Glycomic Analysis of Biomedically Important O-glycoconjugates

A thesis submitted for the Degree of Doctor of Philosophy of
Imperial College London

Submitted by

Mohd Nazri Ismail

Imperial College London
Division of Molecular Biosciences
Department of Life Sciences
Faculty of Natural Sciences
South Kensington
London SW7 2AZ
United Kingdom
Declaration

I hereby declare that the work presented in this thesis has not been previously or concurrently submitted for any other degree, diploma or other qualification at any other university, and it is the result of my own independent research unless otherwise indicated.

Mohd Nazri Ismail
Acknowledgements

This PhD work would not have been possible without the support of many people. I would like to express my deepest gratitude to my supervisor, Prof. Anne Dell who was abundantly helpful and offered invaluable assistance, support and guidance. Deepest appreciation is also due to my second supervisor, Dr. Stuart Haslam, without whose knowledge and assistance this thesis would not have been successful.

Special thanks also to all former and current group members, Prof. Howard Morris, Dr. Maria Panico, Dr. Berangere Tissot, Dr. Mark Sutton-Smith, Dr. Aristotelis Antonopoulos, Alana Trollope, Rebecca Harrison, Kevin Canis, Dr. Poh-Choo Pang, Dr. Paul Hitchen, Dr. Rositsa Karamanska, David Damerell, Dr. Simon Parry, Dr. Alessio Ceroni and Dr. Simon North, for sharing knowledge and invaluable assistance. Not forgetting to my friends who always been there and making my life journey in this wonderful but challenging city of London much smoother.

It was an honour for me to collaborate with some of the best scientists in this field, including Prof Jamey Marth, Dr. Erica Stone, Prof Minoru Fukuda, Dr. Celso Reis, Ana Magalhaes and Dr. Frederic Bard.

I would also like to convey my appreciation to the Malaysian Ministry of Higher Education and the Universiti Sains Malaysia for providing the financial means throughout my study.

Lastly, I wish to express my love and gratitude to my beloved families; for their understanding & endless love, through the duration of my study.

Thank you - تَرْمَا كَأْسِيْهُ - Alhamdulillah
Abstract

It has been established that all cells carry an array of glycans attached to proteins and lipids that are crucial in the interaction between cells and the surrounding matrix. Proteins are mainly glycosylated on asparagines (N-glycosylation) and serine or threonine residues (O-glycosylation). Compared to N-glycans, O-glycans offer a higher degree of structural ambiguity due to the existence of several types and cores. This is believed to contribute to the relative lack of knowledge on these molecules. Therefore, improvement to the current methodologies of structural studies is a prerequisite to complement the immense functional findings of O-glycoconjugates in biological systems. This thesis discusses the structural characterisation, regulation and biological roles of O-glycans. The overall aim was to optimise O-glycomic mass spectrometric analysis to help illuminate the phenotypic findings from our collaborators in three separate but related projects. The methodologies utilised involving MALDI-TOF/TOF-MS, GC-EI-MS, ESI-QTOF-MS and MALDI-QIT-TOF-MS.

The first project investigated the effects of core 2 GlcNAc transferase (C2GnT) deficiency in mice. This enzyme exists in three isoforms which are expressed differently in different tissues. Analysis of the single knockout of each of these isoenzymes as well as the triple knockouts has allowed the investigation of their unique and overlapping functions. The outcomes of this study include characterisation of alterations of not just mucin-type O-glycans but also O-mannose glycan, which could be associated with several organ lesions and system failures. The second project focused on the gastric mucosa of mice with deficiency in α1,2-fucosyltransferase (FuT2). This enzyme plays an important role in decorating the mucosal mucins with ABH-blood group and Lewis antigens that are known to interact with various gut flora including the pathogen *Helicobacter pylori*. It has been shown that the binding of *H. pylori* via BabA adhesins was significantly impaired with the loss of H antigens and Lewis y on O-glycans. The third project investigated the regulation of mucin-type O-glycosylation. The protein Src has been recognised to play an essential role in the localisation of ppGalNAc transferases, the initiating enzyme of O-glycosylation, in the endoplasmic reticulum and Golgi apparatus. Therefore, it could be inferred that Src influences the regulation of protein O-glycosylation. The NIH3T3 and NBT-II cell lines with different levels of Src or different localisation of ppGalNAcT-2 have been analysed in order to identify the changes on the structures of O-glycans and the relative abundances of cores 1 and 2. Valuable information has been gathered which could lead to further investigative work to better understand the role of Src in the regulation of protein O-glycosylation.
Thesis publications


Contents

DECLARATION .......................................................................................................................... 2

ACKNOWLEDGEMENTS .......................................................................................................... 3

ABSTRACT .............................................................................................................................. 4

THESIS PUBLICATIONS ....................................................................................................... 5

CONTENTS ............................................................................................................................. 6

INDEX OF FIGURES .................................................................................................................. 9

INDEX OF TABLES .................................................................................................................. 12

1 INTRODUCTION .......................................................................................................................... 16

1.1 Sweet interactions: biomedically important glycoconjugates ................................................. 16

1.2 Protein glycosylation ........................................................................................................... 19

1.2.1 N-glycosylation .............................................................................................................. 20

1.2.2 O-glycosylation .............................................................................................................. 21

1.2.2.1 β-O-galactose, α-O-fucose, β-O-glucose and α-O-mannose ........................................... 22

1.2.2.2 Mucin type O-glycans .................................................................................................. 24

1.2.2.3 Core 2 β1,6-N-acetylglucosaminyltransferase ............................................................. 26

1.2.3 Glycan extension, branching and terminal epitopes ......................................................... 29

1.2.3.1 Lewis antigens .......................................................................................................... 32

1.2.3.2 Functional significance of fucosyltransferases ............................................................ 33

1.2.3.3 Carbohydrate recognition by lectins ......................................................................... 35

1.2.4 Mucins ............................................................................................................................ 40

1.2.5 Helicobacter pylori: a specialist in human stomach colonisation .................................. 44

1.2.6 Regulation of protein O-glycosylation .......................................................................... 46

1.2.7 Glycomics and glycoproteomics .................................................................................... 52

1.3 Animal models in functional glycomic research .................................................................. 53

1.3.1 Mouse models .............................................................................................................. 54

1.4 Mass spectrometry: principles and applications ................................................................ 55

1.4.1 History .......................................................................................................................... 55

1.4.2 Components ................................................................................................................. 56

1.4.3 MALDI-TOF/TOF mass spectrometry ........................................................................ 58

1.4.4 ESI-QTOF mass spectrometry ...................................................................................... 61

1.4.5 GC-EI mass spectrometry ............................................................................................ 63

1.4.6 MALDI-QIT-TOF mass spectrometry ........................................................................ 64

1.4.7 Other prominent mass spectrometry instrumentation .................................................. 66

1.4.8 Strategies for glycan structural analysis ......................................................................... 67

1.4.8.1 Release of oligosaccharides from glycoproteins .......................................................... 68

1.4.8.2 Chemical and enzymatic digestions ......................................................................... 70

1.4.8.3 Permethylation .......................................................................................................... 70

1.4.8.4 Glycan fragmentation ................................................................................................ 72

1.4.8.5 Glycan linkage analysis ............................................................................................. 74

1.5 Thesis aims ......................................................................................................................... 76

2 MATERIALS AND METHODS .................................................................................................... 78

2.1 Materials ............................................................................................................................ 78

2.1.1 Biological samples ......................................................................................................... 78

2.1.1.1 High sensitivity O-glycomic analysis of wild type and C2GnT knockout murine tissues .......................................................................................................................... 78

2.1.1.2 O-glycosylation profiling of gastric mucosa from FuT2-null mice in relation to Helicobacter pylori adhesion ........................................................................................................ 78

2.1.1.3 Investigation of cellular regulation of protein O-glycosylation by Src ................................ 79
## 2 Methods

### 2.1 General chemicals and reagents

### 2.2 Enzymes, antibodies and proteins/peptides

### 2.3 Labware and Equipment

### 2.4 Isolation of glycopeptides/peptides

- **2.4.1** Tissue homogenisation (without glycolipid removal)
- **2.4.2** Tissue homogenisation (with glycolipid removal)
- **2.4.3** Cell sonication
- **2.4.4** Protein reduction and carboxymethylation
- **2.4.5** Cleavage into peptides and glycopeptides

### 2.5 Digestion of N-glycans from glycopeptides

### 2.6 Release of O-glycans from glycopolypeptides

### 2.7 Permethylation and purification for non-sulphated glycan analysis

### 2.8 Permethylatation and purification for sulphated O-glycan analysis

### 2.9 Fractionation of sulphated O-glycans

### 2.9.1** Preparation of matrices

### 2.9.2 MALDI-TOF/TOF mass spectrometry and tandem mass spectrometry

### 2.9.3 ESI-QTOF tandem mass spectrometry

### 2.9.4 GC-EI mass spectrometry

### 2.9.5 MALDI-QIT-TOF mass spectrometry and tandem mass spectrometry

## 3 HIGH SENSITIVITY O-GLYCOMIC ANALYSIS OF WILD TYPE AND C2GNT KNOCKOUT MURINE TISSUES

### 3.1 Background

### 3.2 Results

### 3.3 Discussion

## 4 O-GLYCOSYATION PROFILING OF GASTRIC MUCOSA FROM FUT2-NULL MICE IN RELATION TO *HELICOBACTER PYLORI* ADHESION

### 4.1 Background

### 4.2 Results

- **4.2.1** Mass spectrometric strategy
- **4.2.2** Glycomic profiling of wild type stomach and mucosal scrapings
- **4.2.3** Glycomic profiling of FuT2-null stomach and mucosal scrapings
- **4.2.4** Differentiation of fucosylated histo-blood group antigens
- **4.2.5** Determination of N-acetyllactosamine unit type
- **4.2.6** Glycomic analysis of immunopurified stomach lysates

### 4.3 Discussion
5 INVESTIGATION OF CELLULAR REGULATION OF PROTEIN O-GLYCOSYLATION BY SRC

159

5.1 Background ........................................................................................................................ 159
5.2 Results .................................................................................................................................. 159
  5.2.1 Mass spectrometric strategy ........................................................................................ 159
  5.2.2 O-Glycomic analysis of NIH3T3 cells without N-glycan digestion ............................. 160
  5.2.3 O-Glycomic analysis of NIH3T3 cells with N-glycan digestion ..................................... 163
  5.2.4 O-Glycomic analysis of NBT-II cell lines ................................................................. 165
  5.2.5 Trimming NBT-II cell glycomes to basic core 1 and core 2 structures ......................... 170
5.3 Discussion .......................................................................................................................... 173

6 CONCLUDING REMARKS ........................................................................................................ 180

APPENDIX I ........................................................................................................................................ 186

REFERENCES .................................................................................................................................... 187
Index of figures

Figure 1.1. Schematic representation of types of interactions mediated by extracellular lectins. ........ 18
Figure 1.2. Symbol nomenclature for the most common monosaccharides found on mammalian
glycoproteins. ................................................................................................................. 20
Figure 1.3. Summary of the biosynthetic pathways of N-glycans ........................................... 22
Figure 1.4. The biosynthetic pathways of mucin type O-glycans ........................................... 26
Figure 1.5. Relative expression levels of C2GnT2 and C2GnT3 in various murine tissues. ........... 27
Figure 1.6. The biosynthesis and elongation of I-antigen/I-branch ......................................... 30
Figure 1.7. The most common terminal epitopes found on mammalian glycans. ..................... 31
Figure 1.8. Schematic representation of typical mucin domains ............................................. 40
Figure 1.9. H. pylori adherence to gastric mucosa prior to colonisation on epithelial cells. .......... 46
Figure 1.10. Interaction of the most studied H. pylori adhesins with host cell receptors leading to the
translocation of CagA. ........................................................................................................ 46
Figure 1.11. Biosynthesis of N-glycans, O-GalNac and O-mannose on proteins from endoplasmic
reticulum (ER) to trans-Golgi ....................................................................................... 48
Figure 1.12. COP-I-mediated retrograde and COP-II-mediated anterograde transportation between
cis-Golgi and ER ............................................................................................................ 49
Figure 1.13. MALDI-TOF mass spectra of O-glycans derived from (TOP) cells with inactive Src (SYF)
and (BOTTOM) cells with active Src (SYFsrc) .............................................................. 51
Figure 1.14. Illustrative model for Src-induced COP-I-mediated retrograde trafficking of ppGalNacTs.
 ........................................................................................................................................... 52
Figure 1.15. Schematic view of a MALDI-TOF/TOF mass spectrometer .................................. 59
Figure 1.16. A schematic diagram of an ESI-QTOF-MS ..................................................... 61
Figure 1.17. Schematic view of the ion source based on electron ionisation and the single quadrupole
mass analyser typically found in a GC-MS instrument .................................................. 64
Figure 1.18. Schematic diagram of a MALDI-QIT-TOF-MS ................................................... 65
Figure 1.19. Reaction steps for reductive elimination of O-linked GalNAc ............................... 69
Figure 1.20. Common fragmentation pathways of permethylated glycans ............................... 74
Figure 1.21. The conversion of permethylated oligosaccharides to PMAAs ............................. 75
Figure 2.1. Generation of mice deficient in all three C2GnT genes .......................................... 79
Figure 2.2. Western blotting of Leα and Muc5AC immunoprecipitated gastric mucosa from wild type
and FuT2-null mice after gel electrophoresis ............................................................... 80
Figure 3.1. MALDI-TOF/TOF-MS and MS/MS spectra of the molecular ion m/z 1331 [M+Na]+ detected
in the mass fingerprinting of the stomach from wild type and C2GnT2 knockout mice .......... 94
Figure 3.2. MALDI-TOF-MS profile of O-glycans from the colon of wild type and knockout mice. ... 95
Figure 3.3. The full range MALDI-TOF-MS spectra of permethylated O-glycans from the colon of wild
type and knockout mice ................................................................................................. 101
Figure 3.4. MALDI-TOF-MS profile of sulphated O-glycans from the colon of wild type and C2GnT2
knockout mice .............................................................................................................. 103
Figure 3.5. The MALDI-TOF/TOF-MS/MS spectra of the sulphated O-glycans showing the unusual reducing GalNAc sulphation. ................................. 104
Figure 3.6. MALDI-TOF-MS spectra of O-glycans from the stomach of wild type and knockout mice. .......................................................... 105
Figure 3.7. The full range MALDI-TOF-MS spectra of permethylated O-glycans from the stomach of knockout mice. .................................................. 111
Figure 3.8. MALDI-TOF-MS profile of O-glycans from the small intestine of wild type and knockout mice. .......................................................... 113
Figure 3.9. The full range MALDI-TOF-MS spectra of permethylated O-glycans from the small intestine of wild type and knockout mice. .................................................. 118
Figure 3.10. MALDI-TOF-MS profile of O-glycans from the thyroid/trachea of wild type and knockout mice. .......................................................... 120
Figure 3.11. MALDI-TOF-MS profile of O-glycans from the kidneys of wild type and knockout mice. .......................................................... 121
Figure 3.12. MALDI-TOF-MS profile of O-glycans from the thymus of wild type and knockout mice. 122
Figure 3.13. Determination and verification of the I-branching site and position. .................................................................................. 123
Figure 3.14. Structural confirmation of the RM2-like epitope found in the colon tissues. .................................................................................. 124
Figure 4.1. The MALDI-TOF-MS O-glycomic profile of reduced and permethylated O-glycans [M+Na]+ detected in the stomach of wild type mice. ........................................................................ 138
Figure 4.2. The MALDI-TOF-MS O-glycomic profile of reduced and permethylated O-glycans [M+Na]+ detected in the mucosal scrapings of wild type mice. ........................................................................ 140
Figure 4.3. The MALDI-TOF-MS O-glycomic profile of reduced and permethylated O-glycans [M+Na]+ detected in the stomach of FuT2-null mice. ........................................................................ 142
Figure 4.4. The MALDI-TOF-MS O-glycomic profile of reduced and permethylated O-glycans [M+Na]+ detected in the mucosal scrapings of FuT2-null mice. ........................................................................ 143
Figure 4.5. The MALDI-TOF/TOF-MS/MS spectrum of the molecular ion m/z 708 [M+Na]+ from the stomach of FuT2-null mice. .................................................................................. 144
Figure 4.6. The ESI-QTOF-MS/MS spectrum of glycans at m/z 590 [M+2Na]+ from the stomach of (A) wild type and (B) FuT2-null mice. .................................................................................. 145
Figure 4.7. The MALDI-QIT-TOF analysis of the glycans at m/z 1157 [M+Na]+ from the stomach of FuT2-null mice. .................................................................................. 146
Figure 4.8. The MALDI-TOF/TOF-MS/MS spectrum of glycan at m/z 2578 [M+Na]+ with the putative Lea antigen from the stomach of wild type mice. .................................................................................. 147
Figure 4.9. Determination and verification of N-acetyllactosamine unit type. .................................................................................. 148
Figure 4.10. The O-glycomic profile of permethylated glycans from the stomach of wild type mice before and after β1,4-galactosidase digestion. .................................................................................. 149
Figure 4.11. The MALDI-TOF-MS O-glycomic profile of reduced and permethylated O-glycans [M+Na]+ detected in the experimental control murine IgA bands. .................................................................................. 150
Figure 5.1. MALDI-TOF-MS spectra of O-glycans from 3T3 WT (1) cells without N-glycan removal 161
Figure 5.2. MALDI-TOF-MS spectra of O-glycans from 3T3 vSrc (1) cells without N-glycan removal. .............................................................................................................................................................................................. 162
Figure 5.3. MALDI-TOF-MS spectra of O-glycans from 3T3 WT (2) cells with N-glycan removal .... 164
Figure 5.4. MALDI-TOF-MS spectra of O-glycans from 3T3 vSrc (2) cells with N-glycan removal.... 165
Figure 5.5. MALDI-TOF-MS spectra of O-glycans from NBT-II WT (1) cells................................. 166
Figure 5.6. MALDI-TOF-MS spectra of O-glycans from NBT-II WT Golgi (1) cells ..................... 167
Figure 5.7. MALDI-TOF-MS spectra of O-glycans from NBT-II WT ER (1) cells.......................... 168
Figure 5.8. MALDI-TOF-MS spectra of O-glycans from NBT-II H2 ER (1) cells.......................... 169
Figure 5.9. Combined O-glycomic spectrum of NBT-II cell lines showing only major peaks. ...... 170
Figure 5.10. MALDI-TOF-MS spectra of enzyme- and TFA-treated NBT-II cell lines. .............. 171
Figure 5.11. MALDI-TOF-MS spectra of TFA-treated and non-treated wild type kidneys.......... 172
Index of tables

