed evenly with the matrix on the plate. When the solution
was dry, a drop of pure ethanol was applied onto the well and the sample and
matrix re-crystallized. MS^n analyses were performed on methylated-reduced
glycans using an ESI-LTQ instrument (Thermo, Waltham, MA).

### 3.4 Results

**MALDI-MS Profiles of Dimer and Tetramer N-glycans**

To enhance the release efficiency of PNGase F, both dimer and
tetramer NAs were subjected to trypsin digestion prior to de-glycosylation.
Aliquot of N-glycans were reduced and per-methylated in preparation for the
MS^n analyses [Ashline 2005]. MALDI-MS profiles of total native N-glycans
from both forms were shown in Figure 3.5. Both forms have relatively simple
high mannose type glycoforms. The abundant peak in tetramer was m/z
935.2, corresponding to the pentasaccharide core structure of the N-glycan: \(\text{Man}_3\text{GlcNAcGlcNAc}\) (Figure 3.5a). The following abundant peak was \(m/z\) 1081.4, had an identical composition with an additional fucose: \(\text{FucMan}_3\text{GlcNAcGlcNAc}\). All other peaks appear to be a series of low abundant mannose type glycans, \(m/z\) 773.3 (\(\text{Man}_2\text{GlcNAcGlcNAc}\)), 1097.6 (\(\text{Man}_4\text{GlcNAcGlcNAc}\)), 1259.7 (\(\text{Man}_5\text{GlcNAcGlcNAc}\)), 1421.9 (\(\text{Man}_6\text{GlcNAcGlcNAc}\)), 1584.2 (\(\text{Man}_7\text{GlcNAcGlcNAc}\)), and 1746.5 (\(\text{Man}_8\text{GlcNAcGlcNAc}\)) (Figure 3.5a and 3.5c). No hybrid and complex type N-glycans were observed. In contrast to the tetramer, the most abundant peak in the dimer profile was \(m/z\) 1081.4 (\(\text{FucMan}_3\text{GlcNAcGlcNAc}\)) followed by \(m/z\) 935.3 (\(\text{Man}_3\text{GlcNAcGlcNAc}\)) (Figure 3.5b). Three other low abundant fucosylated glycans were also present at \(m/z\) 919.2 (\(\text{FucMan}_2\text{GlcNAcGlcNAc}\)), \(m/z\) 1243.9 (\(\text{FucMan}_4\text{GlcNAcGlcNAc}\)), and \(m/z\) 1284.8 (\(\text{FucMan}_5\text{GlcNAcGlcNAc}\)) (Figures 3.5b and 3.5d), which was absent in the tetramer. Furthermore, two low abundance high-mannose type glycans \(\text{Man}_7\text{GlcNAcGlcNAc}\) and \(\text{Man}_8\text{GlcNAcGlcNAc}\) were present in the tetramer form, but absent in the dimer.

**Excess Fucosylated Glycans in Dimer NA**

The relative abundance ratio of each composition was calculated according to their MALDI-MS peak intensities (Table 3.2) [Harvey 1999]. These calculations clearly showed that fucosylated structures account for 73.6% of total glycans in the dimer, compared to 18.1% in the tetramer. In
order to ascertain the structural difference between both forms, a complete structural analysis of all compositions was undertaken by ESI-MS on permethylated glycans.

Table 3.2. The relative ratio of each composition of both dimer and tetramer NA were obtained from MALDI-MS spectra. (F: Fuc; H: Man; N: GlcNAc). The ratio of fucosylated glycans are summed in the last row.

<table>
<thead>
<tr>
<th>Native (m/z)</th>
<th>Composition</th>
<th>Dimer Signal (I)</th>
<th>Tetramer Signal (I)</th>
<th>Relat. Ratio Dimer (%)</th>
<th>Relat. Ratio Tetramer (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>773.3</td>
<td>H2N2</td>
<td>5.2</td>
<td>0.7</td>
<td>1.34%</td>
<td>1.35%</td>
</tr>
<tr>
<td>919.2</td>
<td>FH2N2</td>
<td>35</td>
<td>0</td>
<td>8.99%</td>
<td>0</td>
</tr>
<tr>
<td>935.3</td>
<td>H3N2</td>
<td>87</td>
<td>38</td>
<td>22.34%</td>
<td>73.22%</td>
</tr>
<tr>
<td>1081.4</td>
<td>FH3N2</td>
<td>249</td>
<td>9.4</td>
<td>63.94%</td>
<td>18.11%</td>
</tr>
<tr>
<td>1097.5</td>
<td>H4N2</td>
<td>8.9</td>
<td>0.3</td>
<td>2.29%</td>
<td>0.58%</td>
</tr>
<tr>
<td>1243.4</td>
<td>FH4N2</td>
<td>0.6</td>
<td>0</td>
<td>0.15%</td>
<td>0</td>
</tr>
<tr>
<td>1259.4</td>
<td>H5N2</td>
<td>1.2</td>
<td>1.3</td>
<td>0.31%</td>
<td>2.50%</td>
</tr>
<tr>
<td>1284.8</td>
<td>FH3N3</td>
<td>1.9</td>
<td>0</td>
<td>0.49%</td>
<td>0</td>
</tr>
<tr>
<td>1421.5</td>
<td>H6N2</td>
<td>0.6</td>
<td>1.5</td>
<td>0.15%</td>
<td>2.89%</td>
</tr>
<tr>
<td>1583.5</td>
<td>H7N2</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0.96%</td>
</tr>
<tr>
<td>1745.6</td>
<td>H8N2</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0.39%</td>
</tr>
<tr>
<td>Σ of Fuc</td>
<td></td>
<td></td>
<td></td>
<td>73.6%</td>
<td>18.11%</td>
</tr>
</tbody>
</table>

**Detail Structural Analyses by Positive ion MS**

An aliquot of reduced N-glycans from both forms were permethylated. This derivative not only provides more structural detail upon collision induced disassociation, but aides in sample purification and ionizes more efficiently [Reinhold 1983]. Due to the complexity of the glycan structures, we will majorly discuss only to two abundant compositions Man3GlcNAcGlcNAco/
Figure 3.5. MALDI-MS profiles of native N-glycans from tetramer and dimer NA. (a) Tetramer, (b) dimer, (c) Zoom of tetramer (1200 -1800), (d) Zoom of dimer (1200 -1800). All calculated monosaccharides compositions are annotated in the spectra (Abbreviations, F: Fuc; H: Man; N: GlcNAc). All peaks contain one sodium atom.
and Fuc Man$_3$GlcNAcGlcNAco/ shared by both forms, along with a uniquely fucosylated structure in the dimer form (Fuc Man$_2$GlcNAcGlcNAco/).

**Structural Analysis of Composition of Man$_3$GlcNAcGlcNAco/**

The abundant ion in the tetramer NA with the composition (Man$_3$GlcNAcGlcNAco/) suggested it may be an N-glycan core structure. This was confirmed by MS$^n$ using the disassembling pathway, (MS$^5$, →1187.7→894.5→676.4→431.3; Figure 3.6a-d). The proposed fragmentation ions that accounted for the major peaks were set adjacent to each spectrum. MS/MS analysis of the molecular ion m/z 1187.7 provided an abundant product ion at m/z 894.5 due to a neutral loss of the reducing end GlcNAco/ (Δ = 293, Figure 3.6a). Isolation and subsequent MS$^3$ analysis of this fragment, m/z 894.5, provided an abundant ion, m/z 676.4, indicating the presence of a terminal mannose (Δ = 218), and also distinguished a cross-ring cleavage fragment ion $^{3.5}$ A, m/z 737.4 (Δ = 157). This ion and the absence of a 0.4A-type fragment characterized a 1-4 linkage between the core mannose and the penultimate GlcNAc (Figure 3.6b). The MS$^2$ precursor also fragments to provide a B- and C-type fragment (m/z 649.4, 667.4, Figure 3.6b) due to the neutral loss of the penultimate GlcNAc residue. Three diagnostic fragment ions, m/z 301.1, 329.2 and 343.2 were present in MS$^5$ spectrum (Figure 3.6d), confirming both Man-6 and Man-3 arms [Ashline 2005]. Therefore, the composition of Man$_3$GlcNAcGlcNAco/ was determined
as an N-glycan core structure. No other isomeric structures have been detected in MS^n spectra.

**Structural Analysis of Composition of FucMan₃GlcNAcGlcNAco/**

MS² analysis of the ion m/z 1361.7 (FucMan₃GlcNAcGlcNAco/) from the tetramer, provided a prominent B-type ion, m/z 894.4, and a minor neutral loss ion at m/z 490.3 (Figure 3.7a). These ions composition support a fucose residue linked to the core reducing end GlcNAco. The intervening linkages of this disaccharide fragment, FucGlcNAco, are known to be either at the 3- or 6-O-GlcNAc position, but the latter was indicated by the known specificity of the enzyme, PNGase F. Importantly, the disassembling pathway following the neutral loss, MS⁵: →1361.7→894.5→676.4→431.2 (Figure 3.7b – 3.7d), proved to be identical to the pathway of the core structure Man₃GlcNAcGlcNAco as discussed above (Figure 3.6b – 3.6d), indicating structural similarity. The presence of three diagnostic fragment ions at m/z 301.1, 329.2, and 343.2, derived from a cross-ring cleavage of the core mannose, confirmed the linkages on both the Man-3 and Man-6 arms.

Therefore, the structure was assigned as an N-glycan core structure with an extra fucose linked to the 6-O-position of the reducing end GlcNAco (N4-A, table 3.3).

A small ion, (m/z 1068.5, arrow insert, Δ = 293, figure 3.7a), suggested the presence of an isomeric structure without any fucosylated modification on the reducing end GlcNAc. Further disassembly of m/z 1068.7 (Figure 3.8a),
the fragment ions \( m/z \) 880.6 and 850.6 indicated the presence of fully methylated terminal Fuc and Man residues, respectively. Following that pathway (\( MS^4: \rightarrow 1361.7 \rightarrow 1068.7 \rightarrow 880.6 \)), the abundant fragment ion \( m/z \) 676.4 (\( \Delta = 204 \)), derived from a neutral loss of internal hexose, indicated that the terminal Fuc residue was attached to an internal hexose (Figure 3.8b), most likely the core mannose. Another fragment ion \( m/z \) 662.4 (\( \Delta = 218 \)) also confirmed the presence of the terminal hexose residue (Figure 3.8b). Further disassembly of that fragment, \( m/z \) 431.2, provided a disaccharide doubly scarred on the reducing terminus and a 6-O-linkage defined by the paired ions at \( m/z \) 301.1 and 329.2. Two diagnostic ions at \( m/z \) 301.1 and 329.2 via cross-cleavage fragmentation mechanism suggests that the terminal mannose was located on 6-arm (Figure 3.8d). The lack of ion \( m/z \) 343.2 indicated the Fuc-Man disaccharide substitute on 3-arm most likely. Unfortunately, the linkage between Fuc and Man was not possible to characterize due to the low ion current. Therefore, this isomeric structure was characterized as N4-B shown in table 3.3. No other isomeric structures have been detected for this composition FucMan\(_3\)GlcNAcGlcNAcol.

**Structural Analysis of Composition FucMan\(_2\)GlcNAcGlcNAcol**

The profiled ion \( m/z \) 1157.7 provided a composition, FucMan\(_2\)GlcNAcGlcNAcol, which was unique in the dimer NA. \( MS^2 \) disassembly of the molecular ion showed a prominent peak \( m/z \) 690.4 (Figure 3.9a), corresponding to the neutral loss of the reducing end of
Fuc-GlcNAco moiety (Δ = 467), indicated the terminal core fucosylation.

Assuming a core structure, the antennal positioned mannose moiety must be either 3-, or 6-linked to the central core Manα(1-4)GlcNAc. Further loss of the internal GlcNAc residue leaded to a dimer moiety at non-reducing end (MS4: →1157.7→690.4→445.2; Figure 3.9c). Two cross-ring cleavage ions m/z 329.2 and 301.1 showed that the terminal residue is the branching 6-linked mannose. The 3-linked mannose residue appears missing due to the absence of the cross-ring cleavage fragment ion at m/z 343.2 such as seen in other compositions. Thus the major isomeric structure of this composition FucMan2GlcNAGlcNAco was characterized as N2-A listed in table 3.3.
Figure 3.6. Sequential MS$^n$ disassembly of the composition Man$_3$GlcNAc$_2$ prepared as the methyl-reduced derivate, Man$_3$GlcNAcGlcNAcol. Pathways are indicated in the top left corner of the spectra (a – d). Fragments accounting for the spectrum are proposed on the structure to the right as dotted lines. In spectra (c) and (d), two possible isomeric fragments are proposed to follow the same pathway. All ions are single charged sodium (Na$^+$) adducted ions.
Figure 3.7. Sequential MS^n disassembly of the composition FucMan_3GlcNAc_2 prepared as the methyl-reduced derivate, FucMan_3GlcNAcGlcNAcol. Pathways are indicated in the top left corner of the spectra (a – d). Fragments accounting for the spectrum are proposed on the structure to the right as dotted lines. In spectra (c) and (d), two possible isomeric fragments are proposed following the same pathway. All ions are single charged sodium (Na^+ ) adducted ions.
Figure 3.8. Sequential MS\textsuperscript{n} disassembly of the composition Fuc\textsubscript{3}Man\textsubscript{3}GlcNAc\textsubscript{2} prepared as the methyl-reduced derivate, FucMan\textsubscript{3}GlcNAccGlcNAcol. Pathways are indicated in the top left corner of the spectra (a – d). Fragments accounting for each ion are proposed on the structure to the right as dotted lines. All ions are single charged sodium (Na\textsuperscript{+}) adducted ions.
There was a small peak showing a neutral loss of 293 Da (m/z 864.4, Figure 3.9a), which indicated the absence of core fucosylation. Isolation of this ion and further fragmentation provided two prominent product ions at m/z 676.4 (Δ = 188) and 472.3 (Δ = 188+204), indicating the presence of a non-reducing terminal Fuc-Man moiety (Figure 3.10a and 3.10b). The diagnostic ion at m/z 315.1 suggests this disaccharide moiety exclusively locating at Man-6 arm (Figure 3.10b). Therefore this low abundant isomeric structure was characterized as Fuc-Man(1-6)Man(1-4)GlcNAc(1-4)GlcNAc.

Another abundant ion m/z 646.3 (Δ = 218) in MS³ spectra of m/z 864.4 (Figure 3.10a) showed a neutral loss of 218 Da and suggested the presence of a terminal Hex residue. This cannot be explained from the one proposed structure and indeed it indicates the presence of another isomeric structure. This terminal hexose residue was also detected in the subsequent MS⁴ spectra of m/z 676.4 due to the presence of daughter ion m/z 458.2 (Δ = 218; Figure 3.10b). The cross-ring cleavage ion at m/z 343.1 indicated the terminal hexose was located on the Man-3 arm (Figure 3.10b). It also suggested that the terminal fucose most likely was attached to C-6 of the core mannose (Figure 3.10b) because an intact cleavage of the internal GlcNAc was observed at ions m/z 419.2 (Δ = 227), and 401.2 (Δ = 245) in an alternative MSⁿ pathway (MS⁴: →1157.7→864.4→646.4; Figure 3.10c). Therefore, this isomer was proposed as N2-C (Table 3.3).
All the compositions from both forms of NA were analyzed by MS^n method as discussed above except three large ions (Hex_6GlcNAc_2; Hex_7GlcNAc_2; and Hex_8GlcNAc_2), which couldn't be analyzed due to the weak signals. As for the compositions shared by both forms, no unique isomeric structure has been found between them. All characterized structures were proposed in table 3.3.

**N-glycans from the Stalk Region**

The five N-glycosylation sites (Asn55, Asn58, Asn63, Asn68, and Asn88) were located in a single large tryptic peptide (Ile7 to Lys102) with the size over 10K Da (Figure 3.3). The rest of the tryptic peptides were relatively small, less than 6k Da making the stalk peptide easy to isolate from other peptides via 10K MWCO membrane dialysis. An N-glycan differential analysis of these stalk peptides were prepared and the native MS profiles were obtained from MALDI-MS (Figure 3.11). The fucosylated ion (m/z 1079.4, FucMan_3GlcNAc_2) was the most abundant one and the core ion m/z 933.3, was around 50% abundant in the dimer NA (Figure 3.11A). In contrast to the dimer, the core ion m/z 933.3 was the most abundant peak in the tetramer NA, and the fucosylated ion, m/z 1079.4, was barely visible (Figure 3.11b). The other smaller ions in the total glycan profile were not observed from this fraction due to the low abundance.
Figure 3.9. Sequential MS<sup>n</sup> disassembly of the composition FucMan<sub>2</sub>GlcNAc<sub>2</sub> prepared as the methyl-reduced derivate, FucMan3GlcNAcGlcNAcol. Pathways are indicated in the top left corner of the spectra (a – c) as fellows: MS<sup>2</sup>→1157.7→690.4→445.2. Fragments accounting for the spectrum are proposed on the structure to the right as dotted lines. All ions are single charged sodium (Na<sup>+</sup>) adducted ions.
Figure 3.10. Sequential MS^n disassembly of the composition FucMan₂GlcNAC₂ prepared as the methyl-reduced derivate, FucMan₃GlcNACGlcNACol. Pathways are indicated in the top left corner of the spectra (a – c) as fellows: MS^4→1157.7→864.4→676.4 and MS^4→1157.7→864.4→646.4. Fragments accounting for the spectrum are proposed on the structure to the right as dotted lines. All ions are single charged sodium (Na^+) adducted ions.
Table 3.3: The Characterized N-glycan structures from dimer and tetramer NA by MS$^n$ analysis. (H: Man; N: GlcNAc; F: Fuc)