Table 1.1. Fucosyltransferases in mice and humans with their respective activities...........................34
Table 1.2. Families of the most common murine lectins and their properties. .....................................36
Table 1.3. List of known murine mucin genes with respective human orthologues and expression sites. .............................................................................................................................................42
Table 3.1. List of tissues analysed in this study derived from the wild type (WT), C2GnT2 KO (KO2), C2GnT3 KO (KO3) and C2GnT triple KO (TKO) mice. .................................................................................................................................92
Table 3.2. O-glycan structures from the colon of wild type and knockout mice ...................................96
Table 3.3. O-glycan structures from the stomach of wild type and knockout mice ..............................106
Table 3.4. O-glycan structures from the small intestine of wild type and knockout mice. ....................114
Table 4.1. List of all samples analysed in this study ........................................................................135
Table 5.1. List of cell line samples analysed and their descriptions ..................................................160
Table 5.2. Summary table of O-glycan structures observed in both sets of experiments of NIH3T3 cell line. ................................................................................................................................................................................163
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>Alpha</td>
</tr>
<tr>
<td>α-DG</td>
<td>Alpha-Dystroglycan</td>
</tr>
<tr>
<td>β</td>
<td>Beta</td>
</tr>
<tr>
<td>µg</td>
<td>Micrograms</td>
</tr>
<tr>
<td>µl</td>
<td>Microlitres</td>
</tr>
<tr>
<td>°C</td>
<td>Degrees centigrade</td>
</tr>
<tr>
<td>Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>Ambic</td>
<td>Ammonium bicarbonate</td>
</tr>
<tr>
<td>Asn</td>
<td>Asparagine</td>
</tr>
<tr>
<td>Asp</td>
<td>Aspartic acid</td>
</tr>
<tr>
<td>BabA</td>
<td>Blood group antigen-binding adhesin</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>C1GalT</td>
<td>Core 1 β1,3-galactosyltransferase</td>
</tr>
<tr>
<td>C2GnT</td>
<td>Core β1,2-N-acetylglucosaminyl-transferase</td>
</tr>
<tr>
<td>CagA</td>
<td>Cytotoxin-associated gene A</td>
</tr>
<tr>
<td>CFG</td>
<td>Consortium for Functional Glycomics</td>
</tr>
<tr>
<td>CHCA</td>
<td>α-cyano-4-hydroxycinnamic acid</td>
</tr>
<tr>
<td>CID</td>
<td>Collision induced dissociation</td>
</tr>
<tr>
<td>COP</td>
<td>Coat protein complex</td>
</tr>
<tr>
<td>CRD</td>
<td>Carbohydrate recognition domain</td>
</tr>
<tr>
<td>Da</td>
<td>Daltons</td>
</tr>
<tr>
<td>DABP</td>
<td>3,4-diaminobenzophenone</td>
</tr>
<tr>
<td>DC-SIGN</td>
<td>Dendritic cell specific intercellular adhesion molecule grabbing non-intergrin</td>
</tr>
<tr>
<td>DSS</td>
<td>Dextran sodium sulphate</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>ECD</td>
<td>Electron capture dissociation</td>
</tr>
<tr>
<td>ETD</td>
<td>Electron transfer dissociation</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediamine tetraacetic acid</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>EI</td>
<td>Electron impact</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionisation</td>
</tr>
<tr>
<td>FAB</td>
<td>Fast-atom bombardment Fourier transform</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>Fuc</td>
<td>Fucose</td>
</tr>
<tr>
<td>FuT</td>
<td>Fucosyltransferase</td>
</tr>
<tr>
<td>FW</td>
<td>Formula weight</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>GalNAc</td>
<td>N-acetylgalactosamine</td>
</tr>
<tr>
<td>GalT</td>
<td>Galactosyltransferase</td>
</tr>
<tr>
<td>GC</td>
<td>Gas chromatography</td>
</tr>
<tr>
<td>Glc</td>
<td>Glucose</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetylglucosamine</td>
</tr>
<tr>
<td>Glu</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine triphosphate</td>
</tr>
<tr>
<td>H. pylori</td>
<td>Helicobacter pylori</td>
</tr>
<tr>
<td>Hex</td>
<td>Hexose</td>
</tr>
<tr>
<td>HexNAc</td>
<td>N-acetylhexitosamine</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>HLys</td>
<td>Hydroxylysine</td>
</tr>
<tr>
<td>IAA</td>
<td>Iodoacetic acid</td>
</tr>
<tr>
<td>ICR</td>
<td>Ion cyclotron resonance</td>
</tr>
</tbody>
</table>
Ig  Immunoglobulin  POMGnT1  Protein-O-mannos-β1,2-
    β1,3-N-acetylglucosaminyl  N-acetylglucosaminyl-
iGnT  transferase  transferase  1
    β-1,6-N-  ppGalNAcT  Polypeptide-α-N-
    acetylglucosaminyl  acetylgalactosaminyl
    transferase  transferase
    KO  Knockout  Pro  Proline
    L  Litres  PTS  Proline/Thrreonine/Serinein
  LacNAc  N-acetyllactosamine  e domain
  LC  Liquid chromatography  Q  Quadrupole
  Le  Lewis  QIT  Quadrupole ion trap
  Leα  Lewis a  QTOF  Quadrupole Time of Flight
  Leβ  Lewis b  RF  Radio frequency
  Le¢  Lewis x  S  Sulphate
  Le£  Lewis y  SabA  sialic acid-binding adhesin
  Leº  Lewis y  SDS  Sodium Dodecyl-sulphate
  Leu  Leucine  SDS  Serine
  Lys  Lysine  Serine  Serine
  M  Molar  SLeα  Sialyl-Lewis a
  m/z  Mass over charge  SLe¢  Sialyl-Lewis x
  MALDI  Matrix assisted laser  ST  Sialyltransferase
desorption ionisation  TFA  Trifluoroacetic acid
  Man  Mannose  Thr  Threonine
  MCP  Multichannel plate  TOF  Time of Flight
  mg  Milligrams  Tris  Tris(hydroxymethyl)amine
  mg/ml  Milligrams/millilitres  methanone
  ml  Millilitres  UDP  Uridine diphosphate
  mM  Millimolars  v/v  Volume/volume
  MS  Mass spectrometry  w/v  Weight/volume
  MS/MS  Tandem mass  WT  Wild type
  spectrometry
  MS"  Multiple stage mass
        spectrometry
  Muc  Mucin
  MW  Molecular weight
  NeuAc  N-acetylneuraminic acid
  NeuGc  N-glycolyneuraminic acid
  PNGaseF  Peptide: N-glycosidase F
  POMT  Protein-O-
        mannosyltransferase
Chapter 1:

Introduction
1 Introduction

1.1 Sweet interactions: biomedically important glycoconjugates

Glycosylation, the enzymatic synthesis of oligosaccharides attached to biomolecules, is an event that reaches beyond the genome and is controlled by factors that differ greatly among cell types and species (Spiro, 2002). Glycosylation produces a diverse and abundant set of covalently linked oligosaccharides, referred to as glycans. This repertoire of glycans is collectively known as the glycome, and is a result of combinatorial expression of, for instance in the mammalian genome, more than 200 glycosyltransferases and glycosidases (Marth & Grewal, 2008). The rising interest in glycoconjugates is largely influenced by the progressive revelations of diseases that can be associated with glycosylation and glycans. Indeed, aberrations in glycosylation can be correlated to more than 30 types of genetic diseases in humans encompassing neurological disorders, congenital muscular dystrophies, connective tissue disorders and immune deficiencies (Freeze, 2006; Hennet, 2009). By acknowledging how distressed biological systems can become due to errors in glycosylation pathways, it is becoming increasingly clear that glycoconjugates play undeniably important roles in ensuring functional integrity. Investigative work on the regulations, structures and functions of glycans spans various research fields including immunology, genetics, microbiology etc. with the hub discipline being known as glycobiology.

Glycobiology is the field that investigates the biological functions of glycans attached to proteins and membranes and determines how these functions are carried out (Taylor & Drickamer, 2006). Glycans, together with nucleic acids, lipids and proteins, are important macromolecules in a living system. Of these four, glycans are the most abundant and structurally diverse biopolymers. Living cells synthesise these molecules, either to be expressed on their surface or to be secreted into the blood stream, in the form of glycoconjugates such as proteoglycans, glycolipids and glycoproteins.

The high abundance of glycans on cell surfaces illustrates their premier position as the interface between cells and interacting cells or molecules (Varki & Sharon, 2009). The structural diversity of carbohydrates underlies the potential of this class of biomolecules for storing biological information. Turning this potential into an operative sugar code entails the existence of efficient decoding devices (Soils et al., 2009). Recognition of glycans by proteins has been shown to be central to a myriad of intra- and extra- cellular physiological and pathological processes.
Glycans interact with various types of proteins including carbohydrate active enzymes and glycan-binding proteins (GBPs). The latter represents the major way in which the information contained in glycan structures is recognised, deciphered and transformed into biological actions (Cummings & Esko, 2009).

A family of GBPs known as lectins constitute the major mediators or modulators of specific glycan biological roles. Lectins were first discovered in plants about 100 years ago but their presence is now known to span most living organisms including animals and microbes (Varki et al., 2009). The sugar binding activity of typical lectins can be ascribed to a specific domain known as the carbohydrate recognition domain (CRD) (Weis & Drickamer, 1996). The classification of lectins was initially based on their CRD sequence motifs, which led to the designation of C-type lectins (requires calcium ion for recognition) and galectins (galactose binding) (Drickamer & Taylor, 1993), both of which are animal lectins. The subsequent discovery of a substantive number of other types of lectins, has resulted in various categories of classification and at the present time, there is no single universally accepted classification method (Varki et al., 2009).

Many lectins are either soluble or membrane integrated proteins with their CRDs located in the extracellular space. These lectins are the counterparts of membrane-bound or secreted glycoconjugates, with both being the mediators of adhesion and signalling events at the cell surface. Upon CRD binding to its substrates, other lectin domains or the substrate itself will mediate responses to the recognition events. This outside-in signalling is similar to transmembrane signalling from hormones and growth factors to their intracellular targets (Taylor & Drickamer, 2006). Outcomes of these signalling processes most commonly involves changes in the levels of gene expression, leading to diverse biological functions. These lectin-ligand interactions can be classified into several types, as illustrated in Figure 1.1. Interaction A in the figure represents antigen capturing by antigen-presenting cells. For instance, dendritic cell-specific intercellular adhesion molecules-3-grabbing non-integrin (DC-SIGN) binds and induces endocytosis of glycosylated pathogens (Svajger et al., 2010). The internalised antigens are then digested and the fragments are presented on the cell surface to interact with other immune cells (Pöhlmann et al., 2001). The B portion of the figure illustrates selectins and their membrane-bound glycoprotein ligands. As an example, E-selectin on endothelial cells transiently binds to P-selectin glycoprotein ligand-1 (PSGL-1) on neutrophils and triggers a signalling cascade through Src family kinases to induce integrin expression, leading to reduced neutrophil rolling and
eventually migration to sites of inflammation (Yago et al., 2010). C and D denote galectin mediated cell-adhesion and glycoprotein receptor clustering, respectively. For instance, galectin 3 regulates adhesive interaction between breast carcinoma cells, via binding to poly-N-acetyllactosamine chains on surface glycoproteins, and elastin on connective tissues (Ochieng et al., 1999). Galectin 3 also plays a critical role in cancer metastasis by clustering Muc1 glycoprotein on cancer cell surface via adhesion to core 1 O-glycans thus exposing the cell surface for adherence to endothelial cells and then migrate (Yu et al., 2007). The various types of animal lectins and their specificities are discussed in more detail in Section 1.2.3.3.

Figure 1.1. Schematic representation of types of interactions mediated by extracellular lectins.

Interaction at A represents lectin-mediated endocytosis leading to antigen presentation on the cell surface. Lectin binding at B, C and D generate signalling cascades that are usually interpreted in the form of gene expression.

Many eukaryotic glycoconjugates directly interact with diverse microbial species by binding to lectins that are known as adhesins. For example, human and animal pathogens including viruses, bacteria, fungi and parasites have been shown to utilise cell surface glycans as
‘keys’ to enter and invade host cells leading to various types of diseases (Corfield et al., 2000; Nizet & Esko, 2009). *Helicobacter pylori* for instance, a common opportunistic bacterium in the human gut, utilises glycan binding adhesins that bind to fucosylated and sialylated glycans (Ilver et al., 1998; Mahdavi et al., 2002). These glycans are very abundant on gastric mucosa specifically on mucin glycoproteins that naturally function to protect the inner epithelial cells from their harsh environment (Corfield et al., 2001). The interaction between *H. pylori* and host glycans is relevant to one of the projects in this thesis (Chapter 4), and therefore will be discussed in more detail in Section 1.2.5.

From the examples given above, it is clear that glycococonjugates are undeniably biomedically important. Nevertheless, there are still many missing pieces of information that are required to assemble a complete picture of the biochemical mechanisms underlying carbohydrate mediated interactions and their functional consequences. This is partly due to the underdevelopment of glycomics technologies relative to proteomics and genomics (Pilobello & Mahal, 2007). Improvements in glycomics methodologies, in parallel with biochemical and phenotypic studies, are essential for better understanding of the structure-function attributes of glycoconjugates in health and disease. Of all glycoconjugates, glycoproteins are the most abundant and can highly influence how a biological system works due to the fact that they include functionally important biomolecules such as hormones, enzymes, antibodies and receptors (Brooks, 2002).

### 1.2 Protein glycosylation

Glycosylation is the most functionally recognised and complex form of protein posttranslational modification. The first glycopeptide linkage was described by Johanssen et al. in 1961 on ovalbumin (Johansen et al., 1961). There are two main types of protein glycosylation depending on the way they are linked to a protein, namely N-glycosylation and O-glycosylation (Kornfeld & Kornfeld, 1976). Both of these types of glycosylation share a number of similar building blocks but are catalysed through different metabolic pathways, resulting in structure and function diversification. At the moment, according to the calculation by the PTM browser (www.ptmbrowser.org) based on the biochemical data from Swiss-Prot (www.uniprot.org) (Consortium, 2010) and Ensembl databases (www.ensembl.org) (Flicek et al., 2010), it has been estimated that about 20% and 15% of all proteins in humans and mice are glycosylated, respectively. However, based on the rate of glycosylation site occupancy from Swiss-Prot, it has
been estimated that more than half of all proteins in nature are glycosylated (Apweiler et al., 1999). The relatively very low glycosylation percentage calculated by the PTM browser is mostly contributed by the limited information on glycosylated proteins available in the database, for instance the lack of data on O-GlcNAc modification meaning that many cytoplasmic and nuclear glycoproteins, which are very commonly modified with O-GlcNAc, are not recorded as such.

Throughout this thesis, glycan residues are mostly depicted using cartoon structures (Figure 1.2) which were standardised for the Essentials of Glycobiology textbook (Varki et al., 1999) and later adopted and modified by the Consortium for Functional Glycomics (CFG) (www.functionalglycomics.org).

![Figure 1.2. Symbol nomenclature for the most common monosaccharides found on mammalian glycoproteins.](image)
The cartoon structures are according to the guidelines in the Essentials of Glycobiology (Varki et al., 1999) and modifications by the CFG (www.functionalglycomics.org).

### 1.2.1 N-glycosylation

N-glycosylation is the attachment of oligosaccharides via an N-acetylglucosamine (GlcNAc) residue to the amide nitrogen of asparagine (Asn) side chains. This modification can determine or influence protein folding, stability, localisation, trafficking and oligomerisation (Freeze, 2006). Unlike O-glycans, N-glycans are almost invariably found in a specific target sequence, Asn-Xaa-Ser/Thr, where Xaa can be any amino acid except proline (Hart et al., 1979). Furthermore, N-glycosylation is exclusively initiated in the lumen of endoplasmic reticulum (ER) with en bloc transfer of a Glc$_3$Man$_9$GlcNAc$_2$ oligosaccharide from a lipid dolichol donor by the oligosaccharyltransferase complex (Silberstein & Gilmore, 1996).

Figure 1.3 summarises the key steps in the biosynthetic pathway of N-glycans from the ER to trans-Golgi. After the transfer of the N-glycan precursor to a protein, the processing and maturation of N-glycans begins with the removal of all glucose residues and some of the mannose residues in the ER and cis-Golgi, forming Man$_5$GlcNAc$_2$Asn. These actions are carried
out by glucosidases I and II, ER mannosidase I and Golgi mannosidases IA, IB and IC. Some glycans remain in the high-mannose state, containing between five to nine mannose residues, while others are processed to more complicated structures (Taylor & Drickamer, 2006). The first branch of an N-glycan is initiated in the cis-Golgi and this is followed by further digestion of mannose residues in the medial-Golgi by mannosidase II. Another GlcNAc residue is then added forming the biantennary structure Man$_3$GlcNAc$_4$Asn. More GlcNAc residues may be added to form triantennary and tetraantennary structures. N-glycans are then ready for modifications and elongations which continue in the trans-Golgi and trans-Golgi network (Stanley et al., 2009). N-glycans can be classified into 3 types based on their peripheral structures: high-mannose structures which contain only mannose residues, complex structures which contain antennae initiated by GlcNAc residues and hybrid structures which contain only mannose residues on the Man$_6$ arm of the core and GlcNAc initiated antennae on the Man$_3$ arm. All N-glycans share the common core sequence Man$_6$(Man$_3$)Man$_4$GlcNAc$_4$GlcNAc$_1$-Asn, which is retained from the conserved precursor.

### 1.2.2 O-glycosylation

O-glycans are attached via an α or β-O-glycosidic linkage to the hydroxyl group of an amino acid residue, usually serine, threonine or hydroxylysine (HLys) on a protein. Peptide O-glycosylation creates a number of O-glycan classes depending on the first sugar attached to the peptide chain, namely O-N-acetylgalactosamine (O-GalNAc), O-mannose, O-GlcNAc, O-glucose, O-galactose (O-Gal), O-fucose and O-xylose (glycosaminoglycans). Most O-glycosylation occurs in the ER/Golgi. The exception is β-O-GlcNAcylation which is a common modification of nucleocytoplasmic proteins (Hu et al., 2010), and will not be discussed here. O-GalNAc-linked glycans are known as mucin type O-glycans, due to the fact that mucin glycoproteins are highly glycosylated with O-GalNAc. Mucin-type glycosylation is the focus of this thesis and will be discussed in detail in Sections 1.2.2.2 (glycans) and 1.2.4 (mucins). First, less common classes of O-glycosylation are briefly reviewed (Section 1.2.2.1).
Figure 1.3. Summary of the biosynthetic pathways of N-glycans
Processing of N-glycans from the ER to trans-Golgi results in three types of N-glycan, namely complex, hybrid and high-mannose. Asn, asparagine; Dol-P-P, dolicholpyrophosphate. Based on (Dennis et al., 2009; Stanley et al., 2009).

1.2.2.1 β-O-galactose, α-O-fucose, β-O-glucose and α-O-mannose

In the mammalian system, O-galactosyl glycans have been found only on protein collagen domains (Freeze & Haltiwanger, 2009). Instead of serine and threonine, they are attached to hydroxylysine residue usually as a disaccharide Glcα1,2Gal. The quantities and types of O-galactose vary considerably among different types of collagen and even the same collagen
from different expression sites (Wopereis et al., 2006). O-galactose also has been identified in secreted proteins of *Schistosoma mansoni* eggs and larvae (Jang-Lee et al., 2007).

O-Fucosylation, on the other hand, is a common modification of epidermal growth factor (EGF)-like domains that are defined by six conserved cysteine residues which form three disulfide bonds. EGF-like domains can be found in several secreted and membrane-bound proteins. The consensus site where O-fucose is likely to be glycosylated by protein-O-fucosyltransferase 1 has been proposed as C\textsubscript{2}XXGGS/TC\textsubscript{3}, where C\textsubscript{2} and C\textsubscript{3} are the 2\textsuperscript{nd} and 3\textsuperscript{rd} conserved cysteines of the EGF repeat, S/T is the modified residue and X can be any amino acid (Harris & Spellman, 1993). More recently, O-fucose was detected in another cysteine-rich motif known as thrombospondin type 1 repeat (Luo et al., 2006). On this domain, protein O-fucosylation is catalysed by protein-O-fucosyltransferase 2 on serine or threonine. O-Fucose can exist either as a single monosaccharide or modified for elongation up to tetrasaccharides with GlcNAc, galactose (Gal) and sialic acid (Luther & Haltiwanger, 2009).

Similar to O-fucose, O-glucosyl glycans are generally found at EGF-like domains, but are β- instead of α-linked and only to serine. This glycosylation is initiated by protein-O-glucosyltransferase and the glycan typically exist as a monosaccharide or can be modified with xylose residues to form a trisaccharide (Luther & Haltiwanger, 2009). In addition, the putative consensus sequence for O-glucosylation also has been suggested as C\textsubscript{1}XSXPC\textsubscript{2}, where C\textsubscript{1} and C\textsubscript{2} are the first and second conserved cysteines of the EGF repeat, respectively, S is the glycosylation site and X can be any amino acid (Harris & Spellman, 1993).

*O*-Linked mannosyl glycans were first identified in 1969 in yeast and were found to be very abundant in the cell wall (Sentandreu, 1969), hence it used to be known as yeast-type modification (Kukuruzinska, 1987). In mammals, brain proteoglycans were the first found to be O-mannosylated (Krusius et al., 1986). To date, O-mannosylation occurrences in humans are relatively limited and have only been reported to be present on a small number of glycoproteins in the brains, nerves and skeletal muscle (Wopereis et al., 2006). The biosynthesis of O-mannosyl glycans is initiated in the ER by protein-O-mannosyltransferase-1 and -2 (POMT-1 and -2) which add mannose from dolichol-phosphate-Man to serine or threonine residues (Van den Steen et al., 1998). Manya and colleagues have proposed a consensus sequence preferred for O-mannosylation that is IXPT(P/X)TXPXXXXPTX(T/X)XX, where I is isoleucine, P is proline, T is threonine and X can be any amino acid (Manya et al., 2007). However, another two more
recent studies argued that O-mannosylation is regulated in a much more complicated manner than just a simple local sequence (Breloy et al., 2008; Stalnaker et al., 2010). Modification on O-linked mannose is initiated by protein-O-mannose-β1,2-N-acetylglucosaminyltransferase 1 (POMGnT1) which adds β1,2-GlcNAc. The resulting disaccharide then can be further elongated and branched with sugar compositions including GlcNAc, Gal, sialic acid, glucuronic acid (GlcA) and fucose. Unlike POMTs and POMGnT1, it is not known whether the other participating enzymes are unique to this pathway (Freeze & Haltiwanger, 2009). Although it is not fully understood yet, correct modifications, particularly glycosylation of the glycoprotein α-dystroglycan (α-DG) by POMT2 and POMGnT1 and two putative glycosyltransferases, fukutin-related protein (FKRP) and LARGE, were found to be vital in ensuring the integrity of muscle tissues (Endo et al., 2010). Accordingly, O-mannosylation of α-DG has been shown to be critical in stabilising the link between the extracellular matrix and cytoskeleton in muscle (Moore & Hewitt, 2009). α-DG isolated from human and rabbit skeletal muscles has been mapped and characterised to have at least 3 and 9 sites modified with O-mannose-initiated glycans, respectively (Nilsson et al., 2010; Stalnaker et al., 2010). Campbell and co-workers recently have identified a novel phosphorylated O-mannose sequence \[\text{GalNAc}\beta1,3\text{GlcNAc}\beta1,4(6\text{-PO}_3)\text{Man}\] on recombinant human α-DG at which LARGE is speculated to participate in an unidentified post-phosphoryl glycosylation important for binding to the extracellular matrix protein laminin (Yoshida-Moriguchi et al., 2010).