<table>
<thead>
<tr>
<th>N-glycan ID</th>
<th>Glycan Structure and MS$^n$ pathway</th>
<th>Dimer</th>
<th>Tetramer</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1 (H2N2)</td>
<td>MS$^n$ Pathways: 983.6 $\rightarrow$ 690.5 $\rightarrow$ 445.3</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="N1 diagram" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N2-A (H2N2F)</td>
<td>MS$^n$ Pathways: 1157 $\rightarrow$ 690 $\rightarrow$ 445</td>
<td>+++</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="N2-A diagram" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N2-B (H2N2F)</td>
<td>MS$^n$ Pathways: 1157 $\rightarrow$ 864 $\rightarrow$ 676 $\rightarrow$ 472</td>
<td>+</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="N2-B diagram" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N2-C (H2N2F)</td>
<td>MS$^n$ Pathways: 1157 $\rightarrow$ 864 $\rightarrow$ 646 $\rightarrow$ 458</td>
<td>+</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="N2-C diagram" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**N3 (H3N2)**

MS Pathways:
1187 → 894 → 649 → 431

**N4-A (H3N2F)**

MS Pathways:
1361 → 894 → 676 → 431

**N4-B (FH3N2)**

MS Pathways:
1361 → 1068 → 880 → 676

**N5 (H4N2)**

MS Pathways:
1391 → 1098 → 880 → 635
<table>
<thead>
<tr>
<th></th>
<th>MS&lt;sup&gt;n&lt;/sup&gt; Pathways:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N6</strong></td>
<td>1565 → 1098 → 880</td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td>(H4N2F)</td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>MS3 1098</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS4 880</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Diagram]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N7-A</strong></td>
<td>1595 → 1302 → 880</td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td>(H5N2)</td>
<td></td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>MS3 1302</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS4 880</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Diagram]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N7-B</strong></td>
<td>1595 → 1302 → 1084 → 866 → 648</td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td>(H5N2)</td>
<td></td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>MS3 1302</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS5 866</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS6 648</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Diagram]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N8</strong></td>
<td>1606 → 1139 → 880 → 676 → 431</td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td>FH3N3</td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>MS3 1139</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS6 431</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Diagram]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Putative structure, no MS(^n) analysis due to weak signals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------------------------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>N9 H6N2</td>
<td>![Diagram of N9 H6N2 structure]</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>N10 H7N2</td>
<td>![Diagram of N10 H7N2 structure]</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>N11 H8N2</td>
<td>![Diagram of N11 H8N2 structure]</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

N/A: Not applicable for this form. +++: major component; +: minor component.
Figure 3.11. MALDI-MS profiles of N-glycans were released from the stalk region tryptic peptide. Native glycans from dimer of NA (A) and tetramer of NA (B). The fucosylated peak m/z 1079.4 is almost invisible in the tetramer of NA. All peaks are single sodiated adducts.

**Incorporation of $^3$H-Fuc onto Dimer and Tetramer NA *in Vitro***

Human fucosyltransferase 8 (FUT8) [Ihara 2007], can add fucose onto the C-6 position of the reducing end GlcNAc. It was chosen to incorporate $^3$H-fucose onto dimer and tetramer NA *in vitro*. The reaction yield was measured by filter binding assay as described in the experimental section. The tetramer NA incorporated almost 4 times more fucose than the dimer (Figure 3.12), and assayed for enzyme release activity that the tetramer NA still maintains its sialic acid release ability (Figure 3.13). The Addition of extra fucose onto core GlcNAc seems to neither affect the tetrameric structure, nor lead the tetramer to de-oligomerize.
Figure 3.12. $^3$H-Fuc was incorporated onto dimer and tetramer NA by FUT8 enzyme in vitro. The radioactivity was measured by filter binding assay. D: dimer, T: tetramer, FUT8: human fucosyltransferase 8.

Figure 3.13. The tetramer NA enzyme activity after fucosylated with FUT8. No enzyme activity change was observed after the tetramer NA was added extra core fucose by FUT8.

3.5 Discussion

Fucosylation Seems the Key Difference between Dimer and Tetramer

The glycoprotein NA of 1918 influenza viruses (H1N1) exists in monomeric, dimeric, and tetrameric forms, but only the tetramer demonstrates the enzyme activity (the ability to release sialic acid residue from glycoconjugates). Previous report has showed that N-glycosylation
pattern was the only difference between tetramer and dimer/monomer [Wu 2009]. In order to seek the structural features which are responsible for the formation of the tetramer NA, N-glycans from both dimer and tetramer NAs were enzymatically released and characterized by MS^n.

As expected, recombinant NA expressed simple pauci-mannose type and high mannose type N-linked glycans (Table 3.3) because inset cells do not carry the enzymes required for more complex type glycans [Kost 2005]. However, the distinct difference between dimer and tetramer is the features of activity and a differential measure of fucosylated glycans. Dimer and tetramer share one fucosylated composition, FucMan\(_3\)GlcNAc\(_2\) (63.9% in dimer vs. 18.1% in tetramer). This is the only fucosylated composition found in the tetramer. Other three fucosylated compositions, FucMan\(_2\)GlcNAc\(_2\), FucMan\(_4\)GlcNAc\(_2\), and FucMan\(_3\)GlcNAc\(_3\), were also found in the dimer, but not in the tetramer. Fucosylated glycans accounts for 73.6% of the total in the dimer, but only 18.1% in the tetramer (Table 3.2). Detail characterization via MS^n showed that the majority of fucosylated structures occurred on the 6-position of the reducing end GlcNAc, although the trace amounts were found on the outer antennae of mannose such as N4-B (Table 3.3). For the compositions shared by both dimer and tetramer forms, no structural difference was found. Even there were two large high-mannose type glycans (H\(_7\)N\(_2\) and H\(_6\)N\(_2\)) in tetramer, absent in dimer, but both of them only consists less than 1.5% in total (Table 3.2). Thus it appeared that excess fucosylation,
especially 6-Fuc modification on the reducing GlcNAc, was the key difference between tetramer and dimer.

**Fucosylation Fails to Alter Tetramer Enzyme Activity**

Fucose can be enzymatically (FUT8) added *in vitro* to the core GlcNAc of the dimer and tetramer NAs. The tetramer when compared to the dimer can accept almost four times more fucose when measured by a filter binding assay (Figure 3.12). This was in agreement with the mass spectrometry results which indicated that fucosylated glycans were 73.6% in the dimer but only 18.1% in the tetramer (Table 3.2). To our surprise, the tetramer NA retains its enzyme activity after the incorporation of excess fucosylation, suggesting that the \(N\)-glycans were not directly involved in the receptor binding or influence the tetrameric three-dimensional structure once they are formed. It does seem that \(N\)-glycans play a key role in the oligomerization process to regulate the formation of the active tetrameric shape before it reaches the cell surface, but has no influence once formed.