### 1.2.2.2 Mucin type O-glycans

Mucin type O-glycans (α-O-GalNAc) are the most commonly found O-glycans on secreted and membrane-associated proteins in mammalian tissues, especially on a family of glycoproteins known as the mucins (see section 1.2.4). The biosynthetic pathway is initiated by the addition of a GalNAc residue from a UDP-GalNAc on serine or threonine by a family of UDP-GalNAc: polypeptide-α-N-acetylgalactosaminyltransferases (ppGalNAcTs). At least 20 and 13 ppGalNAcTs have been identified in humans (Perrine et al., 2009) and mice (www.informatix.jax.org) (Bult et al., 2008), respectively, and they are usually located in the cis-Golgi. Some isoforms are expressed ubiquitously (ppGalNAcT-1 and -2), but some are expressed specifically in a subset of tissues (Ten Hagen et al., 2003). All ppGalNAcTs are composed of a catalytic domain and a lectin domain. It has been suggested that the lectin domain modulates the glycosylation of glycopeptide substrates and may contribute to the strict specificities of certain ppGalNAcT isoforms. ppGalNAcT specificities are not fully understood yet but it seems that they
have similar and unique specificities. For instance, ppGalNAcT-1 & -2 have been shown to prefer glycosylating peptides with a common motif (Gerken, 2006), but not ppGalNAcT-10 (Perrine et al., 2009). Due to the complex regulation of the addition of the first GalNAc, there is unlikely to be a general consensus sequence for mucin type glycosylation. However, a number of statistical studies have yielded some calculations to predict O-glycosylation sites, for instance the NetOGlc server (Hansen et al., 1998) and the Random Forest algorithm (Hamby & Hirst, 2008) that demonstrated over 80% and 90% of accuracy, respectively, for serine and threonine.

The serine/threonine-linked GalNAc residue (known as the Tn antigen) can be further modified or elongated to produce up to 8 core-types in mammals (Figure 1.4). The attachment of a Gal residue by core 1 β1,3-galactosyltransferase (C1GalT/T-synthase) on position 3 of the GalNAc produces core 1 (also known as Thomsen-Friedenreich or T antigen). The core 1 enzyme has to compete with α2,6-sialytransferase I which sialylates the Tn antigen at position 6 forming sialyl-Tn (sTn) antigen thus preventing substitution at position 3 (Sewell et al., 2006). The T, Tn and sTn antigens are largely restricted to mucins secreted by normal and malignant epithelial cells and have been correlated with aggressive tumour growth and poor prognosis in a number of tumours (Springer, 1989; O’Boyle et al., 1992). Once formed, the core 1 O-glycan can be sialylated/disialylated, α-galactosylated, fucosylated, elongated and/or converted to the core 2 O-glycan. The latter is formed by the attachment of a GlcNAc residue on position 6 of the GalNAc of core 1 by core 2 β1,6-N-acetylglucosaminyltransferases (C2GnTs). C2GnTs will be discussed in the next section.

On a separate pathway, the Tn antigen can also be extended by the attachment of a GlcNAc residue on position 3 of the GalNAc to form core 3, and an additional GlcNAc on position 6 to form core 4, by Core 3 β1-3-N-acetylglucosaminyltransferase (C3GnT) and core 2 β1,6-N-acetylglucosaminyltransferase 2 (C2GnT2), respectively (Vavasseur et al., 1995; Schwientek et al., 1999). O-glycan cores 5 to 8 have rarely been observed. Furthermore, enzymes involved in their biosynthesis remain to be fully characterised. Core 5 O-glycans were first reported from human meconium (Vincent et al., 2008) and are believed to exist in colonic tissues and colonic adenocarcinoma (Brockhausen et al., 2009). Core 6 O-glycans were also first reported from human meconium (Perrine et al., 2009) and core 7 O-glycans in bovine kappa-casein (Gerken, 2006), whereas core 8 O-glycans were from human respiratory mucin (Torres & Hart, 1984).
Figure 1.4. The biosynthetic pathways of mucin type O-glycans.

Eight cores of mucin type O-glycan have been described in mammals. All cores can be further elongated with LacNac units and modified with sialylation, fucosylation, blood groups etc.

All core structures can be further elongated by the repetitive addition of N-acetyllactosamine (LacNac) units type 1 (Galβ1-3GlcNAc) or type 2 (Galβ1-4GlcNAc). The chain may also be branched on a Gal residue by I β-1,6-N-acetylglucosaminyltrasferase (IGNT) and C2GnT and/or modified with fucosylation, sialylation and other capping moieties (Section 1.2.3).

1.2.2.3 Core 2 β1,6-N-acetylglucosaminyltransferase

Core 2 β1,6-N-acetylglucosaminyltransferase (C2GnT), which exists in 3 isoforms, C2GnT1, C2GnT2 and C2GnT3, is one of the key enzymes in the O-glycan biosynthetic pathway. Recently, it has been shown that in the mouse C2GnT2 is highly expressed mainly in the stomach and colon, whereas C2GnT3 is highly expressed mainly in the small intestine, liver.
and spleen (Stone et al., 2009) (Figure 1.5). On the other hand, C2GnT1 is more widely expressed in most tissues (Yeh et al., 1999).

![Figure 1.5. Relative expression levels of C2GnT2 and C2GnT3 in various murine tissues. Based on data from (Stone et al., 2009).](chart)

The C2GnT gene (C2GnT1) was first cloned from humans in 1992 (Bierhuizen, 1992) and since then many efforts have been directed towards fully understanding core 2 O-glycans and the enzymes involved. C2GnT1-deficient mice have been shown to suffer severe but selective defects in selectin ligand biosynthesis, demonstrated by the reduction in neutrophil rolling but not lymphocyte homing (Ellies et al., 1998). These early results already indicated that there is a possibility of more than one C2GnT enzyme existing. In 2004, Hiraoka et al. showed that core 2 O-glycans, via the sialyl Lewis\(^x\) (SL\(e^x\)) terminal antigen, are crucial for lymphocyte homing in mice (Hiraoka et al., 2004). This finding was further clarified by Gauguet et al. in 2004, who suggested that C2GnT1 exerts differential control over B- and T-lymphocyte homing (Gauguet et al., 2004). Recently, human C2GnT1 has been demonstrated to prefer a core 1 substrate on a hydrophilic glycopeptide rather than those having aromatic groups, suggesting a role of the peptide moiety in regulating C2GnT1 activity (Brockhausen et al., 2009).

In 1999, Yeh et al. and Schwientek et al. separately cloned a second C2GnT gene from humans (C2GnT2), primarily expressed in mucus secretory tissues, and proposed the idea that C2GnT2 could also be involved in synthesizing core 4 O-glycans and I-branches on O-glycans (Schwientek et al., 1999; Yeh et al., 1999). I-branches were originally thought to be solely produced by IGnT isoenzymes (see Section 1.2.3). Murine C2GnT2 also has been characterised
to have similar enzymatic properties as the human counterpart (Hashimoto et al., 2007). In line with the fact that most cancer cells express simple O-glycan structures (Brockhausen, 1999; Burchell et al., 1999), human C2GnT2 has been shown to suppress the adhesion, motility and invasion of colorectal cancer cells. It was evident that this enzyme is frequently downregulated in cancer patients and cancer cell lines (Huang et al., 2006). C2GnT3, which was first cloned from humans in 2000, has been postulated to play an important role in T-cell biology due to its relatively high expression in human thymus (Schwientek et al., 2000). It was demonstrated that EGF preferentially inhibited C2GnT2 whereas C2GnT1 was moderately affected in the human airway adenocarcinoma cell line, contributing to the shifting of the carbohydrate structures of MUC1 in the cells from core 2-based towards core1-based (Beum et al., 2003). These findings have highlighted potential differences in biological functions and regulation amongst the C2GnTs. Accordingly, a human airway epithelial cell line exhibited upregulation of C2GnT2 upon stimulation by interleukin-4 and -13 and retinoic acid. However C2GnT3 was shown to be only moderately upregulated by the interleukins (Beum et al., 2005).

Although oligosaccharide functions are mostly determined by terminal antigens, it has been demonstrated that their synthesis is often not regulated through terminal or intermediate glycosyltransferases, but by the branching glycosyltransferases, such as C2GnT, that allow the generation of the scaffold for terminal epitopes. For example, it has been reported that C2GnT1, but not C2GnT2 or C2GnT3, regulates the synthesis of SLeα structures that are preferentially downregulated during human B cell differentiation (Kikuchi et al., 2005). Even though many biological functions of C2GnTs have been suggested, particularly of human C2GnT1, previous studies have provided little evidence to associate their findings specifically with glycan structures that are initiated by the different isoenzymes.

In addition to the ambiguities regarding the roles of C2GnT2 and C2GnT3, it is also still unclear how much these three isoenzymes compensate each other’s activities in vivo. Therefore, it is important to investigate the correlation between C2GnT properties and the aetiology of diseases and the necessities of having multiple enzymes to produce core 2 O-glycans. To this end, mice with separate deficiencies of C2GnT2 and C2GnT3 as well as all three C2GnTs have been generated by our collaborators Jamey Marth and colleagues from the University of California San Diego, USA. Results of comprehensive O-glycomic analysis of these mice correlated with the phenotypic findings of our collaborators are presented in Chapter 3 of this thesis.
1.2.3 Glycan extension, branching and terminal epitopes

Both N- and O-glycans can be extended with poly-N-acetyllactosamine (poly-LacNAc) chains and decorated with various types of sialylation and fucosylation. There are two types of poly-LacNAc chain based on the N-acetyllactosamine (LacNAc) constituent units. The type 1 LacNAc unit sequence is Galβ1-3GlcNAc, whereas the type 2 unit is Galβ1-4GlcNAc. The slight distinction of monosaccharide linkages between these two units when polymerised in a chain can highly influence the conformation of the overall glycan structure, thus providing different binding surfaces. Furthermore, different poly-LacNAc chains could lead to the synthesis of different terminal epitopes. Poly-LacNAc chains are produced by concerted actions of β1,3-N-acetylglucosaminyltransferase (i-extension enzyme or iGnT) and β1,3/4-galactosyltransferases (β3/4Gal-T) (Seko et al., 1996; Ujita et al., 2000). It has also been proposed that poly-LacNAc chains on core 2 O-glycans and core 4 O-glycans are most efficiently synthesised by β4Gal-T IV and β4Gal-T I (Ujita et al., 2000). However, to date only type 2 LacNAc units have been shown to be capable of polymerisation to form poly-LacNAc chains on N- and O-glycans (Ujita et al., 1998; Ujita et al., 1999). Elongated type 1 LacNAc units have only been demonstrated on the oligosaccharides of lactosylceramides (Fan et al., 2008).

Linear LacNAc chains can be branched on position 6 of a Gal residue by IGnT and C2GnT (Yeh et al., 1999). There are two types of I-branching activities. Distally-acting IGnT (dIGnT) initiates branching at sites closer to non-reducing termini whereas centrally acting IGnT (cIGnT) acts at internal sites of poly-LacNAc chains (Figure 1.6). It has been demonstrated in vitro that on a branched oligosaccharide, extension is more common on the backbone chain than on the I-branch when incubated with iGnT and β4Gal-T I (Ujita et al., 1999). Originally, the importance of I-branches was demonstrated by the analysis of band 3 glycoprotein of human erythrocytes. The conversion of i-antigen (linear LacNAc chain) to I-antigen (branched LacNAc chains) is critical in the development of erythrocytes (Fukuda et al., 1979). The I-branch allows multivalent terminal epitopes to be synthesised, therefore replicating binding possibilities. For example, H antigens present on both termini of branched LacNAc chains provide better avidity to anti-ABO antibodies than a single epitope on linear LacNAc chains (Romans et al., 1980).
Figure 1.6. The biosynthesis and elongation of I-antigen/I-branch.

I-antigen is biosynthesised from i-antigen by C2GnT2 or IGnTs and the elongation is initiated by $\beta_1,4$-galactosyltransferase. dIGnT and cIGnT activities are independent of each other.

Terminal epitopes are often the functionally important component of a glycan molecule. Certain terminal epitopes are classified based on the type of LacNAc chain they are residing on, for instance histo-blood group antigens type 1 and type 2. Histo-blood group antigens include A, B, H/O, Lewis (Le) antigens and their derivatives (Lewis antigens will be discussed in more detail in the next section). In addition to these antigens, Sd$, Gal\alpha1,3Gal$, LacdiNAc, and sialylated LacNAc are also common terminal epitopes on N- and O-glycans. The structures of these terminal epitopes are shown in Figure 1.7. Protein-glycan interactions, as has been discussed earlier, are mostly mediated by these sequences. For example, P and E-selectins have been shown to bind to SLe$^x$ structures on O-glycans of PSGL-1 and this is important for leukocyte homing on endothelial cells (Kawar et al., 2008).
Figure 1.7. The most common terminal epitopes found on mammalian glycans.

Type 1 Lewis antigens and the Galα1,3Gal epitope are known to be absent in mice and humans, respectively. Some of the epitopes including Lea/x and Leb/y could be duplicated on the penultimate LacNAc unit. Sialic acids depicted could be either N-acetylneuraminic acid (NeuAc) or N-acetylglycolylneuraminic acid (NeuGc). Red and blue numbers/letters represent LacNAc type 1 and type 2, respectively, whereas those in black are for general annotations. S, sulphate group.

Glycan epitopes on host cell surfaces are also specifically recognised by microbial agents to initiate colonisation. These microbes sometimes exploit molecular mimicry by expressing host glycan epitopes to avoid being recognised as foreign by the host immune systems. As an example, the human gut is a home for millions of bacteria that mutually and continuously interact with the mucosal layer (mostly in the intestine) which have important effects, including host immune function and nutrient processing (Hooper & Gordon, 2001). There are also pathogenic interactions for instance, as mentioned earlier, those utilised by *H. pylori* aided by its host-glycan
mimicry to infect human stomach. Another example would be *Pseudomonas aeruginosa*, which infects human lungs and kidneys. This bacterium expresses the lectin LecB, that binds to SLe\(^a\), and, interestingly, with much lower affinity to SLe\(^x\) (Imberty et al., 2004). A number of species of human parasitic helminths also express Le\(^x\) epitopes on glycoproteins and glycolipids found on their surface or within their secretory products in all life stages (van Die & Cummings, 2010).

### 1.2.3.1 Lewis antigens

Lewis (Le) antigens are biochemically related to the ABH blood group antigens, that are formed as terminal epitopes on carbohydrate chains type 1 or type 2 of both glycolipids and glycoproteins (Moran, 2008), as shown in Figure 1.7. The biosynthesis of Lewis antigens starts from the competitive actions of glycosyltransferases including α1,2-fucosyltransferases, α1,3/4-fucosyltransferases and α2,3-sialyltransferases. The first enzyme leads to the synthesis of H antigens type1/2, which then serve as substrates for the synthesis of A, B, Le\(^b/y\), and A/B-Le\(^b/y\) antigens. The actions of α1,3-fucosyltransferase and α1,4-fucosyltransferase on the penultimate GlcNAc produces Le\(^x\) and Le\(^a\) antigens, respectively. If the terminal Gal is readily sialylated prior to the fucosylation, the antigens Sialyl-Le\(^x\) (SLe\(^x\)) and Sialyl-Le\(^a\) (SLe\(^a\)) are formed instead. These determinants could also carry sulfate esters, mainly at the 3-hydroxyl of Gal and/or the 6-hydroxyl of GlcNAc (Brockhausen, 1999).

The most common Lewis antigens in human secretions are Le\(^a\), Le\(^b\), Le\(^x\), Le\(^y\), SLe\(^x\) and SLe\(^a\). However in mouse, the orthologous gene for human FuT3, the only enzyme known to produce 4-linked fucose, is found to be a pseudogene (Gersten et al., 1995). Therefore, it has been assumed that mice do not synthesise Le\(^a\), Le\(^b\), SLe\(^a\) and their sulphated derivatives although this assumption is now being questioned due to preliminary findings by our collaborators, Celso Reis et al. from University of Porto, Portugal. Anti Le\(^a\) and anti Le\(^b\) monoclonal antibodies were found to be able to bind to murine gastric mucosa, suggesting the existence of a functional α1,4-fucosyltransferase in mice.

Structural evidence that assertively distinguishes Le\(^a\) and Le\(^b\) from Le\(^x\) and Le\(^y\) in murine tissues has not been applied to previous *H. pylori* adhesion studies, therefore leaving ambiguities in their binding/staining-based characterisation. Parry et al., as an example, have proven that mass spectrometric analysis is able to identify SLe\(^x\) in mice (Parry et al., 2007). It is anticipated that with further improvement and modification, this technique in addition to chemical and
enzymatic digestion, can shed light on the Lewis antigen repertoire in the highly complex glycosylation of the murine GI tract.

Classical models assume that Le$^b$ and Le$^y$ structures are synthesised from H antigen type 1 and 2, respectively. However, it has been shown in some human cancer tissues and cells that there is an unusual α1,2-FuT activity that could add fucose to monofucosylated antennae, i.e. allowing the synthesis of Le$^b$ from Le$^a$ or Le$^y$ from Le$^x$ (Wang et al., 1999). It has been revealed that in mucinous tissues, FuT2 (secretor enzyme) is the main enzyme that catalyses the addition of α1,2-fucose residues (Hurd et al., 2005), however there is no evidence so far on which pathway is the most dominant in the murine GI tract. In order to better understand the biosynthesis of the fucosylated Lewis antigens in the mouse, it is best to first explore the fucosyltransferase population in this organism.

1.2.3.2 Functional significance of fucosyltransferases

Fucosylation appears to be one of the most frequent modifications in eukaryotic cells involving oligosaccharides including N- and O-glycans, but rather limited in prokaryotic organisms (Ma et al., 2006). Changes in the levels of fucosylation in mammalian cells point to its significance in a number of pathological conditions, including inflammation and cancer, and the biological functions of adhesion molecules such as selectin ligands (Miyoshi et al., 2008). The regulation of fucosylation seems to be complicated with the existence of fucosyltransferases (FuTs) with distinct but overlapping specificities and different tissue expressions, in addition to other enzymes involved in the synthesis and transport of GDP-fucose, the sugar nucleotide donor for fucosylation (Becker & Lowe, 2003). Generally, FuTs are a group of enzymes that catalyse the addition of fucose from GDP-fucose, either to amino acid residues or to other sugars. On oligosaccharides, fucose can be linked at position 2 of Gal or 3, 4 or 6 of GlcNAc, contributing to the vast variety of glycan epitopes and functional properties. Even though mice have become a very important model in human biological research (see Section 1.3.1), little is known regarding their FuTs and they are often confused with the relatively better studied human FuTs. To date, there are 8 FuTs in mouse that have been cloned or described in publications (Table 1.1); nonetheless, there are still many ambiguities with respect to their specificities. Humans, on the other hand, have at least 11 FuTs with known biological activities (Becker & Lowe, 2003).
Table 1.1. Fucosyltransferases in mice and humans with their respective activities.
The list of fucosyltransferases were extracted from the database of Murine Genome Informatics (www.informatics.jax.org) (Bult et al., 2008) without including protein-O-fucosyltransferases.

<table>
<thead>
<tr>
<th>Name</th>
<th>Enzymatic activity</th>
<th>Major biological activity(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FuT1 (Humans &amp; mice)</td>
<td>α1,2 fucosylation</td>
<td>Synthesises blood group H antigens in erythrocytes (Larsen et al., 1990; Domino et al., 2001)</td>
</tr>
<tr>
<td>FuT2/ Secretor (Humans &amp; mice)</td>
<td>α1,2 fucosylation</td>
<td>Synthesises blood group H antigens in secretions (Kelly et al., 1995; Domino et al., 2001)</td>
</tr>
<tr>
<td>FuT3 (Humans)</td>
<td>α1,3/4 fucosylation</td>
<td>Synthesises Lewis antigens type 1 and 2 (Kukowska-Latallo et al., 1990)</td>
</tr>
<tr>
<td>FuT4 (Humans &amp; mice)</td>
<td>α1,3 fucosylation</td>
<td>Controls Le× synthesis (Kumar et al., 1991; Homeister et al., 2001)</td>
</tr>
<tr>
<td>FuT5 (Humans)</td>
<td>α1,3 fucosylation</td>
<td>Synthesises Lewis antigens (Weston et al., 1992)</td>
</tr>
<tr>
<td>FuT6 (Humans)</td>
<td>α1,3 fucosylation</td>
<td>Direct the synthesis of Leα and SLex (Weston et al., 1992)</td>
</tr>
<tr>
<td>FuT7 (Humans &amp; mice)</td>
<td>α1,3 fucosylation</td>
<td>Controls leukocyte trafficking via E, L and P-selectin ligand formation- SLex (Sasaki et al., 1994; Malý et al., 1996)</td>
</tr>
<tr>
<td>FuT8 (Humans &amp; mice)</td>
<td>α1,6 fucosylation</td>
<td>N-glycan core fucosylation (Yanagidani et al., 1997; Becker &amp; Lowe, 2003)</td>
</tr>
<tr>
<td>FuT9 (Humans &amp; mice)</td>
<td>α1,3 fucosylation</td>
<td>Controls Leα and Leγ but not SLex synthesis in brain and stomach (Kaneko et al., 1999; Kudo et al., 2007)</td>
</tr>
<tr>
<td>FuT10 (Humans &amp; mice)</td>
<td>α1,3 fucosylation</td>
<td>Core fucosylation of biantennary N-glycans (Baboval &amp; Smith, 2002; Mollicone et al., 2009)</td>
</tr>
<tr>
<td>FuT11 (Humans &amp; mice)</td>
<td>α1,3 fucosylation</td>
<td>Core fucosylation of biantennary N-glycans (Baboval &amp; Smith, 2002; Mollicone et al., 2009)</td>
</tr>
</tbody>
</table>

In humans and mice, FuT2 is responsible for the synthesis of α1,2-fucose on a terminal Gal and also the associated structures derived from this sequence in epithelial cells and body fluids. FuT2 was shown in vitro to have higher affinity for type 2 LacNAc chains but lower affinity for type 1 LacNAc chains, as compared to FuT1 (Sarnesto et al., 1992). Individuals with inactivating mutations in the FuT2 gene are known as non-secretors and they appear to have limitations in synthesising ABO blood group antigens in their epithelial secretions. Unlike non-secretors, who can only produce Leα and Leα histo-blood group antigens and their sialylated counterparts, secretors (normal individuals) have the capability to also produce H antigens and difucosylated Lewis antigens which can be found along the GI tract (Lindén et al., 2008).
However, it has been reported that MUC5B mucin from the non-secretors demonstrated higher expression levels of sialylated epitopes (Thomsson et al., 2005).

An interesting variation in susceptibility towards *H. pylori*’s infection relating to the secretor status in humans has been observed (Ikehara et al., 2001). It was shown that secretor individual are more prone to be infected by blood group antigen-binding adhesin (BabA) positive strains than the non-secretors, the latter consisting of about 20% of the western population (Azevedo et al., 2008). Infected secretor animals have been shown to have higher gastric mucosal density of *H. pylori*, gastritis and sialylation compared to the non-secretors (Lindén et al., 2008). Furthermore, individuals with blood group O phenotype have a higher incidence of gastric ulcers than other blood group individuals, which might indicate the importance of exposed fucosylated terminal epitopes in *H. pylori* infection (Borén et al., 1994). However, incapability of producing α1,2-fucose is not always a health advantage. A statistical analysis has shown significant data that non-secretor individuals have higher susceptibility towards Crohn’s disease (McGovern et al., 2010).

To help resolve ambiguities concerning the H and Lewis antigens and the adhesion of *H. pylori*, a mouse model with deficiency in FuT2 has been generated and characterised. Furthermore, this model closely resembles the non-secretors, which could provide tools to better understand the differences between this phenotype and the secretors (wild type) as well as the diverse pathogenicity of *H. pylori* (see Section 1.2.5). Results of this study are presented in Chapter 4.

### 1.2.3.3 Carbohydrate recognition by lectins

As discussed earlier, glycans are capped with a variety of functionally important terminal structures. Turning biological information stored in these epitopes into an operative sugar code entails the existence of efficient decoding devices such as lectins (Soils et al., 2009). Lectins comprise various intracellular and extracellular receptors that bind to carbohydrate molecules. Lectins are most commonly classified according to their CRD properties (Taylor & Drickamer, 2006). Because this thesis focuses on murine glycosylation, it is pertinent to consider which lectins are found in this animal. Known murine lectins are listed in Table 1.2 with their respective properties. In the context of the mouse as a model animal (see Section 1.3.1), it is important to bear in mind that not all human lectin genes have an orthologous partner in the murine genome.
For instance, the human genome codes for at least 13 functional siglecs (Crocker et al., 2007), in contrast to only 9 in mice. Furthermore, at least 10 functional galectins have been identified in humans (www.uniprot.gov) (Consortium, 2010) whereas in mice there are only 7 known so far.

Table 1.2. Families of the most common murine lectins and their properties.
The list of lectins were extracted from the database of Murine Genome Informatics (www.informatics.jax.org) (Bult et al., 2008) and the CFG, Consortium for Functional Glycomics glycan array data (www.functionalglycomics.org).

<table>
<thead>
<tr>
<th>Family</th>
<th>Typical members</th>
<th>Expression sites</th>
<th>Main ligands</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C-type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Selectins:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-selectin (CD62L)</td>
<td>Lymphocytes (Lasky et al., 1989).</td>
<td>SLex on core 2 and sulphated tyrosines on PSGL-1 and endoglycan (Leppänen et al., 2010).</td>
<td></td>
</tr>
<tr>
<td>P-selectin (CD62P)</td>
<td>Activated endothelial cells (EC), platelets (Sanders et al., 1992).</td>
<td>SLex on core 2 and sulphated tyrosines on PSGL-1 (Leppänen et al., 2010).</td>
<td></td>
</tr>
<tr>
<td>E-selectin (CD62E)</td>
<td>Activated EC (Becker-Andre et al., 1992).</td>
<td>SLex on N-glycans of CD44 and ESL-1 and core 1 and core 2 O-glycans of PSGL-1 (Ellies et al., 1998; Yago et al., 2010).</td>
<td></td>
</tr>
<tr>
<td><strong>Collectins:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surfactant proteins A and D.</td>
<td>Alveolar (Motwani et al., 1995)</td>
<td>Glucosyl residues of cells, virus and microbes (Kishore et al., 2006).</td>
<td></td>
</tr>
<tr>
<td><strong>SIGNR1 and SIGNR5</strong></td>
<td>Macrophages (Galustian et al., 2004)</td>
<td>High-mannose N-glycans, Le, Le on microbial surface polysaccharides (Galustian et al., 2004; Powlesland et al., 2006).</td>
<td></td>
</tr>
<tr>
<td><strong>SIGNR3 (DC-SIGN)</strong></td>
<td>Dendritic cells, macrophages, monocytes (Nagaoka et al., 2010).</td>
<td>High-mannose N-glycans, fucosylated glycans (Galustian et al., 2004; Powlesland et al., 2006).</td>
<td></td>
</tr>
<tr>
<td><strong>SIGNR7</strong></td>
<td>Testis, liver (Powlesland et al., 2006).</td>
<td>SLex and Sulfo-SLe (Powlesland et al., 2006).</td>
<td></td>
</tr>
<tr>
<td>Galectins</td>
<td>Galectin 1</td>
<td>Thymus (Earl et al., 2010).</td>
<td>β-galactosides including H antigen (CFG).</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Galectin 3</td>
<td>Various immune cells (Liu, 2005).</td>
<td>β-galactosides including H antigen (CFG).</td>
<td></td>
</tr>
<tr>
<td>Galectin 4 and 6</td>
<td>Small intestine, colon, kidney, liver (Markova et al., 2006).</td>
<td>α-GalNAc and α-Gal (Markova et al., 2006).</td>
<td></td>
</tr>
<tr>
<td>Galectin 7</td>
<td>Epithelial cells (Sato et al., 2002).</td>
<td>β-galactosides including H antigen (CFG).</td>
<td></td>
</tr>
<tr>
<td>Galectin 9</td>
<td>Lymphocytes, liver, small intestine, thymus (Wada &amp; Kanwar, 1997).</td>
<td>β-galactosides including H antigen (CFG).</td>
<td></td>
</tr>
<tr>
<td>Galectin 12</td>
<td>Adipocytes (Hotta et al., 2001).</td>
<td>β-galactosides including H antigen (CFG).</td>
<td></td>
</tr>
<tr>
<td>I-type</td>
<td>Siglec-1/ sialoadhesin</td>
<td>Macrophages (Crocker et al., 1994)</td>
<td>NeuAcα2,3Galβ1,3GalNAc (Crocker et al., 1994).</td>
</tr>
<tr>
<td></td>
<td>Siglec-2/ CD22</td>
<td>B cells (Crocker et al., 2007).</td>
<td>α2,6-linked sialic acid (Hoffmann et al., 2007).</td>
</tr>
<tr>
<td></td>
<td>Siglec-3/ CD33</td>
<td>Myeloid precursors in bone marrow (Brinkman-Van der Linden et al., 2003).</td>
<td>Short O-glycans and sialyl-Tn (Brinkman-Van der Linden et al., 2003).</td>
</tr>
<tr>
<td></td>
<td>Siglec-5/ Siglec F</td>
<td>Eosinophils (Zhang et al., 2004).</td>
<td>α2,8-linked disialic acid (Zhang et al., 2004), 6-sulfo-SLe^a (Tateno et al., 2005).</td>
</tr>
<tr>
<td></td>
<td>Siglec-E</td>
<td>Various cells of innate immune system (Zhang et al., 2004).</td>
<td>α2,3/6-linked sialic acid (Hoffmann et al., 2007).</td>
</tr>
<tr>
<td></td>
<td>Siglec-G</td>
<td>B cells (Hoffmann et al., 2007).</td>
<td>α2,6-linked sialic acid (Nitschke, 2009).</td>
</tr>
</tbody>
</table>
The selectins are arguably the best characterised family of C-type lectins (which bind sugars in a Ca\(^{2+}\)-dependent manner). They are expressed on endothelial cells, leukocytes and platelets, and have extensively documented roles as mediators for leukocyte and platelet rolling. In humans, selectins bind to sulphated/non-sulphated sialyl (NeuAc) Le\(^x\) and (in some instances) sulphated amino acids on selective carriers (Carlow et al., 2009). Recently, it has been proven that selectins can bind to both N-acetylneuraminyl and N-glycolynueraminyl forms of SLe\(^x\) and 6-sulfo SLe\(^x\), the latter form is the most prominent in murine tissues (Mitoma et al., 2009). The same study also highlighted that a number of anti-Le\(^x\) antibodies raised against human cells do not recognise murine SLe\(^x\) ligands, illustrating disparities in the binding mechanism between anti-carbohydrate antibodies and lectins. Weak α1,3-fucosylation activity by FuT7 was shown to have higher impact on E-selectin than on P-selectin or L-selectin binding efficiency on PSGL-1, which is explained by the latter two selectins’ requirement of tyrosine sulphation in addition to core 2 O-glycans for proper binding (Prorok-Hamon et al., 2005). The molecular basis for differential ligand recognition by selectins has started to be elucidated. For instance it was revealed that the substitution of a single residue in the CRD of L-selectin increases its affinity towards the sulphated peptide binding site but not to the sulphated glycan binding site (Klopocki et al., 2008).

As lectin receptors play central roles in mediating and modulating immune responses, elucidation of their ligand-binding properties is definitely of importance. For instance, C-type lectins have been identified to play crucial roles in regulating immune responses towards Candida albicans and thus are now an important target for vaccine development (Ferwerda et al., 2010). Similarly, galectin-1, but not galectin-3, increased human immunodeficiency virus-1 (HIV-1) infectivity on macrophages, most likely due to the distinction in virus adsorption kinetics
with galectin (Mercier et al., 2008). This finding is in parallel with previous analysis on the
binding thermodynamics of galectin-1 and -3, with one of the disparities being the former’s lower
affinity towards 2,6-sialylated diLacNAc (Ahmad et al., 2002). In addition, HIV-1 also has been
shown to utilise the binding to DC-SIGN to induce kinase Raf-1 signalling cascades leading to
the generation of full-length viral transcript for its replication (Gringhuis et al., 2010).

DC-SIGN plays critical functions in immune regulation by recognising a wide variety of
pathogens not just viruses but also bacteria, yeast and parasites via the binding of mannose and
Le\(^x\) structures. Additionally it serves as an adhesion receptor via binding to intracellular adhesion
molecule 2 (ICAM 2) and ICAM 3 present in the membrane of leukocytes and endothelial cells
(van Kooyk & Geijtenbeek, 2003). For instance, mice deficient in DC-SIGN, but not those that
are deficient in SIGNR-1 or -5, exhibited reduced resistance towards Mycobacterium tuberculosis
infection, illustrating its unique contribution to host protection (Tanne et al., 2009). Detailed
specificity of human DC-SIGN has been demonstrated with the highest binding being attributed
to Le\(^x\) (preferably multivalent type) followed by high-mannose and fucosylated biantennary N-
glycans, but no binding to core fucosylated N-glycans was observed (van Liempt et al., 2006).

Siglecs have been regarded as positive and negative regulators of the immune system
expressed mainly on various subsets of leukocytes (Crocker & Redelinghuys, 2008). For instance
human MUC2 was shown to bind to siglec-3 on the surface of monocyte-derived dendritic cells
and induces apoptosis (Ishida et al., 2008). As mucins are commonly expressed by cancer tissues,
this mechanism confers protection from immune cells. Several other siglecs including siglec-1
and -5, on the other hand, have been postulated to mediate the interactions between macrophages/
neutrophils and sialylated bacteria lipopolysaccharide thereby playing a potentially protective role
during infections (Jones et al., 2003). The restricted expression of several siglecs to specific cell
types makes them attractive targets for cell-directed therapies (O'Reilly & Paulson, 2009).

Some lectins play important intracellular roles and are mostly involved in protein sorting
and dislocation in the ER, illustrating the importance of proper glycosylation on the fate of newly
synthesised glycoproteins. Calreticulin, a soluble protein in the lumen of the ER, together with its
membrane-bound homologue, Calnexin, independently associates with a disulfide isomerise
ERp57 and functions in capturing misfolded or incompletely folded proteins tagged with
Glc\(_1\)Man\(_9\) N-glycans. This is to ensure proper protein folding by Erp57 before further
glycosylations and modifications on the glycan structures (Maattanen et al., 2010). This event has
been indirectly demonstrated by the analysis of C2GnT1 with a mutation in one of its two N-glycosylation sites by Prorok-Hamon et al. The N-glycan on Asn-95 was found to be essential for Golgi targeting and the ability of the enzyme to initiate the formation of a functional PSGL-1 (Prorok-Hamon et al., 2005). Recent findings have suggested that Calreticulin might be secreted into the extracellular milieu where it could perhaps regulate a wide range of cellular responses including immune responses (Gold et al., 2010).

1.2.4 Mucins

Mucins are large and filamentous glycoproteins that can be either membrane-bound or secreted. Typically, the backbone of a membrane-bound mucin (apomucin) contains a long extracellular domain, a hydrophobic transmembrane domain and a short cytoplasmic tail (Taylor & Drickamer, 2006). Membrane bound mucins also contain several other conserved domains including epidermal growth factor (EGF)-like, sea urchin sperm protein Enterokinase and Agrin (SEA) and tandemly repeated proline/threonine/serine (PTS) domains (Dekker et al., 2002; Van Seuningen & Jonckheere, 2008) (Figure 1.8). Secreted mucins, on the other hand, can be found in secretory cells such as gastric mucous cells and intestinal goblet cells and can be either gel-forming (for instance Muc2, Muc5AC, Muc5B, Muc6 and Muc19) or non-gel-forming (for instance Muc9). The polypeptide backbone can be grouped into several domains including the amino terminal disulphide-rich von Willebrand factor (VWF)-D-like and VWF-C-like domains, the cystine-rich CK (cystine-knot) and CYS domains and PTS-rich tandem repeat domains (Perez-Vilar & Hill, 1999; Dekker et al., 2002) (Figure 1.8). The PTS domain is an important substrate for O-glycosylation especially O-GalNAc.

![Figure 1.8. Schematic representation of typical mucin domains](image)

Top, membrane-bound mucins (MUC1); bottom, secreted mucins (MUC2). Empty spaces represent unique sequences.
The main characteristic of mucins is the variable number of tandem repeat (VNTR) regions that are rich in serine and threonine (PTS domain) and are highly glycosylated with O-GalNAc glycans. These regions lack any secondary structure, thus facilitating O-glycosylation of the folded protein. The O-glycans are hydrophilic and usually negatively charged and therefore promote the binding of water and salts which are the main contributors to the viscosity and adhesiveness of mucus (Brockhausen et al., 2009). Oligosaccharides which often comprise 50–80% of mucin mass also play important roles in providing binding sites for microorganisms, mimicking those expressed by epithelial cells, thereby promoting their removal by mucus flow (Hurd et al., 2005). Glycosylation patterns contribute different properties to the mucins and thus alter the potential for their biological interactions (Patsos & Corfield, 2009). The biophysical properties of mucins are speculated to be related to their extensive O-glycosylation rather than directly to their polypeptide backbones (Dekker et al., 2002). For these reasons, comprehensive characterisation of glycans attached on the mucins as well as mucinous tissues is necessary to better understand their biological roles. However, many of these functionally and structurally important glycans have only relatively recently become the focus of rigorous characterisation. Despite this, considerable progress has been made. For example, the application of sensitive mass spectrometry techniques has enabled detailed structural profiling on isolated human MUC16/CA125- and MUC2-derived glycans (Kui Wong et al., 2003; Holmen Larsson et al., 2009). In addition, Issa and colleagues recently described rigorous sequencing of oligosaccharides derived from semi-purified salivary agglutinin, a mucin-like glycoprotein, using LC-ESI-MS^n in the negative ion mode (Issa et al., 2010). Nevertheless, analyses on mucinous tissues/organs so far have provided data mostly restricted to glycan mass and compositions without rigorous structural profiling, for example analyses on human intestinal tract (Robbe, 2004; Robbe-Masselot et al., 2009) and murine colon (Hurd et al., 2005).

The nomenclature of mucins is based on their protein backbones, which are encoded by MUC genes. Nevertheless, a standard criteria to define a glycoprotein as a mucin is not readily available, which has resulted in some discrepancies in the number of designated MUC genes (Dekker et al., 2002; Rose, 2006). According to the Human Genome Organization Gene Nomenclature Committee (www.genenames.org) (Bruford et al., 2008), currently there are 20 MUC genes in humans. However, their homologs in the murine genome are poorly characterised and are associated with a number of ambiguities (Escande et al., 2004) (Table 1.3). For example, earlier characterisation of murine Muc5B and Muc6 showed expression of these mucins in stomach and submaxillary glands, respectively (Escande et al., 2002; Desseyn & Laine, 2003),
but they were not detected in more recent work (Escande et al., 2004). Furthermore, a previous study revealed that murine Muc3 has higher sequence identity to human MUC17 than human MUC3 (Moniaux et al., 2006), thus raising the possibility of discrepancy in other human-mouse orthologue assignments that can complicate the understanding of their respective biological functions.

Table 1.3. List of known murine mucin genes with respective human orthologues and expression sites.

This information was extracted from the database of Murine Genome Informatics (www.informatics.jax.org) (Bult et al., 2008).

<table>
<thead>
<tr>
<th>Mucin name</th>
<th>Human orthologue</th>
<th>Expression sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muc1</td>
<td>MUC1/polymorphic</td>
<td>Colon, breast (Spicer et al., 1991; Van Seuningen &amp; Jonckheere, 2008)</td>
</tr>
<tr>
<td></td>
<td>epithelial mucin (PEM)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(70% homology)</td>
<td></td>
</tr>
<tr>
<td>Muc4</td>
<td>MUC4 (60% homology)</td>
<td>Human trachea-bronchial mucosa (Desseyn et al., 2002; Van Seuningen &amp; Jonckheere, 2008)</td>
</tr>
<tr>
<td>Muc12</td>
<td>MUC12</td>
<td>Colon (Williams et al., 1999; Van Seuningen &amp; Jonckheere, 2008)</td>
</tr>
<tr>
<td>Muc13</td>
<td>MUC13</td>
<td>Intestines, trachea, stomach, kidney (Van Seuningen &amp; Jonckheere, 2008)</td>
</tr>
<tr>
<td>Muc16</td>
<td>MUC16/CA125 (48%</td>
<td>Female reproductive organ, cornea, trachea (Cheon et al., 2009)</td>
</tr>
<tr>
<td></td>
<td>homology)</td>
<td></td>
</tr>
<tr>
<td>Muc3/ Muc17</td>
<td>MUC17 (60% homology)</td>
<td>Intestine (Moniaux et al., 2006)</td>
</tr>
<tr>
<td>Muc20</td>
<td>MUC20 (61% homology)</td>
<td>Kidney, placenta, lung, liver (Van Seuningen &amp; Jonckheere, 2008)</td>
</tr>
<tr>
<td>Epiglycanin</td>
<td>MUC21</td>
<td>Lung, large intestine, thymus, testis (Itoh et al., 2008)</td>
</tr>
</tbody>
</table>

| Secreted mucins | Muc2                | Small & large intestines (Van Klinken et al., 1999; Escande et al., 2004)       |
|                 | MUC5AC              | Stomach (Escande et al., 2004)                                                |
|                 | MUC5B               | Laryngo-tracheal tissue (Escande et al., 2004)                                 |
|                 | MUC6                | Stomach (Escande et al., 2004)                                                |
|                 | MUC9                | Ovary tract (Lyng & Shur, 2009)                                               |
The properties of a mucin are highly influenced by internal and external factors. For example, the composition of mucinous layers is regulated by various genes and epigenetic factors such as DNA methylation and histone modifications (Vincent et al., 2008). The thickness of the layers has been shown to vary throughout the gastrointestinal tract (Atuma et al., 2001), probably to adapt to the specific environment of each section. It is also believed that mucin expressions are influenced by other exogenous factors, for instance the colonisation by commensal and temporary microorganisms (Marcos et al., 2008).