**Fucosylation in the Stalk Region**

Active tetrameric NA consists of a thin stalk region and a large globular binding head. The interaction among the four identical monomers in the stalk region is essential to retain active tetrameric structure [Xu, 2008]. It is currently not known whether the glycans in stalk area contribute directly to that configuration but given the compact shape of the stalk, the glycans in this
area may affect the correct folding of NA subunits prior to the formation of the active NA. Five potential \(N\)-glycosylation sites (Asn50, 58, 63, 68 and 88) occur in the stalk region, when isolated and profiled for \(N\)-glycosylation only a trace amount of the fucosylated structure (FucMan3GlcNAc2, \(m/z\) 1079.4) was detected in the tetramer NA which was in contrast to the dimer (Figure 3.11a and 3.11b). Since the stalk region contained four glycosylation sites (Asn50, 58, 63, and 68), it can be suspected that the trace amount of fucosylation may come from the site Asn88. Core fucosylation may have a definitive effect on the correct folding of each subunit due to its proximity to the protein backbone, comparing to other sugar residues. The four \(N\)-glycosylation sites in the stalk region are shared by all known avian N1 type viruses, and most human H1N1 type viruses [Xu 2008; Reid 2000], implicating its significance to the structural formation. Unlike HA, which acquired up to seven extra glycosylation sites for human viruses under selective pressure, human H1N1 NA doesn’t acquire any extra glycosylation site in the stalk region, and only sporadic cases reported with one or two additional site in the globular head [Reid 2004]. Thus, it may suggest that \(N\)-glycans in the stalk region is essential of the tetramer NA formation, and extra glycosylation site may inhibit the formation of that structure, even if it may benefit the virus to escape the host antibody neutralization.
**Fucosylation is Ubiquitous in Human Cells**

Although human cells produce more complex-type *N*-glycans rather than simple high-mannose type glycans, the fucosylation is ubiquitous in the human cells [Ma 2006]. Core fucosylation may carry the same property of inhibiting the tetramer formation in human cells regardless of the extensive modification on the antennae. Fucosylation has been reported to take part in many important biological functions, such as: blood transfusion reactions, selectin-mediated leukocyte-endothelial adhesion, host-microbe interactions, and numerous ontogenic events [Ma 2006]. Alternations in the expression of fucosyltransferase activity have also been observed in many pathological processes, such as liver disease, and cancer developing, etc [Mehta 2008; Comunale 2009; Miyoshi 2008; Imre 2008; Shah 2008]. Thus we suggest another example where fucosylation may control the formation of influenza virus NA. This is the first report to characterize the structural details of *N*-glycosylation of the viral NA. With this assumption, it exposes an alternative mechanism for the host system to regulate the influenza virus infection which has never been reported previously. Most importantly, this result may bring to focus a new anti-virus strategy for consideration.

**Fucosylation may Explain the High Death Rate in Young Adults**

It has been reported that NA was one of three factors critical to its highly pathogenic nature of the 1918 virus [Pappas 2008]. Given the severity of that pandemic, it is reasonable to speculate that the 1918 viruses can grow
efficiently in human respiratory cells. An intriguing nature of the 1918 viruses was its higher death rate among the young adult age group 25-40. This can not be explained by common immunological considerations. Hence, it may be the case less fucosylation enzyme in this age group 25-40 makes them more susceptible to virus infection due to the production of more tetrameric NA. However, there is no direct evidence to correlate fucosylation activity to various age groups at the present, especially those enzymes in the respiratory epithelial cells. This is only a hypothesis.

**Summary**

In this chapter, a glycosylation difference between the active NA (tetramer) and the inactive NA (dimer) of the 1918 influenza virus has been identified. It appears that fucosylation on N-glycans, especially those located on the stalk region, prevents oligomer formation to the active NA tetramer. Since addition of fucose onto the innermost GlcNAc of N-glycans in tetramer did not affect tetramer’s enzyme activity, it may suggest that this kind of regulation is occurred in the early stages of glycoprotein biosynthetic pathway. Moreover, this is the first report to characterize N-glycan structures of the NA from influenza A virus, which will hopefully expand our understanding to this notorious virus, and help us to prepare for future pandemics.
CHAPTER 4

Sial-Gal Linkage Analysis of the Human Receptor of Influenza Virus

4.1 Introduction

Influenza Virus Receptor

Avian influenza, a highly contagious virus, has been able to cross species barriers and infect a broad range of mammals including human, pig, horse, tiger, whale, etc (Table 3.1 in Chapter 3). As discussed in Chapter 3, the major surface glycoprotein haemagglutinin (HA) is the molecule that recognizes the host receptor and initialize viral entry into the host cell [Weis 1988]. A terminal disaccharide moiety sialic acid-galactose (Sial-Gal) either from glycoprotein or glycolipid on the host cell surface is the receptor for HA attachment. It has been long believed that avian influenza viruses prefer binding to a Sialα(2,3)Gal moiety, whereas the human viruses favors binding to Sialα(2,6)Gal epitope [Pekosz 2009; Russell 2008; Shinya 2006; Debby 2006]. Because of this recognition barrier, avian influenza will not infect humans directly, and vice versa, human viruses do not transmit to avian species. Pigs were traditionally considered as “the mixing vessels” to produce recombinant viruses, which are able to infect both human and avian
species as pigs have both $\alpha(2,3)$ and $\alpha(2,6)$ Sial-Gal linkages in their respiratory tracts [Gambaryan 2005; Ha 2001; Kida 1988].

Recently direct avian-to-human infections have been reported on several subtype viruses: such as H5N1 [Webster 2005; Guan 2004], H7N7 [Fouchier 2004], H9N2 [Wan 2007], associated with the enhanced pathogenicity. The emergence of these highly pathogenic viruses increases the likelihood of future influenza pandemics. Most importantly, these direct avian-to-human infection cases are controversial in light of the previous proposal of HA binding specificity. Thus, the presence of Sial$\alpha(2,3)$Gal moiety in the human respiratory tract is the key issue and needs to be addressed indefinitely.

It will be necessary to understand the basic human anatomical distribution of respiratory tract tissues and organs before considering mechanism of influenza virus infection. Generally the human respiratory system can be divided into an upper and a lower respiratory tract (Figure 4.1). The upper tract includes the nasal passages, pharynx and the larynx, while the lower comprises trachea, bronchi and lungs (Figure 4.1).

Previously, Sial$\alpha(2,3)$Gal epitope has been detected in the human lower tract but not the upper by using classical lectin methods (Figure 4.2) [Shinya 2006]. The detection of Sial$\alpha(2,3)$Gal is almost invisible in the upper tract, which were dominated by Sial$\alpha(2,6)$Gal moiety (Figure 4.2, a-c). In contrast, Sial$\alpha(2,3)$Gal is present sporadically in bronchiole tissues, and has
significant amounts in alveolus tissues of the lower respiratory tract (red color, Figure 4.2, d-e). From these results, it was proposed that avian virus such as H5N1 can infect and replicate efficiently in the human lower respiratory tract if the virus get the opportunity to enter the lower respiratory tract (such as farm workers who may have direct and repeat contact to sick poultry). The paucity of Sialα(2,3)Gal epitope in the human upper respiratory tract may prevent readily human-to-human transmission of avian viruses, which are not usually able to reach the lower tract. Moreover, avian H5N1 virus was also reported to infect human type 2 pneumocytes and alveolar macrophages, two primary cells from the human lower respiratory tract [Riel 2006].

**Human respiratory tract**

![Diagram of the human respiratory tract.](image)

*Figure 4.1. The anatomical distribution of human respiratory tract.*
Figure 4.2. Reactivity of human respiratory tissues with lectins specific for different sialic acid linkages [Adapted from Shinya 2006]. (a) Nasal mucosa; (b) paranasal sinuses; (c) bronchus; (d) bronchiole; (e) alveolus. Res: respiratory bronchiole (adjacent to alveoli); Ter, terminal bronchiole (distal to alveoli); Alv, Alveolus. Green colors, reaction with *Sambucus nigra* lectin, indicate the presence of Sialα(2,6)Gal. Red colors, reaction with *Maackia amurensis* lectin, indicate the presence of Sialα(2,3)Gal.

In contrast to the above conclusions, Nicholls et al. has found that the human nasopharyngeal, adenoid and tonsillar tissues (cultured *ex vivo*) can be infected with avian H5N1 viruses regardless of the paucity of Sialα(2,3)Gal in these tissues [Nicholls 2007a]. In addition the viral yield was increasing with the incubation time, indicating an effective infection and replication. This was further supported from a report that H5N1 virus has been detected in trachea epithelial cells of a human patient [Gu 2007]. These results strongly suggested that avian influenza virus receptors must be
present in the human upper respiratory tract. These opposing data to this most important question prompted us to apply mass spectrometry-based method to clarify this issue avoiding the biological constraints inherent with lectins.