Mucins are the main component of the mucosal barrier in the respiratory, reproductive and gastrointestinal (GI) tract. The highly O-glycosylated domains of mucin (the tandem repeats) form a semi-rigid and relatively inflexible structure (Perez-Vilar & Hill, 1999), making these molecules excellent as a component of protective barriers. The mucosal barrier is made up of several layers (Patsos & Corfield, 2009). The uppermost layer is the mucus gel which constitutes many secreted mucinous proteins and peptides. Beneath the mucus gel is the glycoalyx which mainly consists of membrane-bound mucin glycoproteins and glycolipids. The final component of the mucosal barrier is comprised of the mucosal cells that supply mucins and serve as a physical barrier against invasion by bacteria. In addition, mucosal layers also integrate both innate and adaptive immune elements (Corfield et al., 2001). Their strategic position places the mucins at centre stage in many disease processes in which interactions between epithelial cells and their surroundings have been interrupted (Dekker et al., 2002), usually involving lectins, as discussed earlier. However, most of these interactions and many other properties of the mucus barrier remain to be fully elucidated.

Mucins have been implicated in many important biological events including inflammation, host-pathogen interactions and cancer metastasis (Hanisch & Muller, 2000). For example, mucins are expressed or secreted by epithelial cells towards the lumen of a healthy vesicular organ, but in malignant cells, mucins are expressed on all aspects of the cells and could enter the extracellular space or body fluids (Varki et al., 2009). Cancer cells also express aberrant amounts or forms of mucin, the knowledge of which is being utilised towards the prevention, detection, control of and cure for carcinoma (Hollingsworth & Swanson, 2004). For instance,
CA125 has provided a useful serum tumour marker for monitoring responses to treatments for ovarian cancer (Bast et al., 2005). The density of glycosylation on mucins expressed by different tumours and normal cells is also highly variable, and is believed to contribute significantly to normal and aberrant functions during pathogenesis of cancer and other diseases (Hanisch & Muller, 2000). However, due to the mucinous biophysical characteristic and the relatively large mass, it has been very challenging to fully define mucin oligosaccharide components, which has resulted in a very limited understanding of their functions and regulations.

### 1.2.5 *Helicobacter pylori*: a specialist in human stomach colonisation

*H. pylori*, a gram negative bacterium, infects more than 50% of the human population and is the only bacterial agent known to cause cancer (Fritz & Van Der Merwe, 2009). There might be no single most prominent method of human infection, but it has been shown that *H. pylori* infects humans from childhood (Dore et al., 2002) and, when present, it is the most dominant microbe in the human stomach (Cover & Blaser, 2009). The outcome of this infection ranges from asymptomatic to gastric lesion to gastric adenocarcinoma. Nevertheless, only a small subset of the infected population develops any disease. The mechanism of *H. pylori* infection transmission remains unclear but appears to be infrequent compared to other microflora, thus making its adherence to host cell surfaces a critical early step in colonisation (Testerman et al., 2001).

The complete genome sequence of *H. pylori* strain 26695 published in 1997 has identified many putative adhesins, lipoproteins and other outer membrane proteins, highlighting the potential complexity of host–pathogen interactions (Tomb et al., 1997). Variation in both host and *H. pylori* in the expression of glycan epitopes and also the types of bacterial adhesins may explain the wide range of clinical outcomes following the colonisation of *H. pylori* (Victor & Esko, 2009). Although the mechanism of how *H. pylori* infects humans and causes diseases has been intensively studied, it remains to be fully understood. The existence of various strains of *H. pylori* with up to 25% variations in gene content (Fritz & Van Der Merwe, 2009) adds to the complexity of these studies. The lipopolysaccharide (LPS) of *H. pylori* is replete with fucosylated Lewis antigens including Le\(^a\), Le\(^b\), Le\(^x\) and Le\(^y\) (Wang et al., 2000), and this could facilitate the evasion of the host immune response (Appelmelk et al., 1996). This phenomenon is known as bacterial mimicry, as it is trying to imitate the surface of stomach mucosa that is highly O-glycosylated and often decorated with ABO blood group and Lewis antigens. This illustrates how evolution has well adapted *H. pylori* to colonise the human stomach.
Mucin type O-glycans play important roles in the pathogenesis process of *H. pylori* infection. The establishment of a successful infection is highly dependent on the bacterium’s capability to tightly adhere to the highly O-glycosylated gastric mucosal layers and epithelial cells (Blaser, 1993). This binding is mediated by *H. pylori* adhesins that recognise glycan structures on host glycoconjugates. The inability to adhere to the mucosal layer will result in the bacteria’s rapid removal by host-defense mechanisms including peristalsis and ciliary activity (Borén et al., 1994) (illustrated in Figure 1.9). The two most studied adhesins of *H. pylori* are blood group antigen-binding adhesin (BabA) and sialic acid-binding adhesin (SabA). A few lipoproteins of *H. pylori* have been speculated to also contribute to the adherence capacity of this organism, for example *H. pylori* adhesin A (HpaA) and HpaA orthologue (Tomb et al., 1997). Successful adhesion is critical for efficient delivery of bacterial virulence factor, for instance the cytotoxin-associated gene A (CagA), via type IV secretion system that will damage host tissues and induce pathological conditions (Hatakeyama, 2004). The binding of the most studied *H. pylori* adhesins and the translocation of its virulence factor is illustrated in Figure 1.10.

BabA-mediated binding of *H. pylori* is important for initial adhesion to mucosal and epithelial layers. BabA is encoded by the babA2 gene and has been shown to bind to fucosylated histo-blood group antigens Le\(^b\) and H type 1 but not to H type 2 (Borén et al., 1994; Ilver et al., 1998). However, this binding was only demonstrated in vitro using synthetic oligosaccharides and antibody staining. Furthermore, the influence of these antigens in successful adhesion of the whole bacterium using BabA has not been demonstrated. Bovenkamp et al. later clearly showed that *H. pylori* strains which bind to Le\(^b\) also bind to MUC5AC in gastric tissue, indicating that MUC5AC is an important carrier of Le\(^b\) in the stomach (Bovenkamp et al., 2003). *H. pylori* binding to H type 1 and type 2 antigens however remains ambiguous. Recently, loss of expression of BabA during a course of experimental infection was demonstrated in *H. pylori* strain J166, indicating dynamic expression of BabA as a mechanism to adapt to changes in host glycan expression (Styer et al., 2010). Persistent colonisation by *H. pylori* results in inflammation with concomitant expression of sialylated glycans including SLe\(^a\) and SLe\(^x\), which are aimed to guide leukocyte migration via selectins. Sialylated glycans are recognised by SabA. Therefore this adhesin functions mainly in maintaining *H. pylori* binding as the host responds to epithelial damages (Mahdavi et al., 2002), taking over the responsibility from BabA.
Figure 1.9. *H. pylori* adherence to gastric mucosa prior to colonisation on epithelial cells.
Schematic diagram of a stomach cross-section showing *H. pylori* crucial initial binding to the mucins of mucosal layer and later to the surface of epithelial cells to avoid being swept away by gastric flow and mucus flow.

Figure 1.10. Interaction of the most studied *H. pylori* adhesins with host cell receptors leading to the translocation of CagA.
TLR, toll-like receptors; LPS, lipopolysaccharides. Based on (Testerman *et al.*, 2001; Algood & Cover, 2006).

1.2.6 Regulation of protein O-glycosylation

As discussed earlier (Section 1.2), protein N-glycosylation is initiated on newly synthesised unfolded/partially folded proteins in the endoplasmic reticulum (ER) with *en bloc*
transfer of dolichol-linked precursors to Asn. One of the types of O-glycosylation, the O-
mannosylation of serine/threonine, also occurs in the ER. The synthesis of O-GalNAc-linked
glycans, on the other hand, begins in the *cis*-Golgi on matured and already folded proteins, and
only those with serine/threonine containing domains appropriately accessible to the ppGalNAcTs
will be glycosylated with mucin-type glycans. Most extension and capping glycosyltransferases
are localised on the membranes of medial- and trans-Golgi, though some enzymes are distributed
throughout the cisternae. A schematic diagram of N-glycan, O-GalNAc and O-mannose
biosynthesis on polypeptides from the ER to trans-Golgi, illustrating some of the main
glycosylating enzymes involved, is shown in Figure 1.11. The distribution of glycosyltransferases
in specific cisternae of the Golgi favours the synthesis of certain glycans by dictating the order of
action of the enzymes and limiting enzymatic competition (Tu & Banfield, 2010). It is envisaged
that the localisation of these glycosyltransferases greatly influences the final glycosylation
products, which normally varies according to the state of the cell and external inputs.

Cells adapt multiple aspects of their physiology including protein synthesis and
glycosylation following cues from extracellular molecules, for example growth hormones.
Regulation of these processes involves signal transduction molecules forming complexes with
specific subcellular components. A family of signalling molecules known as the Src family
kinases (SFKs) has been shown to form complexes on Golgi membranes (David-Pfeuty &
Nouvian-Dooghe, 1990). SFKs are non-receptor tyrosine kinases that interact and modify
numerous cellular cytosolic, nuclear and membrane proteins by phosphorylating tyrosine
residues. SFKs comprise nine members including Src, Yes and Fyn that are expressed in a broad
range of tissues (Boyer *et al.*, 2002).
Figure 1.11. Biosynthesis of N-glycans, O-GalNAc and O-mannose on proteins from endoplasmic reticulum (ER) to trans-Golgi.

Structures shown are symbolic of the most common products for each compartment with glycosyltransferases involved. OST, oligosaccharyltransferase; GnTs, GlcNAc transferases; ST3Gal, α2,3-sialyltransferase; ST6GalNAc, α2,6-sialyltransferase.

Src is known for its involvement in the regulation of Golgi structure and KDEL-receptor (KDEL-R)-dependent retrograde transport from the Golgi to the ER via coat protein complex-I (COP-I) mediated transport (Yamamoto et al., 2001; Bard et al., 2003). KDEL-R recognises the sequence Lys-Asp-Glu-Leu, an ER-retention sequence which exists in most membrane-bound proteins (Yamamoto et al., 2001). Specifically, it has been demonstrated by Bard et al. that in the absence of Src, the transport of a KDEL-R-bound protein from Golgi to ER was accelerated, which means it was not properly regulated (Bard et al., 2003). However, the exact physiological roles of Src at the Golgi remains poorly characterised. COP-I-coated vesicles are vesicular carriers that function in the secretory pathway. The mechanism of cargo (Golgi resident enzymes)
uptake and delivery by COP-I vesicles involves several cytosolic and transmembrane molecules (Béthune et al., 2006). The retrograde transport of luminal and membrane proteins in the ER-Golgi network is the best characterised function of COP-I-dependent transport (Beck et al., 2009) (Figure 1.12). According to Beck and colleagues (Beck et al., 2009), first, GTPase Arf1 is activated by a GTP exchange factor GEF1 before dissociating from a dimeric p24 (a family of transmembrane proteins) and then Arf1 itself dimerises. Activated Arf1-GTP then recruits coatamer to form a complex together with monomeric p24 on the cytoplasmic surface of Golgi membrane. Coatamer functions to capture Golgi resident transmembrane proteins including KDEL-R. The coat then polymerises and pinches off from the Golgi membrane to form a COP-I vesicle carrying KDEL-R possibly bound to its ligands as well as soluble lumen proteins trapped in the vesicles. Arf1 can be deactivated by the action of GTPase activating protein 1 (GAP1) through GTP hydrolysis, causing the coat to dissociate from the vesicle (Beck et al., 2009). It was also suggested that KDEL-R could induce the interaction between GAP1 and Arf1 (Yamamoto et al., 2001). The ‘nude’ vesicle can then fuse to the target membrane (ER) and deliver the cargo. Retrograde traffic contributes to an ER quality control mechanism that ensures correct assembly of certain membrane proteins (including glycosyltransferases) by recycling material back to the ER in opposition to anterograde traffic by COP-II (Béthune et al., 2006).

Figure 1.12. COP-I-mediated retrograde and COP-II-mediated anterograde transportation between cis-Golgi and ER.
Elevation of Src kinase activity has been implicated in the manifestation of cancer (Ottenhoff-Kalff et al., 1992) and also in metastases relative to normal tissues (Talamonti et al., 1993). A recent study documented the positive correlation of Src activity with the loss of epithelial differentiation associated with the increase of the metastatic potential of carcinoma cells (Boyer et al., 2002). Accordingly, it has been shown many times that O-glycosylation and O-glycan structures vary dramatically during inflammation and in cancer. For example, as has been discussed earlier, inflammation caused by H. pylori infection induces sialylation of the gastric mucosa (Mahdavi et al., 2002). During the progression of carcinomas, O-glycan structures (O-GalNAc) were shown to be simplified with Tn, sialyl Tn, core 1 and sialylated core 1 in colon and breast cancer cells (Itzkowitz et al., 1990; Brockhausen, 1999; Burchell et al., 1999). In addition, the level of ppGalNAcT-3 has been demonstrated to correlate with tumour differentiation, disease aggressiveness and prognosis in patients with colorectal carcinoma (Shibao et al., 2002). The reported changes in glycan structures and enzyme levels are among the cellular responses that are likely to carry a physiological importance by altering cell adhesive properties in order to facilitate the invasion and metastasis through interactions of O-GalNAc glycans with lectins and cell surface receptors (Brockhausen, 2006).

Whether Src activation regulates the aforementioned cellular transformations is yet to be ascertained. However, recent findings from Bard et al. have intriguingly indicated a possible connection between Src activation and the regulation of protein O-glycosylation (Gill et al., 2010). They have demonstrated that growth hormone stimulation redistributes ppGalNAcTs from the Golgi to the ER and this was dependent on Src activation. Simultaneously, the stimulation also induced the redistribution of COPI-I vesicles from Golgi to ER with ppGalNAcTs being found to colocalise with the coat components and remained active in the ER. Furthermore, constitutively active mutant Arf1 was shown to block or delay the redistribution of ppGalNAcTs, indicating its involvement in the regulation of COP-I transport. Redistribution of ppGalNAcTs to the ER allows the enzyme access to unfolded or partially folded protein substrates. Accordingly, ppGalNAcTs were speculated to be able to initiate more O-glycosylation than normally occurs on folded proteins in the cis Golgi. This was supported by an increase of total helix pomatia lectin (HPL) staining (which recognises GalNAc residues) between Src-deficient cells and activated Src cells. In addition, modified Muc1 coexpressed with active Src was shown to contain a significantly higher density of incorporated GalNAc analogue, GalNAz (N-azidoacetyl-galactosamine acetylated), as judged by anti-GalNAz staining on Western blotting relative to coexpression with inactive Src.
Furthermore, initial O-glycomic screening by a colleague at Imperial College of some of the Bard group’s cells (unpublished data of Aristotelis Antonopoulos, reproduced with permission in Figure 1.13), interestingly showed a significant reduction of core 2 O-glycans relative to core 1 O-glycans in cells with active Src (SYFsrc) compared to those with inactive Src (SYF). This observation could be explained by assuming that the postulated higher number of glycosylated serine/threonine residues resulted in O-glycans that are attached nearer to each other on the same peptide and consequently become less accessible for C2GnTs modification in the cis-Golgi. Putting all this information together, it is tempting to postulate that Src responds to signals from extracellular molecules and regulates protein O-glycosylation (O-GalNAc) to alter the properties of the cell surface by influencing the localisation of ppGalNAcTs from cis-Golgi to ER via COP-I-mediated retrograde transport. This model is illustrated in Figure 1.14.

Figure 1.13. MALDI-TOF mass spectra of O-glycans derived from (TOP) cells with inactive Src (SYF) and (BOTTOM) cells with active Src (SYFsrc).
Peaks for core 1 and core 2 O-glycans are coloured green and red, respectively. Inset depicts the T antigen at m/z 534. Reproduced from the work of Aristotelis Antonopoulos, Imperial College London.

Figure 1.14. Illustrative model for Src-induced COP-I-mediated retrograde trafficking of ppGalNAcTs.
Stimulation of EGF via its cell surface receptor activates Src at the Golgi. High levels of activated Src kinase induce relocalisation of ppGalNAcTs from the Golgi to the ER via COP-I mediated retrograde transport. Modified from (Gill et al., 2010).

To provide more evidence for this proposed mechanism of the regulation of protein O-glycosylation and to confirm that their findings are a general phenomenon, not an artefact of the cell line chosen for study, our collaborators, Frederic Bard et al. from the National University of Singapore have generated several cell lines with different levels of activated Src or ER/Golgi-bound ppGalNAcT-2. O-Glycomic analyses of these cell lines is presented in Chapter 5 and discussed in correlation with the data from our collaborators.

1.2.7 Glycomics and glycoproteomics

Glycobiological research has been well supported and enriched by glycomic and glycoproteomic analysis. Glycomics is the profiling of the total glycan population in a sample under study which usually requires prior glycan isolation from glycoconjugates. N-glycans can be released specifically from glycoproteins by using the enzyme N-glycanase whereas O-glycans are preferably removed by using a chemical method, reductive β-elimination, as there is no enzyme known so far that can release all types of O-glycans efficiently. These techniques will be
discussed in detail in Section 1.4.8.1. O-Glycomics, which is the focus of this thesis, is a glycomic analysis optimised to achieve the best characterisation of O-glycans.

Glycoproteomics refers to studies defining glycan repertoires on specific peptides or even individual glycosylation sites in glycoproteins. This type of analysis requires prior purification and isolation of the glycoproteins of interest. In other instances where the glycomic analysis of specific tissue or cells has revealed interesting glycan structures, the glycoproteomic approach is utilised to identify the glycoprotein that carries that particular glycan.

Together with the advancement of mass spectrometry, structural studies of the glycome and glycoproteome are continuing to steer glycobiology to become an essential complement for other life science research. For instance, glycobiology is now integral to some aspects of other longer-established fields such as immunology, parasitology and oncology (Ni & Tizard, 1996; Kui Wong et al., 2003; Lee et al., 2005). A complete structural characterisation of oligosaccharides demands definition of branching, linkages, configurations and the identification of same-mass sugar isomers (Dell & Morris, 2001). All these approaches require continuing optimisation and enhancement to address the increasing complexity and specificity of biological samples under study.

1.3 Animal models in functional glycomic research

Functional glycomics requires the continuous development of reliable and sensitive methods for the characterisation of glycan structures and their integration into studies of structure-function relationships. As described earlier, distortions in glycosylation related genes cause a series of human diseases. To explore the biochemical aetiology and systemic view of these diseases, suitable animal models are required to delineate oligosaccharide functionality. Among available animal models, the mouse model is the most widely used because of its close genetic and physiological similarities to humans. Moreover, its genome is easier to manipulate and analyse. Nevertheless, the use of other animal models especially from non-mammalian organisms are necessary for instance to understand glycan function from an evolutionary point of view (Honke & Taniguchi, 2009).
1.3.1 Mouse models

*Mus Musculus*, the house mouse, appeared as a distinct species at about 8000 years ago. Since the late 19th century, scientists have shown great interest in research on murine genetics. The draft sequence of the mouse genome was published in 2002 and it was discovered that about 40% of the human genome can be directly aligned with the mouse genome, and about 80% of mouse genes have one corresponding gene in the human genome (Waterston, 2002). On top of that, mouse development, body plan, physiology, behaviour and diseases have much in common with those of humans. However, there are still some limitations that need to be taken into consideration when interpreting glycomic data from mouse models, for instance as discussed earlier, there are different numbers and patterns of expression of mucin and glycosyltransferase genes in humans and mice.

With the introduction of more sophisticated technologies, new mouse strains were created, for instance via gene knock-in, knock-down or knockout (KO), which have specific characteristics that resemble a human disease or disorder. The knockout mouse technique was first created between 1987 and 1989 separately by Mario Capecchi, Martin Evans and Oliver Smithies, winners of the 2007 Nobel prize in Medicine (Honke & Taniguchi, 2009). Since then, this technique has been extensively used to reveal important aspects of glycan functions in mammalian systems (Austin *et al.*, 2004). Knockout experiments on glycosyltransferase genes were first reported by two groups separately in 1994 (Ioffe, 1994; Metzler, 1994). Both groups had knocked out the gene *Mgat-1* in the murine genome to produce mice with deficiency in β-1,2-N-acetylgalcosaminyltransferase I, the enzyme that catalyses the initiation of the first complex-type antenna of N-glycans (see Figure 1.11), which resulted in murine embryonic lethality. After more than 20 years, the knockout mouse technique in glycobiological research is still relevant and improving and now even allows investigations on specific isoforms of a glycosyltransferase. For instance, Akama *et al.* have demonstrated the mutual compensatory roles of α-mannosidase-II and α-mannosidase-IIx in N-glycan processing in mice (Akama *et al.*, 2006).

Although mice models have been widely used for almost 100 years, there are still many aspects of murine biological systems, glycosylation for instance, yet to be fully understood. Even for wild type mice, very little work has been done until now to rigorously characterise their O-glycosylation despite the importance of mouse models in human biomedical research. For example, as has been discussed earlier, basic information such as the type of Lewis antigens that
can be synthesised in mice is still debatable. The functional properties of C2GnTs, the enzymes producing the most abundant O-glycan core type in mouse, have not yet been fully elucidated. There has been only a limited amount of detailed O-glycomic profiling of normal murine tissues, for instance the partially characterised O-glycan structures at the CFG database (www.functionalglycomics.org) and also the work of Dell’s group (North et al., 2010), let alone the delineation of glycan alterations in glycosylation-related diseases induced in mouse models.

It is essential then to rigorously characterise the mouse biological system to take full advantage of this model and to have better insights into human biology and diseases. Therefore, while addressing important biological questions, the present work also intended to provide a more comprehensive structural characterisation of the O-glycomes of various wild type murine tissues.

1.4 Mass spectrometry: principles and applications

Mass spectrometry is currently the technology of choice to fulfil the needs of glycomic and glycoproteomic research. This technique has been at the forefront of this research for more than two decades (Haslam et al., 2006) and is still remarkably evolving with advancing modifications. Mass spectrometry can be exploited with various types of instrumentation, including matrix assisted laser desorption ionization-time-of-flight (MALDI-TOF/TOF), electrospray-quadrupole-time-of-flight (ESI-QTOF), MALDI-quadrupole ion trap-TOF (MALDI-QIT-TOF) and gas chromatography-electron ionisation (GC-EI), which will be discussed in the following sections. The combination of various mass spectrometric techniques is very powerful and has advanced our understanding of glycan structures and functions.

1.4.1 History

Mass spectrometry is said to begin with the work by J.J. Thomson in the year 1897 when he discovered the electron and determined its mass-to-charge ratio (m/z) (De Hoffmann & Stroobant, 2007). He won a Nobel Prize in 1906 and six years later he constructed the world’s first mass spectrometer (Thomson, 1913). Electron ionisation (EI), formerly known as electron impact, is the original mass spectrometry ionisation technique introduced by A.J. Dempster in 1918. Since this pioneering work, vast improvements of mass spectrometers, as well as the invention and optimisation of ways to introduce and separate samples in a mass spectrometer, have been made. For instance, GC was first coupled to a mass spectrometer in 1956 by F.W.
McLafferty (McLafferty, 1957) and R.S. Gohlke (Gohlke, 1959). About 30 years later, another important addition to the mass spectrometry repertoire was the discovery of MALDI-MS by two groups separately, Tanaka et al. and Hillenkamp et al. (Karas et al., 1987; Tanaka, 1988). The invention of the time-of-flight mass analyser (TOF) preceded MALDI by 40 years (back in 1946 by W. E Stephens (Wolff & Stephens, 1953), whereas the reflectron mode, which is a major improvement of TOF, was introduced in 1972 (Mamyrin et al., 1973) but not effectively exploited until the 1990’s. Now, MALDI and TOF is a common and powerful combination. ESI, on the other hand, is the brainchild of M. Dole (Dole et al., 1968) but did not become a useful tool until the pioneering efforts of J. Fenn in the 1980s. In a landmark piece of research, Fenn produced the first spectra of proteins above 20 kDa in 1988 (Fenn et al., 1989). Both Tanaka and Fenn shared a chemistry Nobel Prize in 2002 for their excellent work in analysing biomolecules using mass spectrometry. In 2004, Tanaka supplemented his contribution to the mass spectrometry world by inventing the MALDI-QIT-TOF-MS (Ojima et al., 2005).

In the past 100 years, mass spectrometry has progressed from the analysis of elements and small volatile molecules to large biological macromolecules, practically with no mass limitations (El-Aneed et al., 2009). This has been made possible by refinements of every component of mass spectrometers, which have led to spectacular improvements in resolution, sensitivity, mass range and accuracy. In the last 20 years especially, the rapid development of mass spectrometers has seen the invention of entirely new instruments as well as hybrid instruments which were realised by combinations of the same or different analysers (De Hoffmann & Stroobant, 2007). Discussion all of these however, is beyond the scope of this thesis. Therefore, only instruments that are relevant to this thesis plus a few other prominent instruments will be discussed below.

1.4.2 Components

Mass spectrometry instruments generally consist of three main components, namely an ionisation source, one or more mass analysers and one or more detectors. The instruments may include collision cells for ion activation in tandem mass spectrometric analysis. All these components are important in order to form a robust instrument and their different combinations can produce unique analytical capabilities.
An ionisation source is where samples are converted to gaseous ions. How this is achieved is crucial for the successful application of mass spectrometry to diverse samples including those which are neutral and/or large. Some ionisation techniques are very energetic and can cause fragmentation during the ionisation process, for example EI. On the other hand, there are softer ionisation techniques that produce mainly ions of the molecular species, for instance ESI and MALDI. All the aforementioned ionisation sources are applied in the work included this thesis and therefore each will be discussed in more detail in later sections.

The mass analyser is the component of a mass spectrometer that separates gas-phase ions according to their \( m/z \). The separation can be based on many principles; hence, there are several types of mass analysers with different advantages and limitations. Among the important characteristics of a mass analyser include resolution, mass range, scan rate and detection limit (Van Bramer, 1997). The current trend in mass analyser development is to combine different analysers in order to increase the versatility and allow multiple experiments to be performed (De Hoffmann & Stroobant, 2007). The mass analysers that are relevant to this thesis are TOF, quadrupole and QIT and will be discussed in more detail in the subsequent sections.

A detector transforms mass analysed ions into usable signals, usually electric currents, which are proportional to the ion abundances. A detector is selected based on its speed, dynamic range, gain and high voltage. There are two types of detectors utilised in the mass spectrometry instruments relevant to this thesis, microchannel plate (MCP) and photomultiplier. In both detectors, mass analysed ions that strike a conversion dynode (electrode) are first converted to secondary electrons before being transferred to an electron multiplier. The MCP detector consists of a plate in which parallel cylindrical channels have been drilled with the inner sides covered by a semiconductor substances, each serves as a continuous electron multiplier dynode (De Hoffmann & Stroobant, 2007). The secondary electrons are then amplified by a cascade effect every time they hit the inner side of a dynode and are finally collected by a metal anode to produce a current signal and submitted to the data system (Smith, 2004). The photomultiplier detector, on the other hand, contains a phosphorescent screen that converts secondary electrons to photons prior to entering a high vacuum photomultiplier tube in which photons are converted back to electrons by a photocathode. The secondary electrons are then multiplied based on the same principles as the MCP except that the electrons are transferred between several discrete dynodes according to a potential gradient.
Ion activation is a group of numerous techniques where mass separated/selected ions (precursor ions) are given excess energy either by collision or by irradiation to dissociate. Dissociations may either occur spontaneously (metastable ions) or can result from intentionally supplied additional activation in a collision cell (Gross, 2004). Spontaneous fragmentation occurring within the ion source is known as in-source decay, whereas outside the ion source is known as post-source decay (Harvey, 2010). Methods for ion activation include surface-induced dissociation, infrared multiphoton dissociation, electron capture dissociation and collision-induced dissociation (CID). The most prominent ion collision technique is the CID, which generally involves passing an ion beam through a collision cell that contains collision gas, for instance helium, argon and nitrogen, at high pressure. CID is very useful for structural elucidations of ions of low internal energy because it allows for the fragmentation of gaseous ions that were stable before the activating process (Gross, 2004). In tandem mass spectrometric analysis, the precursor ions are normally dissociated between the two mass spectrometric stages. Tandem mass spectrometers can be categorised in two ways: performing tandem MS in space by the coupling of two physically distinct analysers (for example TOF/TOF and QTOF), or in time by performing a sequence of fragmentations in an ion storage device (for example QIT) (De Hoffmann & Stroobant, 2007). In an MS/MS (MS²) experiment, selected precursor ions are cleaved into smaller fragments which then will be separated and detected. In multiple stage MS analyses (MSⁿ), the fragments of the precursor ions can be selected and further fragmented and this process is repeated during the course of each of the activation stages.

1.4.3 MALDI-TOF/TOF mass spectrometry

In oligosaccharide analysis, MALDI-TOF/TOF-MS is currently the most powerful mass spectrometric method for mass fingerprinting because of its sensitivity of detection and ability to analyse glycans from complex mixtures derived from a variety of organisms and cell lines (Haslam et al., 2006; Parry et al., 2007). This instrument consists of a MALDI source, a short linear TOF, a collision chamber for CID, a second TOF with a reflectron and two MCP detectors, one each for the linear and reflectron modes (Figure 1.15).
Figure 1.15. Schematic view of a MALDI-TOF/TOF mass spectrometer.

The schematic shows ion paths of reflectron mode during tandem mass spectrometric experiment. Circles of different sizes represent ions of different $m/z$ while circles of similar size but with different colours represent ions of similar $m/z$ but with different level of kinetic energies. After passing the reflectron, ions of the same $m/z$ but differing initial kinetic energy are eventually recorded at the detector as the same signal.

As the name implies, ionisation in a MALDI source requires biomolecules to be mixed with a low molecular weight ultraviolet-absorbing compound, known as the matrix. As the solvent dries on the target plate, the matrix compound crystallises together with the dissolved biomolecules. Sample spots are then irradiated with a laser beam, that causes evaporation and ionisation, and then the ions are directed towards the TOF analyser (Figure 1.15). The ionisation process is still not fully understood but it is widely believed that the matrix allows the energy from the laser to be dissipated and assists the formation of ions by proton transfer and chemical processes (Dell et al., 2008). It has been suggested that incorporation of analyte into the matrix crystal and the amount of contact between analyte and the matrix surface can influence the strength of MALDI signals (Harvey, 2010). Important features of a matrix, additional to having an absorption frequency compatible with the MALDI laser, include sample solubility, reactivity, volatility and suitable desorption properties (Hossain & Limbach, 2010). Most matrices employed for analysing substances in the positive ion mode are acidic, for example α-cyano-4-hydroxy cinnamic acid (CHCA) and 2,5-dihydroxybenzoic acid (DHB), which helps the ionisation of biomolecules (Kinter & Sherman, 2000). The MALDI ion source produces mainly singly charged molecular ions with minimal fragmentation, making it ideal for accurate overall glycan profiling (Dell et al., 2008). In-source metastable fragmentation was found to be prominent in early MALDI studies of glycans. Although this allows PSD of oligosaccharides in MALDI-TOF instruments it also complicates MS profiling. This drawback was dealt with by increasing the pressure in the MALDI source and by stabilising glycan molecules via
derivatisation (Zaia, 2004). Furthermore, ionisation in MALDI produces a pulsed sample ion current, which is ideally suited to the TOF mass analyser.

In a MALDI-TOF-MS analysis, excited ions from the ion source are attracted towards the TOF analyser where ions of different $m/z$ are dispersed in time during their flight along a field-free drift linear path of known-length; the lighter ions arrive earlier at the detector than the heavier ones (Gross, 2004). Ions generated by hundreds of laser shots are accumulated from different points of laser irradiation, which make the MALDI-TOF technique excellent in mass spectrum reproducibility (Wada et al., 2007). TOF analysers were originally designed for GC (Gohlke, 1959), but at present TOF is very commonly coupled to MALDI with the capability to produce mass spectra of proteins of at least 100 000 Da. Over the years, various modifications have been made to instruments with TOF analysers, for example incorporation of delayed pulsed extraction and reflectrons, both targeted to improve the mass resolution by correcting the energy dispersion of ions with the same $m/z$ but with different kinetic energy so that they arrive at the detector at the same time. In the delayed pulsed extraction, ions are initially allowed to separate according to their kinetic energy in a field-free region before an extraction pulse, thus permitting the initially less energetic ions to receive more kinetic energy and join the initially more energetic ions at the detector. The reflectron, meanwhile, creates a retarding field at the end of the TOF tube and acts by deflecting the ions back through the flight tube. Ions with higher kinetic energy and hence with more velocity will penetrate the reflectron much deeper than ions with less kinetic energy, thus the high energy ions will spend more time in the reflectron (De Hoffmann & Stroobant, 2007). TOF has almost unlimited mass range but the reflectron mode is restricted to analyses below masses of about 10,000 Da (Dell et al., 2008).

Each component of the MALDI-TOF mass fingerprint can be rigorously characterised by subjecting each molecular ion to collisional activation in MS/MS experiments. The combination of a short linear TOF and a reflectron TOF analyser (TOF/TOF) separated by an ion selector and a collision cell (Figure 1.15) enhances tandem mass spectrometric analyses in MALDI-MS and is currently the leading type of mass analyser for MALDI instruments. In MALDI-TOF/TOF-MS/MS, excited ions from the MALDI source are separated in the first TOF analyser and selected with a timed ion selector (TIS) or mass "gate" based on their arrival time at the TIS gate. The TIS is used to isolate specific molecules for fragmentation based on their $m/z$ before entering a collision cell which contains the collision gas. The selected precursor ions are then fragmented in the collision cell before being accelerated by a second source into the second TOF analyser.
with the reflectron capability. Ions and fragmented ions are resolved according to their \( m/z \) before arriving at the detector. This technique has greatly enhanced the sensitivity and resolution of tandem mass spectrometric data which leads to exceptional determination of glycan compositions and sequences.

### 1.4.4 ESI-QTOF mass spectrometry

The ESI-QTOF hybrid mass spectrometer combines the benefits of ESI and the Quadrupole-TOF (QTOF) mass analyser, thus becoming a great complement to MALDI-TOF/TOF tandem mass spectrometry. In this thesis, the ESI-QTOF technique has been most beneficial in sequencing of very low abundance glycans. Figure 1.16 shows a schematic diagram of an ESI-QTOF-MS/MS, which consists of a nano ESI source, three quadrupoles, a reflectron TOF analyser and an MCP detector.

![Figure 1.16. A schematic diagram of an ESI-QTOF-MS.](image)

The schematic shows ion paths during tandem mass spectrometric experiment. Samples can be introduced into the ion source directly after separation in LC or from a needle filled by manual injection using a syringe. Inset diagram (reproduced from www.waters.com) shows the mechanism of ion desolvation in the ESI source. Sample ions are represented by circles of different colours and sizes as described in Figure 1.15.
In the ESI source, samples dissolved in solvent emerging from a liquid chromatography (LC) column or syringe are introduced via a capillary at high voltage and at atmospheric pressure. The sample solution eventually emerges as tiny droplets at the needle tip possessing a strong positive or negative charge due to the strong electric field. When the charged liquid first exits the tip, it briefly forms a cone shape known as a Taylor cone before the droplets burst away into a fine spray (Cole, 2010). The droplets in the spray then pass through a curtain of heated inert gas, most often nitrogen, to remove the remaining solvent molecules. When the electric field on their surface is large enough, ion desorption from the droplet surface will occur before entering the QTOF mass analyser (De Hoffmann & Stroobant, 2007). One of the advantages of this ionisation method is that the molecules remain intact (soft ionisation) (Gross, 2004). The most notable benefit of ESI is the ability to produce multiply charged ions that is very helpful especially in characterising biomolecules with high masses. Most commonly, ESI is connected to an LC system which enables direct analysis of LC separated biomolecules in mass spectrometers. For example, the nano-liquid chromatography interface (nano-LC) on ESI is outstanding for glycosylation site specific analysis in addition to the detection of other protein modifications such as phosphorylation and alklylation (Thomsson et al., 2000; Dell & Morris, 2001). However, manual injection into the ESI needle via a syringe is also excellent for ESI especially when involving very low amounts of samples or when LC separation is not required. Ionisation of glycans and glycoconjugates in conventional ESI used to be relatively poor compared to peptides and proteins, until the introduction of nano ESI that produces smaller droplets with better ion signals (Wilm & Mann, 1996). Smaller droplets reduce the hydrophilicity of oligosaccharides leading to increases in surface activity rather than in volatility (Karas et al., 2000), resembling the effects of glycan derivatisation.

The QTOF mass analyser which was first proposed by H.R. Morris (Morris et al., 1996) is currently the commercially most successful hybrid system (Gross, 2004). The quadrupole analyser consists of four cylindrically shaped rod electrodes with the pairs of opposite rods being each held at the same potential and commonly assembled in three sets to form an instrument with ion activation capability. A direct current voltage and an oscillating radio-frequency are applied to each pair of opposite rods creating an electric field that acts as a mass filter. The quadrupole allows high-speed scanning at relatively high pressure that is perfect for the continuous beam of ions from the ESI source (Dell et al., 2008). The additional reflectron TOF mass analysis immediately after the quadrupole analysis improved the ion detection, transmission, resolution and mass accuracy of a quadrupole instrument (Morris et al., 1997). However, it was not feasible
to directly combine a TOF analyser, a pulsed instrument, with continuous ionisation from electrospray source. This was solved by the orthogonal arrangement of the quadrupole and TOF analysers (Morris et al., 1996), a design that was successful because of the incorporation of an ion modulator/pusher, originally devised by Dawson and Guilhaus (Dawson & Guilhaus, 1989), at the interface.

The first quadrupole analyser, shown in Figure 1.16, is a radio frequency (RF)-only quadrupole (Q0) where ions are collimated and transferred into the adjacent high vacuum region of the second quadrupole, which is known as the mass filter quadrupole (Q1). In the MS mode, Q1 transmits ions over a wide mass range, whilst in the MS/MS mode, ions are resolved according to their m/z. This is followed by the third quadrupole (Q2), which is an RF-only quadrupole filter housed inside a collision cell. In the MS mode, Q2 acts just by focusing the ion beam, similar to Q0. In the MS/MS mode however, Q2 transmits ions through the collision cell that contains the collision gas (nitrogen) for fragmentation to occur. Precursor and/or fragment ions then enter the ion modulator in the TOF analyser. A pushout pulse applied orthogonal to the ion beam direction then extracts the ions to travel through the TOF tube. Ions are separated according to their m/z value before being detected and recorded (Cole, 2010). The TOF analyser contains a reflectron similar to what has been discussed for MALDI-TOF/TOF-MS in the previous section.

1.4.5 GC-EI mass spectrometry

Another essential requirement for a complete characterisation of glycans is linkage analysis. Gas chromatography-electron ionisation-MS (GC-EI-MS) has been shown to be a robust tool for that purpose (Haslam et al., 1997; Haslam et al., 2006; Parry et al., 2006). Fragmentation data from this instrument, when used in concert with enzymatic or chemical digestion, provides information on individual monosaccharide linkages within a sample. Glycan linkage analysis will be discussed in Section 1.4.8.5. Here the principles of GC-MS are described.

GC is known for its capability to perform chromatographic separation of volatile molecules from complex samples prior to mass spectrometric analysis. Because EI results in complex fragment-ion patterns, a high resolution separation technique such as the GC is vital because it ensures that only a limited number of components enter the ion source at any given time even when complex mixtures are examined (Dwek et al., 1993). From the GC, the sample
of interest is vaporised into the EI source and collided with a beam of electrons produced by a filament (commonly rhenium and tungsten wire). When the electron beam is close enough, energy can be transferred to the sample molecules. When there is sufficient energy, an electron can be expelled from the sample molecule, thus generating molecular ions (De Hoffmann & Stroobant, 2007) (Figure 1.17). Excess energy usually can be converted into activation energy that causes single or multiple stages of fragmentation. Several types of mass analysers can be coupled to this instrument, however the single quadrupole is best suited due to its high scan speeds (Gross, 2004). This instrument usually utilises a photomultiplier detector.

Figure 1.17. Schematic view of the ion source based on electron ionisation and the single quadrupole mass analyser typically found in a GC-MS instrument.
Black circles represent neutral sample molecules. Circles of other colours represent charged sample molecular ions and fragments. Reproduced from (Wittmann, 2007).

1.4.6 MALDI-QIT-TOF mass spectrometry

MALDI-QIT-TOF-MS is a relatively new type of instrumentation that utilises the benefits of MALDI-MS in mass fingerprinting and sequencing of biomolecules and brings them to a new level. The instrument’s most prominent advantage is the capability it provides for performing multiple stage mass spectrometric analysis (MS^n). Figure 1.18 shows a schematic diagram of a
MALDI-QIT-TOF-MS, which consists of a MALDI source, quadrupole-ion trap (QIT) and TOF mass analysers and an MCP detector.

**Figure 1.18. Schematic diagram of a MALDI-QIT-TOF-MS.**
The schematic shows ion paths during multiple stage mass spectrometric (MS^n) experiment. Sample ions are represented by circles of different colours and sizes as described in Figure 1.15.

The primary concern of coupling a low pressure MALDI source to a QIT was to be able to efficiently trap the axially injected gas-phase ion inside the analyser. The QIT (also known as Paul trap) mass analyser consists of two end cap hyperbolic electrodes and a ring electrode. The trap is formed when voltages are applied to these electrodes forming a 3D quadrupole. The use of dampening gas (usually helium) improves the trapping ability by cooling the kinetic energy of sample ions to the centre of the trap (Cole, 2010). The mass analysing principle of the QIT is based on creating stable trajectories for ions of a certain m/z or m/z range while removing unwanted ions from the trap due to their unstable trajectories (Gross, 2004). The ability to trap ions means that a sufficient number of ions can be stored for subsequent MS^n analyses with high detection sensitivity (Dell *et al.*, 2008). Other than MALDI, the QIT analyser can also be coupled to ESI and GC. The reflectron TOF analysis after the QIT analysis, similar to what has been described for MALDI-TOF/TOF-MS and ESI-QTOF-MS, enhances ion resolution by efficiently separating the precursor and/or product ions in MS or tandem MS experiments. Koy *et al.* have utilised MALD-QIT-TOF to determine the amino acid sequences of previously unassigned peptides of human haptoglobin and highlighted the advantages of this instrument including easy parent ion creation and minimal sample consumption (Koy *et al.*, 2003).
1.4.7 Other prominent mass spectrometry instrumentation

Prior to the wide usage of modern instruments such as MALDI and ESI, fast atom bombardment (FAB)-MS was at the forefront of oligosaccharide structural analysis (see sections 1.4.8.3 and 1.4.8.4). Advancement in technology has led to the invention of even more sensitive and robust instrumentations, and two of the techniques that are starting to receive much attention within this field are the electron capture dissociation (ECD) and the electron transfer dissociation (ETD), two powerful multiple stage ion activation methods. FAB, ECD and ETD instruments are discussed briefly here.