**Common Lectin Histochemistry Methodology**

Lectins are a class of naturally occurred proteins which are characterized by their properties to bind specific carbohydrate residues. They have been widely used in various biological researches due to their affinities to specific carbohydrate epitopes. In fact, Haemagglutinin (HA) of influenza virus is also a type of lectin, which binds to the sialic acid galactose residue on the host cell surface. Usually lectins are conjugated with fluorescent groups to facilitate detection. There are two types of lectins commercially available to investigate Sial-Gal linkage: *sambucus nigra* agglutinin (SNA) for Sial\(\alpha(2,6)\)Gal and *maacka amurensis* agglutinin (MAA) for Sial\(\alpha(2,3)\)Gal. The later exists in two different isoforms, MAA1 and MAA2 [Nicholls 2007a; Nicholls 2007b]. It was reported that MAA2 is more specific toward Sial\(\alpha(2,3)\)Gal/\(\beta(1,3)\)GalNAc moiety, and MAA1 is more specific to Sial\(\alpha(2,3)\)Gal/\(\beta(1,4)\)GlcNAc residue [Konami 1994]. Since MAA1 lectin has been reported to bind to non-sialic acid components, it has not been used as widely as MAA2 [Nicholls 2007b].

Nicholls *et al.* recently investigated the Sial-Gal linkages of the human respiratory tract using these lectins and tried to determinate the accuracy of
them, and even compared their binding specificities supplied by different vendors [Nicollas 2007b]. The key purpose was to evaluate the lectins accuracy and reliability. Their results are summarized as below:

- **MMA1** have widespread binding areas in both upper and lower respiratory tracts, and more binding affinity to the cells from children than those from adult.

- **MMA2** binding was mainly restricted to the lower respiratory tract, such as alveolar cells, which would be comparable to Shinya et al. result (Figure 4.2).

- **SNA** binding was observed throughout the overall respiratory tract, and stronger signals to children cells and tissues, compared to adults.

- **MMA2** lectins from three different vendors: Roche, EY labs, and Vector labs, showed different binding specificities.

MMA1 binding was observed in the upper respiratory tract, suggesting the presence of Sialα(2,3)Gal. However, MMA1 was able to bind to non-sialic acid residues, thus this result was not definite and reliable. SNA binding was found in the whole respiratory tract, indicating the wide spread of Sialα(2,6)Gal moiety, in agreement with current human HA binding specificity.

MMA2 lectins have distinct binding patterns among different suppliers, indicating the results from the commercial lectins were not reliable. Most importantly, these findings strongly suggest that common lectins were neither
accurate nor reliable to detect the Sialα(2,3)Gal moiety in the human respiratory tract. Moreover, current glycoarray results also found that MMA lectin can't recognize all sialic acid galactose analogs, which are bound by avian H5N1 virus [Steven 2006b]. This suggests even though both of HA and MMA are the same kind of lectin, their specificity may vary, also depending on other factors [Gambaryan 2005b]. An alternative and more-reliable method to determinate Sial-Gal linkage is sorely needed. And mass spectrometry is the key instrument in glycoproteomic field, and more important the results from mass spectrometry are definitive without any biological constraints.

**Mass Spectrometry-Based Strategies**

Since sialic acid is a terminal antennal modification, the common method would be to employ a specific exoglycosidase (called sialidase/neuraminidase) to digest the free glycans, and measure the mass difference before and after enzyme digestion. This method has been widely used to determinate other monosaccharides, such as galactose, mannose etc. Two problems are encountered when identifying Sial-Gal linkage. First, no highly-specific commercial sialidase, that can differentiate α(2,3) or α(2,6) linkages, is available. Currently α(2,3) neuraminidase from streptococcus pneumoniae (Sigma catalog # N7271; Sigma, St. Louis, MO) can hydrolyze α(2,6) linkage but with a relative lower reaction rate. Secondly, glycans with sialic acid are not easily ionized under positive ion mass spectrometry due to
the negative charge of the sialic acid. Also fragmentations have been reported for native sialic acid glycans using MALDI instrument [Harvey 1999].

A second mass spectrometry-based strategy has been employed using MS² spectral comparison of known standards [Ashline 2005]. This was successfully demonstrated to characterize unusual modifications on the innermost GlcNAc on *C. elegans* glycans [Hanneman 2006]. As for Sial-Gal disaccharide residue, it is difficult to obtain this moiety intact due to the extremely labile nature of sialic acid glycosidic bond. However, the loss of the terminal sialic acid will expose either 3- or 6-position hydroxyl group on the penultimate galactose depending on the specific linkage. This fragment ion will have a unique spectrum. This method has been successfully applied to determine the linkage in a recombinant IgG Fc *N*-glycans [Anthony 2008]. The reason for this research was that 100% of the anti-inflammatory function of IgG was completely dependent on one specific *N*-glycan sialylation structure. As shown in the Figure 4.3, the Sial-Gal linkage from IgG Fc glycan was shown to be a α(2,6) linkage since in the product hydroxyl galactose spectrum was closer to that obtained from the 2,6-sialyllactose standard [Anthony 2008].

Another mass spectrometry strategy was published recently for Sial-Gal linkage determination using a chemical cyclization method [Wheeler 2009]. Traditionally stabilization of sialic acid was obtained by selectively methylation of the carboxyl group with methyl iodide, to form the
corresponding methyl ester. This enables the detection of sialic acid glycans in MALDI spectra [Powell 1996]. The new strategy, the sialylo-glycans were esterificated in methanol with 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl morpholinium chloride (DMTMM). The intriguing feature of this method was not only stabilization of sialic acid, but also differentiation of Sial-Gal linkage directly since \( \alpha(2,6) \) linkage will form a methyl -ester, and \( \alpha(2,3) \) linkage will produce a six-member ring lactone between sialic acid and the penultimate galactose. The mass difference between one methyl-ester and one lactone unit is 32 Da, allowing the mass differentiation in a MS profile. This method has been tested successfully on the well-characterized glycoprotein bovine fetuin, which carries both Sial\( \alpha(2,3) \)Gal and Sial\( \alpha(2,6) \)Gal linkages (Figure 4.40. When reacted with methyl iodide, all carboxyl groups were converted to methyl ester, and only one single peak was present for each composition. However, when reacted with DMTMM, multiple peaks were observed for each composition due to the mass difference for Sial-Gal \( \alpha(2,3) \) and \( \alpha(2,6) \) linkages (Figure 4.4).

Due to the significance of Sial-Gal linkage determination on influenza virus infection, both MS\(^n\) spectral comparison and DMTMM/MeOH cyclization methods have been applied to identify the exact Sial-Gal linkage on the human respiratory tract. In this project, two model cell lines were chosen: Detroit 562 and A549, (obtained from ATCC), as representation of the human upper and lower respiratory tract, respectively.
4.2 Experimental Procedures

Materials and Reagents

Human pharynx epithelial cell lines (Detroit 562) and Type II alveolar bronchial epithelial cell line (A549) were purchased from American Type Culture Collection (ATCC) Inc. (Manassa, VA 20108). The glycerol-free PNGase F and associated reagents (10 × denaturing buffer, NP-40, G7 buffer) were purchased from New England Biolabs (Beverly, MA). Proteomic grade trypsin (25 μg per vial), dimethyl sulfoxide, sodium hydroxide, iodomethane, sodium borohydride, dithiothreitol, urea, iodoacetamide, trifluoroacetic acid, tetrahydrofuran, 2,5-dihydroxybenzoic acid (DHB), toluene and ethanol were obtained from Sigma-Aldrich (St. Louis, MO). Membrane dialysis cassette (7000 MWCO) was purchased from Pierce (Prod# 66370, Rockfold, IL).
Figure 4.3: One Strategy for determining Sial-Gal linkage. (A) The IgG Fc glycan is a bisecting core complex glycan with potential one or two sialic acid terminal modification (Structure). Neu5Ac, N-acetylneuraminic acid (sialic acid). (B) Sequential mass spectrometry (MS^n) analysis of SNA-enriched IVIG Fc glycans is performed to determine the sialic acid linkage to the penultimate galactose. The B/Y galactosyl fragment monomer derived from MS^n disassembling was compared to the analogous B/Y fragments from (C) 2,3-sialyllactose and (D) 2,6-sialyllactose standards. The spectrum generated from SNA+ IVIG Fc glycan (B) most closely matches that of 2,6-sialyllactose (D) [Anthony 2008].
Figure 4.4. A second strategy for determining Sial-Gal linkage using DMTMM. N-glycans from bovine fetuin after derivatization with (a) methyl iodide in DMSO, and (b) methanol/DMTMM. The presence of both Sial-Gal α (2,3) and α(2,6) linkages were demonstrated by the multiple peaks for a single glycan in (b) [Wheeler 2009].

Total Protein Extraction

Both cell lines were cultured according to the provider's recommendation (www.atcc.org). The cells were harvested and suspended in 100 mL of an ice-cold lysis buffer. The lysis buffer (1L NP-40) was combined with 30 mL of 5M NaCl, 100 mL of 10% NP-40, 50 mL of 1M Tris (PH 8.0), and 820 mL of water. The lysis buffer was shaken occasionally while standing in an ice-bath for 30 minutes. The nuclei and cellular debris were removed by low-speed centrifuge (1000 rpm for 5 minutes). The pellet was discarded and the supernatant was kept for the subsequent analysis.
Synthesis of DMTMM

4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) was prepared from N-methylmorpholine (NMM, Sigma) and 4-chloro-2,6-dimethoxytriazine (CDMT, Lancaster) in anhydrous tetrahydrofuran [Kunishima 1999]. Briefly, 101 mg NMM (1 mmol) and 193 mg CDMT (1.1 mmol) was dissolved in 3mL THF at r.t. The reaction was stirred for 30 minutes. Then the solution was washed by 0.5 mL THF several times and centrifuged to discard the supernatant. The crude product was dried under vacuum stored in -20°C for the following reaction without further purification.

N-glycan Release by PNGase F

The total proteins extracted from both cell lines were dialyzed against 5 mM NH₄HCO₃/0.01% SDS overnight (7000 MWCO) to remove low molecular salts and impurities. The proteins were recovered, dried and stored in -20 °C for future use. Dried protein sample (around 1 mg) was dissolved in a protein denaturing buffer, and the N-glycans were released following the same procedure as described in Chapter 2.

Reductive O-glycan Release

The total protein (1 mg each) was dissolved in 50 mM NaOH/1M NaBH₄ solution and incubated at 50 °C for 16 hours. The reaction was terminated via the addition of 1 mL acetic acid in an ice bath. Three ML of
ethanol was added into the solution, and the solvents removed by evaporation in a speed-vac. Volatile methyl borates were removed by repeat evaporations with 1 mL of 1% acetic acid in methanol under a stream of nitrogen gas.

**Glycan Purification and Permethylation**

Released glycans were separated from the proteins by a Sep-Pak C18 SPE cartridge (Waters). The glycans were further desalted and purified using a hand-packed graphitized carbon cartridge (Alltech) into neutral (25% acetonitrile fraction) and acidic (30% acetonitrile containing 0.1% trifluoroacetic acid) fractions. The dried glycan sample was dissolved in 300 μL DMSO in a glass tube, followed by the addition of a small amount ground NaOH powder and vortexing for 5 mins. Iodomethane (50 μL) was then introduced into the tube, followed by vortexing extra 40-50 mins, to form a white cloudy suspension solution. The reaction was stopped by the sequential addition of 1 mL dichloromethane and 1 mL of 5% acetic acid in ice bath. The methylated glycans were isolated by liquid-liquid (CH₂Cl₂/H₂O) extraction, repeated 3 times. The combined organic layers were then washed 5-6 times with water. The dichloromethane was evaporated under N₂ gas and ready for mass spectrometry analysis.

**Glycans Modified with DMTMM/MeOH**

Purified acidic N-glycans (from 500 μg protein) were dissolved in 100 μL MeOH with a small amount of DMTMM (~20 mg). The reaction solution
was incubated at 80 °C for 90 minutes. The methanol was removed under a
stream of nitrogen gas, and the mixture was re-dissolved in 50 μL H₂O. 10 μL
chloroform was added repeatedly (three times) to extract the excess
reagents. The product was further purified by a hand-made graphitized
carbon column (30% acetonitrile containing 0.1% TFA fraction). The solvent
was removed and stored in -20°C for mass spectrometry analysis.

Mass Spectrometry

Native glycan MS profiles were performed on a MALDI-TOF
instrument (MALDI-CFR, Shimadazu, Kyoto, Japan). One μL matrix solution
(10 mg DHB in 1 mL of 50% acetonitrile solution containing 0.1% TFA) was
applied on MALDI plate and air dried. Then 1 μL sample solution was applied
to the well and mixed evenly with the matrix on the plate. When the solution
was dry, a drop of pure ethanol was applied onto the well and the sample and
matrix re-crystallized. MSⁿ analyses were performed on the permethylated
reduced glycans (ESI-LTQ, Thermo, Waltham, MA).

4.3 Results

Sialylated Standard Carbohydrates Tested with DMTMM

To fully evaluate this method, two standard sialylated trisaccharides
(2,3 sialyllactose and 2,6 sialyllactose) were tested with DMTMM reagent
(Figure 4.5B). Only one major product peak (methyl-ester: m/z 670.2) was
observed from 2,6 sialyllactose, and no starting material remained. It indicated quantitative conversion of 2,6 linkage to the corresponding methyl-ester product (Scheme 4.1). In contrast, most of 2,3 sialyllactose was converted to the expected lactone product ($m/z$ 638.2, Figure 4.5A), which is 32 Da less than the methyl ester product. However, two small peaks, corresponding to the starting material and the methyl ester were also found ($m/z$ 656.1 and 670.2, respectively, Figure 4.5A) in the spectrum. Although this result may suggest Sial-Gal $\alpha$\textsubscript{(2,3)} linkage may not be able to covert to the lactone quantitatively which was in agreement with the recent report [Wheeler 2009], the major purpose for this project is only to qualitatively detect the presence of 2,3 linkage.
Figure 4.5. Positive MALDI-TOF mass spectra of two standards; 2,3 sialyl-lactose (A) and 2,6 sialyl-lactose (B) after derivatization with methanol/DMTMM. The 2,6 linkage formed the methyl-ester product quantitatively, no starting material and other products were observed. The major product of α2,3 linked sialic acid standard was the lactone, whereas some starting material (S.M.) and methyl-ester products were observed. All ions contain one sodium atom.

**N-Glycans from Bovine Fetuin Reacted with DMTMM**

Further verification of this method was carried out on the N-glycans from bovine fetuin which possessed both sialic acid linkages, and also reported by Wheeler (Figure 4.4). Free and purified fetuin N-glycans were modified with DMTMM in methanol. The result is shown in Figure 4.6, which is similar to Figure 4.4B. The abundant composition was tri-antennary N-glycans with two major peaks (m/z 2879.5 and 2911.3), corresponding to the Sial-Gal combination of (2,3)_2 (2,6) and (2,3)(2,6)_2, respectively (Figure
The second composition was the tri-antennary N-glycan with four sialic acids groups. Two major combinations of this composition were the $(2,3)_2(2,6)_2$ and $(2,3)(2,6)_3$ ($m/z$ 3184.4 and 3216.4), respectively. The single sialic acid composition is almost invisible (Figure 4.6), which was evident in Figure 4.4. This slight difference may come from the glycosylation pattern change among different samples. Most interesting, this result successfully demonstrates that this method is useful to differentiate Sial-Gal linkages, especially for those containing multiple sialic acid residues. In addition, no significant un-reacted material was found in this spectrum.

Scheme 4.1. Expected derivatization products of $\alpha(2,3)$ and $\alpha(2,6)$ linkages of Sial-Gal using the reagent DMTMM in MeOH. The distinct mass difference between the lactone and methyl-ester product is 32 Da per unit which is suitable for mass spectrometry differentiation.
Figure 4.6. Positive MALDI-TOF mass spectrum was obtained on N-glycans from bovine fetuin after derivatization with methanol/DMTMM. The presence of separate peaks correspond to one composition demonstrate the potential use of this method. The combination of α(2,3) and α(2,6) linkages of each single peak was depicted as above. All the ions contain one sodium atom.

Verification of Sialic Acid in N-Glycans of A549 and Detroit 562 Cell Lines

Both A549 and Detroit 562 cell lines have never been studied relative to their glycosylation pattern. The presence of sialic acid was first verified prior to investigate the Sial-Gal linkage types. N-glycans were enzymatically
released from both cell lines and separated into neutral and acidic fractions using a PGC cartridge. Permethylation was carried out in order to remove sialic acid negative charge effect under positive ion MALDI-MS. Both cell lines express a series of complex-type N-glycans containing multiple sialic acid residues (Figure 4.7). In Detroit 562 cell line (Figure 4.7A), the major component was tri-antennary glycan (m/z 3602.8) with three sialic acids. Some other peaks (m/z 3963.9, 3357.7, 3241.7, 2966.4, 2792.4, and 2605.3) were observed containing one to four sialic acids as determined by their mass compositions. In A549 cell lines, the abundant acidic peak was also the tri-antennary glycan (m/z 3602.8) with three sialic acid residues. But the second abundant peak was a di-antennary acidic glycan (m/z 2792.4), which was only a minor component in Detroit 562 cell lines. Some other low abundant acidic glycans (m/z 2185.0, 2430.8, 2605.3, 2880.4, 3037.5, 3241.7, 3357.7 and 3963.9) were also detected as annotated (Figure 4.7B).