The introduction of FAB ionisation in the early 1980’s (Barber, 1981) revolutionised the structural analysis of a wide range of carbohydrate-containing biopolymers (Dell & Ballou, 1983; Dell, 1987). In FAB-MS, a high primary current beam of neutral atoms or ions accelerated from a gun is focused on a droplet of viscous matrix, for example thioglycerol or m-nitrobenzyl alcohol, placed on a small metal target. The sample of interest is dissolved in the matrix droplet prior to bombardment with the atom or ion beam (Barber, 1981). When the beam collides with the droplet, the surface mono-layer is destroyed by the impact. Excess energy is transmitted to the underlying molecules (matrix plus sample) some of which become ionised and are ejected out of the solution into the high vacuum of the ion source. They are then accelerated towards the analyser by a potential difference. In the positive mode, ionisation occurs by protonation or the addition of a cation such as sodium, whilst in the negative mode ions are primarily formed by loss of a proton. Fragmentation of labile bonds is common during the ionisation process, thus allowing both compositional and sequence data acquisition in a single experiment (Dell & Morris, 2001). The most prominent type of mass analyser coupled to a FAB-MS instrument is the double-focusing magnetic sector. Other common analysers include a single-focusing magnetic sector in tandem with a quadrupole (Hogg, 1983). The introduction of high magnetic field technique to FAB-MS by H.R. Morris and colleagues was pivotal to oligosaccharide analysis because of the resulting extended mass range (Morris et al., 1981; Dell & Morris, 1982).

ECD is an ion activation technique that was discovered by Zubarev et al. in 1998 (Zubarev et al., 1998). ECD involves irradiation of multiply charged molecular ions with low energy electrons. Negatively charged electrons can be captured by the cationic molecular ions to create a radical species with one charge less than the original ion. This charged reduced radical then rapidly undergoes dissociation into product ions (Cole, 2010). To date, ECD can only be
implemented in Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers. FT-ICR-MS is a promising mass spectrometry instrument that possesses unprecedented levels of resolution and mass accuracy (Gross, 2004; Zhao et al., 2008). The development of FT-ICR was inspired by ICR spectroscopy and FT-nuclear magnetic resonance and can be traced back to 1973 (Comisarow & Marshall, 1996). MALDI and ESI are both compatible as ionisation sources for FT-ICR-MS. In a cyclotron (also known as a Penning trap), ions are ‘trapped’ on a circular trajectory in a high magnetic field. Ions of different \( m/z \) are separated by the cyclotron frequency. The trapped ions are then excited to a large and coherent cyclotron orbit by an oscillating electric field perpendicular to the magnetic field. An image current generated by the ions as they drop back to their natural orbit can be measured and a Fourier transform is applied to give a mass spectrum (Solouki et al., 1998).

Although ECD is very useful, one of its drawbacks is the difficulty of implementing this technique on instruments other that FT-ICR-MS, which is expensive and requires highly skilled users. D.F. Hunt and colleagues have come up with a solution to this issue by inventing the technique known as ETD (Syka et al., 2004). ETD uses an anion electron carrier rather than a free electron to trigger electron capture by multiply charged cationic molecular ions (Cole, 2010). One of the most commonly used anion electron carriers is anthracene, a singly charged anion of a polycyclic aromatic hydrocarbon (De Hoffmann & Stroobant, 2007). After the electron transfer, similar to ECD, charge state reduction occurs to produce a radical positive sample ion that can then dissociate. ETD can be implemented in instruments including the quadrupole linear ion trap and QTOF (Syka et al., 2004; Xia et al., 2006).

The latest additions to the mass spectrometry family are mostly hybrid instruments. They are constructed by the combination of different types of mass analysers with different advantageous properties in a single machine (Gross, 2004). Examples of hybrid instruments, other than ESI-QTOF-MS and MALDI-QIT-TOF-MS, are quadrupole-FT-ICR-MS (Patrie et al., 2004), linear quadrupole ion trap-FT-ICR-MS (Syka et al., 2004) and magnetic sector-TOF (Strobel et al., 1991).

### 1.4.8 Strategies for glycan structural analysis

Modern analytical techniques such as mass spectrometry, nuclear magnetic resonance, X-ray crystallography and lectin/glycan array have revolutionised structural studies and are among
the popular methods for structural and functional glycan analysis (Dell & Morris, 2001; Drickamer & Taylor, 2002; Wormald et al., 2002). Mass spectrometry is increasingly becoming the method of choice for rigorous structural characterisation of biomolecules since the introduction of desorption based ionisation methods back in the late 1970’s (De Hoffmann & Stroobant, 2007).

Mass spectrometric analysis of glycans can be traced back to about 4 decades ago with the analyses of monosaccharides and hydrolysed small oligosaccharides using GC-MS, for instance the work of the Carlson and Albersheim groups (Sweet et al., 1974; Weber & Carlson, 1982). The analysis of oligosaccharides without hydrolysis, on the other hand, started later with the analysis using FAB instruments pioneered by the Morris and Dell group (Dell, 1987; Dell, 1990; Dell et al., 1993). Since then, tremendous developments and achievements have been made on both analytical methods and instrumentation. Currently, other than MALDI-TOF/TOF and ESI-QTOF that are the main glycan profiling and sequencing instruments applied in this thesis, LC-ESI-MS is another extensively used technique (Karlsson et al., 2004; Wada et al., 2007). The complete structural characterisation of oligosaccharides is more challenging than that of proteins or oligonucleotides due to the existence of additional specific characteristics such as isomeric states, linkage positions and branching capabilities. The determination of all this structural information for a comprehensive characterisation can be obtained by mass spectrometry, although sometimes it may require more than a single mass spectrometric technique. Various sample treatments prior to analysis might also be crucial for a successful experiment. A few important strategies for glycan mass spectrometric analysis are discussed below.

1.4.8.1 Release of oligosaccharides from glycoproteins

In order to perform an efficient glycomic analysis, glycans are released from their glycoconjugates. As mentioned earlier, glycoproteins contain two types of glycans, namely N-glycans and O-glycans. All N-glycan structures can be digested from proteins by peptide:N-glycanase F (PNGase F) except for α1,3 core fucosylated N-glycans commonly found in plants and invertebrates which can be removed by peptide:N-glycanase A (PNGase A) (Tretter et al., 1991). However, recently it has been suggested that the two newly cloned FuT10 and FuT11 can catalyse α1,3-fucosylation to N-glycan core in humans (Mollicone et al., 2009). In contrast to N-glycans, chemical release of O-glycans, known as reductive elimination, is the most practical and preferred method. A number of enzymes that possess the capability to digest O-glycans have been
reported previously, however so far have they been shown only to release simple sialylated and fucosylated O-glycan structures (Umemoto et al., 1978; Abdullah et al., 1992). Reductive elimination, also known as alkaline β-elimination, was first introduced in 1966 in an analysis of oligosaccharides from porcine submaxillary mucins (Carlson, 1966). Currently, this is the most common method used to release O-glycans from proteins. In this technique, unwanted degradation or ‘peeling’ of the oligosaccharides is avoided by reducing the monosaccharide of the reducing terminal to an alditol using reducing agents such as potassium borohydride (KBH₄) in an alkaline solutions such as potassium hydroxide (KOH) (Carlson, 1966). During the chemical degradation hydroxide from the alkali attacks the hydrogen on the backbone of the glycan-linked amino acid leading to the elimination of the O-linked sugar (Figure 1.19). The reducing terminal glycan (O-GalNAc) is then reduced, thus breaking the ring structure. Reductive elimination is permissible with or without prior digestion of N-glycans, but the latter usually results in some contamination with reduced and non-reduced N-glycans especially when high mannose N-glycans are present.

![Figure 1.19. Reaction steps for reductive elimination of O-linked GalNAc.](image)

The reactions lead to the release and reduction of the GalNAc residue and usually the partial destruction of the peptide chain. Reproduced from (Morelle & Michalski, 2007). KOH, potassium hydroxide; KBH₄, potassium borohydride.
1.4.8.2 Chemical and enzymatic digestions

Chemical and enzymatic digestions are an essential adjunct to glycan screening and sequencing because they can provide information on structural features such as stereochemistry and linkage (Dell et al., 2007). Mass spectrometry is an ideal method to monitor the products of these reactions because of its sensitivity. Enzymatic digestions are normally stereochemistry and/or linkage specific but must be carried out on native glycans. For example, β1,4-galactosidase digestion is used to determine the type of terminal Gal population in the sample of interest. On the other hand, chemical digestions are used to degrade sugar residues with broader specificities and can be carried out on either native or permethylated glycans. For example, trifluoroacetic acid (TFA) degradation at room temperature is used to remove terminal fucose residues. By taking advantage of the fact that certain fucose linkages are more labile, time course degradation is possible. Thus, 3- and 4-fucose residues are more likely to be released first than 2- and 6-fucose, allowing easy confirmation of fucose and rough determination of their linkages.

1.4.8.3 Permethylation

As oligosaccharides do not ionise as efficiently as peptides, derivatisation is commonly performed to improve the analysis by most of mass spectrometric techniques including MALDI and ESI (Dell et al., 2008). Derivatisation reduces the hydrophilicity of oligosaccharides so that they are easier to ionise. The types of derivatisation technique utilised in the analysis of oligosaccharides include fluorescence labelling of the reducing terminal via reductive amination (Anumula, 2006) and stable isotope labelling for instance with deuterium (Price, 2006), peracetylation and permethylation (Dell, 1990). Permethylation has become the preferred oligosaccharide derivatisation method because it results in a smaller mass increase and a greater volatility.

In the methylation of carbohydrates, successive base-catalysed ionisation of the hydroxyl groups followed by reaction with the methylating agent occurs, a method that was introduced by Sen-itiroh Hakomori (Hakomori, 1964). The Hakomori permethylation technique uses methylsulfinyl carbanion in dimethyl sulfoxide (DMSO) with methyl iodide being the methyl donor. However, permethylation nowadays is commonly catalysed by less strong bases such as solid sodium hydroxide with rapid derivatisation without the formation of non-sugar products, as initially demonstrated by I. Ciucanu and F. Kerek (Ciucanu & Kerek, 1984). Increased hydrophobicity allows glycans to be extracted from the methylation slurry using chloroform. The
permethylation technique was matured during the era of FAB-MS (Section 1.4.7) and has positioned itself as an essential technique in most oligosaccharide analytical strategies, and what was learnt is now being applied using current instrumentation (Dell, 1987; Dell et al., 1994; Wada et al., 2010).

Sulphated glycans, however, are not compatible with the standard NaOH-permethylation technique because of the low efficiency of chloroform extraction of negatively charged glycans (Zaia, 2004). Instead, a method based on the short Hakomori permethylation technique can be used for this analysis (Dell et al., 1994). Alternatively, the NaOH permethylation procedure can be rendered compatible with sulphated glycans as has been recently demonstrated by Khoo’s group (Yu et al., 2009) (details in Chapter 2). To improve the recovery and stability of sulphated glycans in the latter, the permethylation reaction is carried out in small scale at 4°C but for a longer period. Conventional partitioning against chloroform is replaced by direct reverse phase extraction and followed by a microscale fractionation using an amine column to separate negatively charged from neutral glycans, which gives an advantage to the isolation of sulphated glycans as they are the only remaining charged oligosaccharides.

Permethylated oligosaccharides are most often analysed in the positive ion mode as [M+Na]^+ molecular ions producing reducing and non-reducing terminal product ions with preferential fragmentation to the reducing side of (N-acetylhexosamine) HexNAc residues (Dell, 1987) (see next section). Compared to non-derivatised glycans, permethyl derivatives are better ionised and are capable of forming abundant fragment ions thus increasing the sensitivity of mass spectrometric analysis (Dell & Morris, 2001). In fact, permethylated glycans fragment very selectively and provide structurally diagnostic fragments, hence aiding data interpretation (Dell, 1987; Wada et al., 2007). Ions arising from single and multiple cleavages also can be easily discriminated (Dell et al., 2008). Furthermore, permethylation is a prerequisite for glycan linkage analysis (1.4.8.5). Other than that, permethylation also stabilises sialic acids, thus avoiding errors in glycan profiling. Sialic acid is large family of molecules with various types of naturally occurring derivatives, bearing of N-acetyl and N-glycolyl, and less frequently O-acetyl, O-lactyl, O-methyl and O-sulphate substituents. These added groups are chemically labile and very challenging to detect (Varki, 1992; Schauer, 2000).
1.4.8.4 Glycan fragmentation

In tandem mass spectrometry, sample molecular ions are activated and fragmented in order to provide detailed structural information. The MS$^2$ experiment is the most common type of analysis and can be done in most modern mass spectrometric instruments. However, multiple stage mass spectrometric analysis (MS$^n$), especially MS$^3$, is gaining recognition as a useful strategy for glycomic and glycoproteomic analyses due to the increasing demand for detail structural information. Furthermore, MS$^n$ data could indicate from which fragmentation pathway the MS$^2$ fragment originated, which is particularly useful for analysis on complex glycan structures (Harvey, 2010). The fragmentation pathways of glycans in tandem mass spectrometric analysis are overviewed below. It is important to highlight that most of the current knowledge on glycan fragmentations is derived from the analysis of oligosaccharides using FAB-MS instruments during the 1980’s, for instance the work of Dell & Ballou (Dell & Ballou, 1983).

In contrast to native samples, permethylated glycans fragment in a very predictable manner in both positive and negative ion modes. Generally, the positive ion mode is the method of choice due to the better fragmentations most commonly observed with singly charged sodiated ([M+Na]$^+$) and protonated ([M+H]$^+$) molecular ions in MALDI, or as multiply charged ions in ESI. In the negative ion mode, on the other hand, the molecular ion is deprotonated and therefore this mode of ionisation is best suited to molecules that have an intrinsic negative charge, for example sulphated or sialylated glycans, provided the latter are not permethylated. The negative mode is also well suited to the analysis of underivatised glycans provided the resulting molecular ions are stable during analysis. This is often not the case in MALDI where sulphates and sialic acids are readily lost. On the other hand ESI is well suited to handling fragile molecular ions. For instance, Karlsson et al. have very successfully performed rigorous characterisation of neutral and underivatised O-glycans reductively eliminated from human MUC5B using LC-ESI-MS$^n$ (Karlsson et al., 2004).

In MALDI and ESI instruments, various types of cleavages can be observed in CID analyses of permethylated glycans (exemplified in the fragmentation of a sialylated core 2 O-glycan in Figure 1.20). The most common type of cleavage is the glycosidic cleavage with hydrogen transfer, or also known as β-cleavage, with charge residing on either end (Figure 1.20). This type of cleavage can occur in both positive and negative modes. An important cleavage normally accompanying glycosidic fragmentation with hydrogen transfer is β-elimination of the
substituent from position 3 of a HexNAc residue (Dell, 1987; Dell et al., 2008) (Figure 1.20). These glycosidic and β-eliminated fragment ions can provide information on sites of sialylation, fucosylation and branching as well as the number of LacNAc repeats. The other type of fragmentation commonly observed in oligosaccharide analysis is ring cleavage (cross-ring fragmentation) (Figure 1.20). Cross-ring fragmentation is prominent in multiple stage MS (MS$^3$ and higher) but it can also occur in MS$^2$ experiments provided it is carried out in a high-energy collisional activation. Sodiated molecular ions produce better ring-cleavages relative to their protonated counterparts (Dell & Ballou, 1983; Dell, 1987; Zaia, 2004). In a cross-ring fragmentation, the charge can remain on either the reducing or the non-reducing end, depending on the nature of the sample and the ionisation mode. This type of fragmentation is most helpful for assigning linkages.

In other mass spectrometric techniques, especially FAB instruments, permethylated derivatives most commonly undergo A-type cleavages on the non-reducing side of a glycosidic bond to give oxonium-type fragment ions, which may carry sodium if the parent ion is sodiated. This A-type cleavage is often favoured at HexNAc residues and can only occur in the positive ion mode. Secondary cleavage β-elimination of the substituent from the position 3 of a HexNAc residue is often observed as well (Dell, 1987; Dell et al., 2007). It is important to note that the cleavage of sialic acid residues often leads to the formation of oxonium ions even in MALDI and ESI instruments (Figure 1.20).
By using the O-glycan structure NeuAcα2,3Galβ1,3(Galβ1,4GlcNAcβ1,6)GalNAcitol as a model, various cleavages can be observed following the fragmentation in MALDI or ESI CID. For example, glycosidic cleavages (red lines) with charge retained on the reducing end (fragments Y₁ & Y₂) or non-reducing end (fragments B₁ & B₂). β-elimination of a substituent from the position 3 of a HexNAc residue (green lines) is demonstrated by fragments C₂ and Z₁. Cleavage of sialic acids (purple lines) often results in the formation of oxonium ions (fragment B₁). Examples of ring cleavages (blue lines) often seen in O-glycan fragmentations are 1.5X₁, 0.4X and 0.3A. Based on the nomenclature proposed by (Domon & Costello, 1988).

1.4.8.5 Glycan linkage analysis

Tandem mass spectrometry using MALDI-MS or ESI-MS can provide specific but limited linkage information based on the fragments acquired. To acquire overall glycan linkage data from a particular sample, linkage analysis using GC-MS is a better method. In this analysis, monosaccharides from the glycans of a particular sample are separated according to their mass and the number and position of their former linkages. This technique was first introduced by Albersheim et al. while analysing polysaccharides from plant cell-wall (Albersheim et al., 1967). GC-EI-MS linkage analysis accompanied by chemical or enzymatic digestions has been shown to be a reliable method to aid the overall structural assignment by MS and MS/MS analyses by identifying monosaccharide constituents and glycosidic bond positions (Haslam et al., 1996; Kui Wong et al., 2003). Figure 1.21 illustrates the preparation of permethylated glycans for linkage analysis.
Prior to the analysis, glycans are permethylated in order to permanently tag the hydroxyl groups that are not involved in any linkage with other sugar residues or in ring formation. The permethylated glycans are first subjected to acid hydrolysis releasing the partially methylated monosaccharides. The monosaccharides are then reduced (by breaking the ring structures) producing alditols with borodeuteride and the C-1 that is formerly involved in a linkage is labelled with deuterium (Manzi & van Halbeek, 1999). Free hydroxyl groups are then O-acetylated to label the positions of the former linkages, thus the residues are now partially methylated alditol acetates (PMAA). PMAAs can be identified by comparison of their retention times on the GC column and the EI-MS spectra with those of known standards (Albersheim et al., 1967; Manzi & van Halbeek, 1999).
1.5 Thesis aims

The general aims of this thesis were to optimise current O-glycomic methodologies and to perform more in-depth structural analysis on O-glycoconjugates, in order to help deliver a new level of understanding of the regulation and biological roles of O-glycans, as well as glycosyltransferases involved in their biosynthesis. This thesis encompasses glycomic mass spectrometric analyses of O-glycans derived from glycoproteins of mice and cell cultures from three separate but related projects. The projects are:

1. High sensitivity O-glycomic analysis of wild type and C2GnT knockout murine tissues. This project aimed at characterising the changes in O-glycan structures resulting from partial and complete absence of C2GnT, and to associate the observed structural aberrations with murine pathology.

2. O-glycosylation profiling of gastric mucosa from FuT2-null mice in relation to Helicobacter pylori adhesion. In this study, we performed detailed glycoepitope sequencing of murine gastric tissues in order to better understand the functional importance of fucosylated glycoconjugates in the pathogenicity of H. pylori.

3. Investigation of cellular regulation of protein O-glycosylation by Src. High sensitivity O-glycomic methodology that was optimised in the above projects was applied in this analysis to investigate the significance of Src activation and/or ppGalNAcT relocalisation in modulating the biosynthesis of O-glycans in a number of cell lines.
Chapter 2:

Materials and Methods
2 Materials and methods

2.1 Materials

2.1.1 Biological samples

2.1.1.1 High sensitivity O-glycomic analysis of wild type and C2GnT knockout murine tissues

Murine tissues including stomach, colon, small intestine, thyroid/trachea, thymus and kidneys from 129/SvJ wild type, C2GnT2 knockout (KO), C2GnT3 KO and C2GnT triple KO mice were supplied by our collaborators, Jamey D. Marth and colleagues from the University of California San Diego, USA. Details on the preparation of the murine tissues have been published (Stone et al., 2009). Mice singly deficient for C2GnT1, C2GnT2 and C2GnT3 were cross-bred to produce C2GnT triple KO mice. As illustrated in Figure 2.1, C2GnT1 KO mice were crossed to C2GnT3 KO mice and the progenies were bred to each other to generate mice doubly deficient for both C2GnT1 and C2GnT3 (T1/T3). T1/T3 mice were then bred to C2GnT2 KO mice and the progenies were bred together to produce mice deficient for all three genes (T1/T2/T3).