From the above data, it was clear that both of cell lines consist of many sialylated glycans, which was in agreement with the fact that both cell lines have been used clinically to culture influenza viruses. The cartoons of each peak in the spectra simply represent the most plausible topology according to the biological synthetic pathway, and have not been fully characterized by MS^n analysis.
Figure 4.7. Positive MALDI-TOF mass spectra of N-glycans from Detroit 562 cell line (A) and A549 Cell line (B) after permethylation. The cartoons represent the most plausible topology for each peak according to the biological synthetic passway. The detailed structures have not been characterized by MS\textsuperscript{n} techniques. All the ions contain one sodium atom.
Both $\alpha(2,3)$ and $\alpha(2,6)$ linkages Detected in Detroit 562 Cell Lines

Detroit 562 cell lines were isolated from human pharynx epithelial cells. The presence of Sial\(^\alpha(2,3)\)Gal residue was an important question to evaluate the likelihood of avian influenza virus spreading in human. Thus, the acidic $N$-glycans from Detroit 562 cell lines were reacted with DMTMM/MeOH and MALDI-MS was obtained (Figure 4.8). Although the abundant peak ($m/z$ 2114.7) represented a single sialic acid glycan with $\alpha(2,6)$ linkage, significant portions of $\alpha(2,3)$ linkages were also found, especially for those large multiply sialylated glycans, $m/z$ 2879.9, 2912.0, 3185.2, and 3217.2, (Figure 4.8).

Since Sial\(^\alpha(2,6)\)Gal will not be able to form the lactone product, the unique $m/z$ of each peak indicate the presence of Sial\(^\alpha(2,3)\)Gal epitope unambiguously in Detroit 562 cell line.
Figure 4.8. Positive MALDI-TOF mass spectrum obtained from the acidic N-glycans of Detroit 562 cell lines after derivatization with DMTMM. The combination of $\alpha(2,3)$ and $\alpha(2,6)$ linkages of each single peak was depicted in the figure. All the peaks contain one sodium atom.
Figure 4.9. Sial-Gal linkage was determined on acidic N-glycans and O-glycans from A549 cell line via MS^n fragment Gal B ion comparison to sialylated carbohydrate standards. Internal Gal B ion spectrum (A) from 3'-sialyllactose; (B) from 6'-sialyllactose; (C) N-glycan of m/z 1812.9 from A549 Cell line; (D) O-glycan of m/z 895.6 from A549 Cell line. All the ions contain one sodium atom.

**Sialic Acid Linkage determination in A549 Cells by MS^n**

A549 cell line originated from the human lower respiratory tract. Both $\alpha_{(2,3)}$ and $\alpha_{(2,6)}$ linkages were expected as previously detected by the lectin methods [Shinya 2006]. MS^n disassembly was chosen to analyze the sialic acid linkages for this model cell line. Both N- and O-glycans were
released and permethylated. The internal galactose B ions (1,2 ene and 3/6-open hydroxyl group) from two standards (2,3-sialyllactose and 2,6-sialyllactose) have distinct spectra (Figure 4.9A and 4.9B). The most abundant peaks were fragment ions $m/z$ 153 and 125 from 3-sialyllactose, and the relative ratio between them was around 42% (Figure 4.9A). In contrast, the two most abundant peaks were ions $m/z$ 197.1 and 139 from 6-sialyllactose, and the relative ratio between them was around 70% (Figure 4.9B). Moreover there are two low abundant peaks ($m/z$ 111.1 and 212.0) in the spectrum, which was different from 2,3 linkage fragmentation.

The most abundant composition from acidic N-glycans of this cell line ($m/z$ 3602.8, Figure 4.7B) was sequentially disassembled and the fragment spectrum of the internal Gal B ion was shown (Figure 4.9C). The most abundant ions $m/z$ 153 and 125 clearly indicated the presence of $\alpha (2,3)$ linkage. Since the spectrum was not similar to the standard 3-sialyllactose spectrum (Figure 4.9A), it suggested that a mixture of both $\alpha (2,3)$ and $\alpha (2,6)$ linkage types, but the majority should be $\alpha (2,3)$ linkage. The simplest O-glycan containing one sialic acid residue (Sial-Gal-GalNAc, $m/z$ 895.6) was chosen to be disassembled. The spectrum was more similar to the standard $\alpha (2,3)$ spectrum such as in Figure 4.9A: suggesting the presence of $\alpha (2,3)$ linkage in O-glycan (Figure 4.9D). Therefore, Sial$\alpha (2,3)$Gal linkage was detected in both N- and O-glycans from A549 cell line.
4.4 DISCUSSION

Sialic Acid Linkage Determination by DMTMM Approach

As demonstrated above, DMTMM derivatization provides a unique way to differentiate \( \alpha (2,3) \) and \( \alpha (2,6) \) Sial-Gal linkages because the C-4 hydroxyl group in the galactose prefers to form a six member-ring lactone with the carboxyl group of sialic acid in Sial\( \alpha (2,3) \) Gal moiety (Scheme 4.1). While, no such hydroxyl group in galactose has the same ability in Sial\( \alpha (2,6) \) Gal moiety to form a lactone. The only product for \( \alpha (2,6) \) linkage is a methyl-ester (Scheme 4.1). The mass difference between one methyl-ester moiety and one lactone moiety is 32 Da, distinguishable by mass analysis. Although Sial\( \alpha (2,3) \)Gal may not be converted into the lactone quantitatively from the standard carbohydrates, the remaining starting material was insignificant compared to the products. Also no significant starting material was observed in fetuin and Detroit 562 samples. Therefore, the DMTMM approach seems the most definite and straightforward approach to determine the presence of Sial\( \alpha (2,3) \)Gal linkage.

Relative Quantification of Sial-Gal Linkages

Additional information usually encountered is the relative ratio of each linkage \( \alpha (2,3) \) and \( \alpha (2,6) \) in a specific glycoprotein, or cell, or organism. Without modification of carboxylic acid group, glycans with native sialic acid may form complex fragmentation peaks under positive ion mass
spectrometry, and have different ionization efficiency under negative ion mode mass spectrum due to the various carboxylic acid groups in each individual glycan, making the calculation of relative quantification impossible. The common methylation or permethylation methods provide the stable signal of sialic acid under positive ion mode of mass spectrometry [Powell 1996], which enable the relative quantification of each glycan. However, the evaluation of relative ratio of α (2,3) and α (2,6) linkage was not possible. With this new DMTMM approach, the quantification of sialic acid linkage seems feasible. Once Sial-Gal residue was converted into lactone or methyl-ester, the ionization efficiency of each glycan can be assumed to be approximately the same regardless of the number of sialic acid residues of each glycan. It is worthy to point out that the lactone and methyl-ester may have slight difference in the ionization efficiency, but this difference should be small. Therefore, it provides a structural basis to approximate the relative ratio of specific linkage of sialic acid.

Given that requirement, the first example was to evaluate the ratio of acidic N-glycans from bovine fetuin (Figure 4.6). Since the signal intensity in MALDI spectra is in proportion to mole abundance [Harvey 1999], the relative ratio of each glycan can be approximated from individual signal intensities. As for glycans with multiple sialic acid residues, the intensities of those glycans were multiplied with the number of each specific linkage of sialic acids. For instance, m/z 3184.4 represents a tetra-sialic acid glycan with two
\( \alpha(2,3) \) residues and two \( \alpha(2,6) \) residues (Figure 4.6). Thus, the relative intensity of \( \alpha(2,3) \) linkage in this peak was measured by \( 9.5 \times 2 = 19 \). All the peaks for acidic \( N \)-glycans from fetuin were re-calculated and summarized in Table 4.1. The total relative intensity of \( \alpha(2,3) \) and \( \alpha(2,6) \) linkages is 71.8 and 118.7, respectively (Table 4.1, last row). So the relative ratio of \( \alpha(2,3) \) linkage in the total sialylated glycans from this glycoprotein was calculated as 37.7\% \( \frac{71.8}{71.8+118.7} \times 100\% \). Although this calculation has not been fully tested with standard samples, it does provide a simple way to approximate the relative quantity of sialic acid specific linkage, which is usually challenging to obtain by other methods.

Applying this strategy, the Sial-Gal linkage ratios were calculated and summarized for acidic \( N \)-glycans of Detroit 562 cell line (Table 4.2). The total relative intensity of \( \alpha(2,3) \) linkage is 173, and 299 for total sialic acid glycans (Table 4.2). Thus the ratio of \( \alpha(2,3) \) linkage in the total sialic acid residues for this cell line is 36.7\% \( \frac{173}{173+299} \times 100\% \). According to this calculation, Detroit 562 cell line has significant Sial\( \alpha(2,3) \)Gal epitopes.

Fragment ion comparison method using MS\(^n\) will not be able to provide such information. Since permethylation is required for detail structural analysis on complex glycans. The advantage of MS\(^n\) fragment ion comparison method will be suitable for the limited sample amount, in which detail structural analyses is also needed.
Table 4.1. Measuring relative ratio of Sial\(\alpha(2,3)\)Gal and Sial\(\alpha(2,6)\)Gal residues following derivatization by DMTMM of the acidic \(N\)-glycans from bovine fetuin.

<table>
<thead>
<tr>
<th>(m/z)</th>
<th>Signal Intensity</th>
<th># of ((\alpha2,3))</th>
<th># of ((\alpha2,6))</th>
<th>Relative Intensity of ((\alpha2,3))</th>
<th>Relative Intensity of ((\alpha2,6))</th>
</tr>
</thead>
<tbody>
<tr>
<td>3216.4</td>
<td>8.3</td>
<td>1</td>
<td>3</td>
<td>8.3</td>
<td>24.9</td>
</tr>
<tr>
<td>3184.4</td>
<td>9.5</td>
<td>2</td>
<td>2</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>2943.3</td>
<td>3.5</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>10.5</td>
</tr>
<tr>
<td>2911.3</td>
<td>15</td>
<td>1</td>
<td>2</td>
<td>15</td>
<td>30</td>
</tr>
<tr>
<td>2879.5</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>20.0</td>
<td>10.0</td>
</tr>
<tr>
<td>2638.4</td>
<td>2.5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2606.5</td>
<td>3.7</td>
<td>1</td>
<td>1</td>
<td>3.7</td>
<td>3.7</td>
</tr>
<tr>
<td>2273.5</td>
<td>4.9</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>9.8</td>
</tr>
<tr>
<td>2241.5</td>
<td>5.8</td>
<td>1</td>
<td>1</td>
<td>5.8</td>
<td>5.8</td>
</tr>
<tr>
<td>(\Sigma)</td>
<td></td>
<td></td>
<td></td>
<td>71.8</td>
<td>118.7</td>
</tr>
</tbody>
</table>

Table 4.2. Measuring relative ratio of Sial\(\alpha(2,3)\)Gal and Sial\(\alpha(2,6)\)Gal residues following derivatization by DMTMM of the acidic \(N\)-glycans from Detroit 562 cell line.

<table>
<thead>
<tr>
<th>(m/z)</th>
<th>Signal Intensity</th>
<th># of ((\alpha2,3))</th>
<th># of ((\alpha2,6))</th>
<th>Relative Intensity of ((\alpha2,3))</th>
<th>Relative Intensity of ((\alpha2,6))</th>
</tr>
</thead>
<tbody>
<tr>
<td>3217.2</td>
<td>13</td>
<td>1</td>
<td>3</td>
<td>13</td>
<td>39</td>
</tr>
<tr>
<td>3185.2</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>2912</td>
<td>21</td>
<td>1</td>
<td>2</td>
<td>21</td>
<td>42</td>
</tr>
<tr>
<td>2879.9</td>
<td>23</td>
<td>2</td>
<td>1</td>
<td>46</td>
<td>23</td>
</tr>
<tr>
<td>2607.1</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2419.8</td>
<td>18</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>2387.9</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>2273.8</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>2241.8</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>2114.7</td>
<td>47</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>47</td>
</tr>
<tr>
<td>2082.7</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>1969.7</td>
<td>13</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>(\Sigma)</td>
<td></td>
<td></td>
<td></td>
<td>173</td>
<td>299</td>
</tr>
</tbody>
</table>
Avian Influenza Virus Transmission in Human

There is no doubt about the presence of Sialα (2,6)Gal residues in the human respiratory tract. Whether the presence of Sialα (2,3)Gal linkage in the upper respiratory tract or not is questionable, which was also the main purpose of this research project. Although common lectin method hasn’t detect any α (2,3) linkage in the human upper respiratory tract [Shinya 2006; Nicholls 2007b], avian influenza virus can grow efficiently in the upper respiratory tissue and primary cell lines in vitro [Nicholls 2007a]. Therefore it does suggest that there are avian virus receptors in this area, which are in agreement with our results using Detroit 562 cells which have significant Sialα (2,3)Gal residues, and account for 36.7% of total sialic acid residues by the DMTMM method.

Some recent data suggests that the mechanism of viral entry in human cells is highly complicated, and can’t be simply explained by whether the presence of proper HA receptors or not. For example, in the infection of macrophages cells, the viruses must undergo an extra lectin-like interaction with host mannose receptors after the initial sialic acid binding [Reading 2007]. Furthermore, some cell lines deficient of N-glycosylation (due to a mutation in the N-acetylglucosaminyltransferase I gene) show viral binding to the cell (may bind to sialic acid units of O-glycans or glycolipids of host cells), but no viral entry. These results suggest that influenza virus may require one or more N-linked glycoproteins as a cofactor for cell entry after the initial sialic
acid recognition and attachment [Nizet 2008]. Another interesting report stated that human virus HA prefers binding to Sialα (2,3)Gal moiety with extended chain residues, but not avian virus, such as H5N1 [Chandraselcaran 2008]. In addition, avian influenza virus was found to prefer 32 °C for efficient infection, and have limited infection ability at 37 °C [Scull 2009]. Therefore, these data suggest that the poor transmission of avian virus in human may arise from some other reasons, rather than the absence of the Sialα (2,3)Gal moiety.

In addition to HA binding, NA also plays a key role in the virus infection cycle as discussed in Chapter 3. Newly formed viruses must be released by NA from the infected host cells in order to spread to other cells and tissues. Therefore, the establishment of a productive virus infection is dependent on a delicate balance between both HA and NA. Unfortunately, there are no studies regarding the balance between HA and NA for avian influenza virus. In a word, the effective human-to-human infection of influenza virus is a highly complex situation. The presence of proper HA receptors may be one of the important factors, but certainly not the only determinant.

**Future Experiments**

Theoretically sialic acid galactose linkage investigation should be carried out on real human respiratory epithelial tissues like the above lectin approach. Unfortunately, human tissues are impractical for mass
spectrometry glycan analysis due to either the limited sample amount or strict law regulations on human subjects. Thus, primary human respiratory epithelial cell lines are the only choice which directed their research using Detroit 562 and A549 cell lines. Since HA is also a kind of lectin, and many reports have demonstrated internal residues participation in the lectin recognition, besides of the terminal Sial-Gal moiety [Chandraselcaran 2008; Stevens 2006b]. Therefore, a detail structural analysis of all major acidic glycans via MS^n could clarify this issue, especially those internal residues of the antenna ending with sialic acid. This may be a key factor to evaluate viral HA binding. For comparison, real tissues from other animals could also be carried out to analyze the glycan structures, such as domestic chicken, duck, pig, and the well-recognized animal model for influenza research: the ferret.
LIST of REFERENCES


Harvey, D. J. (2009b) “Application of negative ion MS/MS to the identification of N-glycans released from carcinoembryonic antigen cell adhesion molecule 1” J. Mass Spectrom. 44, 50 – 60.


World Health Organization (WHO) http://www.who.int/topics/avian_influenza/en/


APPENDICES
Figure A.1. Negative ion $MS^2$ and $MS^3$ of glycan 1 (See pp 28).
Figure A.2. GC chromatography of GlcA and GalA standards retention time comparison to EP Glycans. The same retention time 12.13 between EP glycans and GlcA confirm the GlcA identity in EP glycans (See pp 29, 32).
Figure A.4. GC chromatography of Fuc and Man standards retention time comparison to EP Glycans (See pp 29, 32).
Figure A.5. GC chromatography of GlcNAc and GalNAc standards retention time comparison to EP Glycans (See pp 29, 32).
Figure A.6. GC-MS EI spectral comparison of Fuc standard to EP Glycans (See pp 29, 32).
Figure A.7. GC-MS EI spectral comparison of Man standard to EP Glycans (See pp 29, 32).
**Figure A.8.** GC-MS EI spectral comparison of GlcNAc standard to EPG Glycans (See pp 29, 32)
Figure A.9. Negative ion MS$^2$ spectral comparison of glycan 3 to glycan 1 (See pp 45, 46).
Figure A.10. Negative ion MS^n spectra to show distinguished D ion and E ion of glycan 2 (See pp 40).
Figure A.11. Negative ion MS\textsuperscript{n} spectra to show distinguished D ion and E ion of glycan 3 (See pp 45, 46).