2.1.1.2 O-glycosylation profiling of gastric mucosa from FuT2-null mice in relation to Helicobacter pylori adhesion

Murine stomach and gastric mucosa scrapings from C57BL/6 wild type and FuT2-null mice were supplied by Celso A. Reis and colleagues from the University of Porto, Portugal. Details on the preparation of these samples have been published (Magalhaes et al., 2009). In addition, Le\(^a\) and Muc5AC immunoprecipitated glycoproteins from murine stomach lysates of wild type and FuT2-null mice were also supplied. Immunoprecipitated samples were prepared by isolating Le\(^a\) carrier glycoproteins and Muc5AC from stomach lysates separately using Sepharose beads conjugated with antibodies SPM279 and 45M1, respectively. Then the purified samples were confirmed by running on electrophoresis gel followed by Western blotting using similar antibodies. Selected bands for mass spectrometric analysis are shown in Figure 2.2.
2.1.1.3 Investigation of cellular regulation of protein O-glycosylation by Src

NIH3T3 Src mouse fibroblastic and NBT-II rat epithelial cell lines were supplied by our collaborators Frederic Bard and colleagues from the National University of Singapore. The first cell line was induced to have different levels of Src by transforming cells with low Src level (3T3 WT) with viral Src (3T3 vSrc). The latter cell line was prepared with different localisation of exogenous human ppGalNAcT-2 in the ER (H2ER and WT ER) or the Golgi (WT Golgi). Wild type NBT-II cells were also supplied. Each sample consisted of around 35-50 million cells.
### 2.1.2 General chemicals and reagents

The following materials were obtained from the sources indicated: Methanol, chloroform, propan-1-ol, ammonia solution, acetic acid, dimethylsulfoxide (DMSO), trifluoroacetic acid (TFA), and acetonitrile were purchased from Romil (Cambridge, UK). α-cyano-4-hydroxycinnamic acid (CHCA), ethylenediaminetetraacetic acid (EDTA), potassium borohydride (KBH₄), iodoacetic acid (IAA), hexane, 48% (v/v) hydrofluoric acid in water (HF), Tris(hydroxymethyl)aminomethane (Tris), ammonium hydrogen carbonate (Ambic), formic acid, potassium hydroxide, Dowex 1-X8, 2,5-dihydroxybenzoic acid (DHB), sodium hydroxide (NaOH), sodium tetradeuteroborate (NaBD₄), potassium hydroxide (KOH) and phosphate buffered saline tablets were from Sigma-Aldrich Company Ltd (Dorset, UK). Methyl iodide and acetic anhydride were from Alfa Aeser (Lancashire, UK). 8M guanidine-HCl (Gu-HCl) was from Pierce (Northumberland, UK). Dithiothreitol (DTT) and 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) were from Roche (West Sussex, UK). Sodium chloride (NaCl) was from Rose Chemicals (London, UK). 3,4-diaminobenzophenone (DABP) and Macherey-Nagel Nucleosil NH₂ polar phase was acquired from Fisher Scientific (Loughborough, UK). Tubes, bottles and beakers were washed thoroughly.

![Figure 2.2. Western blotting of Leα and Muc5AC immunoprecipitated gastric mucosa from wild type and FuT2-null mice after gel electrophoresis.](image)

The blots were probed using antibody against (A) Muc5AC (45M1) and (B) Leα (SPM279). The yellow boxes indicate the bands of interest. The bands in the red boxes correspond to the antibodies (IgG) that were conjugated with the beads for the IP procedure. Data from Ana Magalhaes, University of Porto, Portugal. WB, Western blotting; IP, immunoprecipitation; wt, wild type.
with ultra high quality water which was generated from Purite Neptune water purification system from Purite Ltd. (Oxfordshire, UK) to avoid contamination.

### 2.1.3 Enzymes, antibodies and proteins/peptides

TPCK treated bovine pancreas trypsin (EC 3.4.21.4), albumin from bovine serum (BSA), IgA kappa from murine myeloma and goat anti mouse IgM peroxidase conjugate was purchased from Sigma-Aldrich Company Ltd. (Dorset, UK). Peptide: N-glycosidase F (PNGase F) containing 50% glycerol (Flavobacterium meningosepticum –recombinant from Escherichia coli, EC 3.5.1.52; 25000 U/mg) was purchased from Roche (West Sussex, UK). Sialidase A (from Escherichia coli, EC 3.2.1.18; > 5 U/ml), α1,3/4/6-galactosidase (from green coffee bean, EC 3.2.1.22; 10 U), α1,2-fucosidase (Xanthomonas sp, 3.2.1.51; activity > 100 mU/ml) and β-N-acetylhexosaminidase (HEXase I) (from E. coli, EC 3.2.1.52; 1.6 U) were purchased from Prozyme (Ely, UK). β1,4-galactosidase (Steptococcus pneumoniae, EC 3.2.1.23; specific activity > 600 pmol/min/µg) was from R&D Systems (Abingdon, UK). The 4700 mass standards kit which contains a peptide mixture of des-Arg 1-bradykinin (FW 904.0), angiotensin I (FW 1296.5), [glu1]-Fibrinopeptide B (FW 1570.68) and adrenocorticotropic hormone 1-17 (FW 2093.4) was obtained from Applied Biosystems (Warrington, UK).

### 2.1.4 Labware and Equipment

Homogenisation and sonication of tissues were performed by using a CAT X120 homogenizer from CAT Ingenieurburo (Denmark) and Vibra-Cell VCX130 ultrasonic processor from Sonic & Materials Inc (Switzerland). Samples were concentrated by using a Savant SPD SpeedVac concentrator and lyophilized by using a Savant ModulyoD freeze dryer, both were from Thermo Fisher Scientific Inc (Basingstoke, UK). Sep Pak C18 cartridges and Oasis HLB extraction cartridges were from Waters Corporation (Hertfordshire, UK). SnakeSkin dialysis tubing (7kDa molecular weight cut off) was from Pierce (Northumberland, UK). Nitrogen and argon gases were supplied by BOC (Guilford, UK). RTX-5MS GC column was purchased from Thames Restek UK Ltd (Saunderton, UK). All aqueous solutions were made up using ultra high quality water which was generated from Purite Neptune water purification.
2.2 Methods

2.2.1 Isolation of glycopeptides/peptides

2.2.1.1 Tissue homogenisation (without glycolipid removal)

This procedure has been described previously (Sutton-Smith & Dell, 2006). Tissues were first weighed and then were disrupted on ice by using a homogenizer in the presence of 5 ml homogenisation buffer (25 mM Tris, 150 mM NaCl, 5 mM EDTA and 1% CHAPS at pH 7.4) (for tissues between 0.1 g to 3 g). Lysates were then transferred into dialysis tubing (pre-soaked in water) dialysed against 50 mM ambic buffer pH 8.5 at 4°C for 48 hours with several buffer changes. After dialysis, lysates were transferred into 15 ml tubes and lyophilised. Then, samples were ready for reduction and carboxymethylation.

2.2.1.2 Tissue homogenisation (with glycolipid removal)

This method has been described previously (Parry et al., 2007). Tissues were first weighed. With assumption that 80% of the tissue weight is water, at least 4 volumes of ice cold water were added to the sample which was then homogenised. After the samples were completely homogenised, based on the total water volume in sample, 2.67 volumes of methanol were added and then the samples were mixed vigorously. Next, 1.33 volumes of chloroform were added, followed by vigorous mixing. The samples were then centrifuged (3000 rpm, 10 min). The supernatants which contain glycolipids were carefully removed. Protein pellets were blown for a few minutes under a nitrogen stream to remove excess methanol and chloroform without completely drying the samples. 50 µl of 0.6 M Tris buffer was added to each sample and then removed under a nitrogen stream without completely drying the samples. Protein pellets were then either reduced and carboxymethylated immediately or kept at -20°C.

2.2.1.3 Cell sonication

A sufficient amount (1- 2 ml) of ice-cold cell lysis buffer (25 mM Tris, 150 mM NaCl, 5 mM EDTA and 1% (w/v) CHAPS, pH 7.4) was added to cell preparations to ensure complete suspension before transferring into 15 ml tubes. The suspensions were sonicated on ice in continuous mode at 40 Amps for 10 seconds. The sonication was repeated for 4 or 5 times with a pause for about 15 seconds between each repeat. Then lysates were transferred to dialysis tubing.
and dialysed against 50 mM ambic buffer pH 8.5 at 4°C for 48 hours with several buffer changes. Samples were finally transferred into 15 or 50 ml tubes and lyophilised.

2.2.1.4 Protein reduction and carboxymethylation

Proteins were reduced and carboxymethylated as previously described (Parry et al., 2007). Lyophilised tissue lysate or protein pellet were reduced by incubation in 0.6 M Tris/4 M Guanidine-HCl buffer pH ~8.5 containing 2 mg/ml dithiothreitol at 50°C for 2 hours. This was followed by carboxymethylation in 0.6 M Tris buffer containing 12 mg/ml of iodoacetic acid incubated in the dark at room temperature for 2 hours. Carboxymethylation was then terminated by dialysis against 50 mM Ambic buffer pH 8.5, at 4°C for 48 hours with several buffer changes, followed by lyophilisation in 15 or 50 ml tubes.

2.2.1.5 Cleavage into peptides and glycopeptides

Protein digestion using TPCK treated bovine pancreas trypsin and purification was performed as previously described (Sutton-Smith & Dell, 2006). For each mg of tissue sample, 1.5 µg of trypsin from a 1 mg/ml trypsin solution in 50 mM Ambic buffer (pH 8.5) was added to the lysates. For cell samples, about 1 mg of trypsin was used for every 2-3 million cells. Incubation was performed at 37°C for 14 hours, followed by purification with Oasis HLB extraction cartridges.

The cartridges were first attached to 10 ml polypropylene/polyethylene syringes and each successively conditioned with 5 ml methanol, 5 ml of 5% (v/v) acetic acid, 5 ml propan-1-ol and 15 ml of 5% (v/v) acetic acid. The digested samples were then loaded onto the cartridges, washed with 20 ml of acetic acid (5% v/v) to remove hydrophilic contaminants, and then eluted sequentially with 5 ml of each 20% and 40% propan-1-ol solution in 5% (v/v) acetic acid. The propan-1-ol elutions, which contain glycopeptides and peptides, were concentrated with the Savant SpeedVac, combined and subsequently lyophilised overnight.

2.2.2 Digestion of N-glycans from glycopeptides

The lyophilised glycopeptide and peptide fractions were dissolved in 200-300 µl of 50 mM ambic buffer (pH 8.4) before adding a total of 3 unit of PNGase F and incubating at 37°C for
24 hours. Another 2-3 unit of PNGase F were added after a few hours of incubation. The samples were then lyophilised and re-dissolved in 200 µl of 5% (v/v) acetic acid for solid phase extraction with Sep-Pak C18 cartridges as described previously (Sutton-Smith & Dell, 2006).

The cartridges were first attached to 10 ml glass syringes and each was successively conditioned with methanol (5 ml), 5% (v/v) acetic acid (5 ml), propan-1-ol (5 ml) and 5% (v/v) acetic acid (15 ml). The samples were then eluted sequentially with 5 ml of 5% (v/v) acetic acid, 4 ml of each 20% and 40% propan-1-ol solution in 5% (v/v) acetic acid. All elutions were concentrated with Savant SpeedVac. The acetic acid fractions, which contain the released N-glycans, were lyophilised and later permethylated. The propan-1-ol eluents, which contain O-glycopeptides, were combined prior to lyophilisation and subsequent release of O-glycans.

2.2.3 Release of O-glycans from glycopeptides

Reductive elimination of O-glycans was performed as described previously (Sutton-Smith & Dell, 2006). 400 µl 0.1 M KOH containing KBH₄ (54 mg/ml) was added to the dried glycopeptide sample and incubated at 45°C for 14-16 hours. The reaction was terminated by adding a few drops of 5% (v/v) acetic acid followed by purification with a Dowex 1-X8 desalting column.

To prepare the desalting column, a 150 mm Pasteur pipette was fitted with a small piece of glass wool inside it and a piece of silicone tubing at its tapered end and then packed with Dowex beads for each of the samples. The columns were first washed with 15 ml of 5% (v/v) acetic acid. Next, the samples were loaded carefully on top of the column and then eluted with 7 ml of 5% (v/v) acetic acid. The volume of the eluents was reduced with a Savant Speedvac followed by lyophilisation overnight. Excess borates in the samples were removed by co-evaporating with 10% (v/v) acetic acid in methanol (4 x 0.5 ml) under a stream of nitrogen at room temperature. The purified native O-glycans were subsequently permethylated.

2.2.4 Permethylation and purification for non-sulphated glycan analysis

Sodium hydroxide (NaOH) permethylation was performed according to an established procedure as described previously (Dell, 1990). Samples were ensured to be completely dried before permethylation. About 7-8 NaOH pellets were ground to a fine powder in a dried mortar
and mixed with 5 ml of anhydrous DMSO. Then, about 1 ml of the resulting slurry was immediately added to each sample, followed by the addition of about 0.7 ml of methyl iodide. The samples were then mixed rigorously on an automatic shaker for 15 min at room temperature. The reaction was quenched by drop wise addition of water, while constantly shaking the sample tube to lessen the exothermic effect. Permethylated glycans were extracted by adding about 1 ml of chloroform to the sample. This was followed by another rigorous shaking before removing the water layer that contains hydrophilic contaminants. The chloroform layer was later washed several times with water by rigorous shaking. Finally, the chloroform layer was dried down under a gentle stream of nitrogen.

Purification of permethylated samples were done by using Sep-Pak C18 cartridges. The cartridges were first attached to 10 ml glass syringes and each successively conditioned with methanol (5 ml), water (5 ml), acetonitrile (5 ml) and water (15 ml). Each sample was dissolved in 200 µl of methanol:water (1:1) solution before loading onto the cartridges. The cartridges were washed with 5 ml of water and then eluted sequentially with 3 ml of each 15%, 35% and 50% acetonitrile/water. All eluents were then concentrated with a Savant SpeedVac and subsequently lyophilised. Samples were then ready for mass spectrometric analysis or derivatisation for linkage analysis.

### 2.2.5 Permethylation and purification for sulphated O-glycan analysis

The methodology was based on the work of K.H. Khoo’s group (Yu et al., 2009). A dried glycan sample in a glass tube was redissolved in a slurry of finely ground NaOH pellets in about 0.2 ml DMSO, similar to that described for the non-sulphated glycans. About 0.1 ml of methyl iodide was added and the sample was gently vortexed for 3 hours at 4°C. The reaction was quenched by the drop wise addition of with 0.2 ml of cold water while placing the glass tube on ice, followed by careful pH neutralisation using 30% aqueous acetic acid by checking on a pH indicator paper. The sample was then purified by using a C18 Sep-Pak cartridge. The cartridge was conditioned as described above (Section 2.2.4). However, after sample application, the cartridge was washed to remove hydrophilic salts and contaminants by using 5 ml each of ultrapure water and 2.5% and 10% acetonitrile/water. This was followed by glycan elution with 5 ml each of 25% and 50% acetonitrile/water and the fractions were concentrated with the SpeedVac. The 25% fraction was then subjected to amine column fractionation.
2.2.6 Fractionation of sulphated O-glycans by an amine column

Amine beads (Nucleosil NH₂, 5 µm particle size) were packed into a 5 ml pipette tip with the tapered end plugged by a small piece of filter paper. The volume of the packed beads was around 5 µl. The column was first conditioned and washed sequentially with 5% acetonitrile/0.1% formic acid, 50% acetonitrile/0.1% formic acid and 95% acetonitrile/0.1% formic acid. The permethylated glycan sample was dissolved in 100% acetonitrile and then loaded to the column, and hydrophobic contaminants were washed off with 95% acetonitrile. Sulphated glycans were eluted with 2.5 mM and 10 mM ammonium acetate in 50% acetonitrile. Eluents were then concentrated by using a SpeedVac prior to mass spectrometric analysis.

2.2.7 Derivatisation of glycans for linkage analysis

Partially methylated alditol acetates for GC-EI-MS analysis were prepared from permethylated glycans according to procedures described previously (Sutton-Smith & Dell, 2006). Samples were first hydrolysed to partially methylated monosaccharides by incubation in 200 µL of 2 M TFA at 121°C for 2 hours. This was followed by a reducing process with 200 µL of 10 mg/ml NaBD₄ in 2 M ammonia solution at room temperature for 2 hours. Samples were then acetylated by incubation in 200 µl of acetic anhydride at 100°C for 1 hour and then dried under a very slow nitrogen stream. The resulting partially methylated alditol acetate monosaccharides were extracted with chloroform the same way as permethylated glycans. The samples were finally dried under a very slow nitrogen stream.

2.2.8 Enzymatic/chemical glycan degradation

2.2.8.1 α-galactosidase, sialidase A and α1,2-fucosidase digestions

Released O-glycan samples after the borate removal step were dissolved in a small amount of ultrapure water and then lyophilised. Each of the samples was dissolved in 100 µl of 50 mM ammonium acetate pH 5.4 (pH adjusted with acetic acid) and 55 µl of ultra pure water. 5 µl (25 mU) of α-galactosidase, 35 µl (175 mU) of sialidase A and 5 µl (0.5 mU) of α1,2-fucosidase were added to each sample. Incubation was done at 37°C overnight and followed by lyophilisation. Digested samples were then purified by using solid phase extraction with Sep-Pak C18 cartridges as described in Section 2.2.2. Only the acetic acid fraction was collected, lyophilised and permethylated.
2.2.8.2 β-N-acetylhexosaminidase/HEXase I digestion

Released O-glycan samples after the borate removal step were dissolved in a small amount of ultrapure water and then lyophilised. 10 µl of the supplied reaction buffer, 40 µl of ultrapure water and 10 µl (80 mU) of the enzyme were added to the each of the samples. Incubation was done at 37°C for 24 hours followed by lyophilisation. The samples were then purified by using solid phase extraction with Sep-Pak C18 cartridges as described in Section 2.2.2. Only the acetic acid fraction was collected, lyophilised and permethylated.

2.2.8.3 β1,4-galactosidase digestion

Released O-glycan samples after the borate removal step were dissolved in a small amount of ultrapure water and then lyophilised. 70 µl of 50 mM ammonium acetate pH6.0 and 40 µl (0.012 µmol) of the enzyme were added to each of the samples. Incubation was done at 37°C for 24 hours followed by lyophilisation. The samples were then purified by using solid phase extraction with Sep-Pak C18 cartridges as described in Section 2.2.2. Only the acetic acid fraction was collected, lyophilised and permethylated.

2.2.8.4 TFA defucosylation

1 ml of TFA was added to native or permethylated glycans and incubated at 20°C in water bath for 20 hours. The samples were then lyophilised, permethylated and purified as explained earlier.

2.2.9 Mass spectrometric analysis

2.2.9.1 Preparation of matrices

2,5-dihydroxybenzoic acid (DHB) solution was prepared by dissolving the powder in either 70:30 methanol: water at a concentration of 20 mg/ml (for all O-glycan analysis in the positive ion mode of MALDI-TOF/TOF) or 80:20 methanol/water at a concentration of 10 mg/ml (for all O-glycan analysis using MALDI-QIT-TOF). α-cyano-4-hydroxycinnamic acid (CHCA) solution was prepared by dissolving the powder in 50:40:10 methanol: water: TFA at a concentration of 10 mg/ml (for peptide calibration standards). 3,4-diaminobenzophenone (DABP) was prepared by dissolving the powder in 75% acetonitrile/0.1% TFA at a concentration
of 10 mg/ml (for sulphated O-glycan analysis in the negative ion mode). The matrices were prepared fresh every week.

2.2.9.2 MALDI-TOF/TOF mass spectrometry and tandem mass spectrometry

Permethylated glycan samples were resuspended in 10 µL of methanol. 2 µl of sample was then mixed with 2 µl of DHB matrix solution. The mixture was then spotted on a stainless steel MALDI target (with a total of 1 µl for each spot) and allowed to dry under vacuum. MALDI mass spectrometric (MS) data were acquired by using the 4800 MALDI-TOF/TOF analyser mass spectrometer (Applied Biosystems, USA) in positive ion and reflectron mode by using DHB as the matrix. The instrument was calibrated by using peptides from the 4700 mass standards kit with CHCA as the matrix. Tandem mass spectrometric (MS/MS) data were acquired by using the same instrument with the collision energy set at 1 kV. Argon was used as the collision gas with a pressure at 3.5 x 10^{-6} Torr.

2.2.9.3 ESI-QTOF tandem mass spectrometry

ESI-QTOF-MS/MS data were acquired by using the Qstar Pulsar Hybrid System mass spectrometer (Applied Biosystems/MDS Sciex, Toronto, Canada) in the positive ion mode and pre-calibrated with 10-100 fmol/µL of [Glu¹]-fibrinopeptide B in 5% (v/v) acetonitrile/acetic acid. A few µl of permethylated samples dissolved in methanol were injected manually via syringe into the instrument. The collision energy used was varied between 30 eV and 90 eV depending on the size and nature of the target glycan. The collision gas used was nitrogen with a pressure at 5.3 x 10^{-5} Torr.

2.2.9.4 GC-EI mass spectrometry

The partially methylated alditol acetate samples were reconstituted in a small amount of hexane. Data were acquired by using the Clarus 560 D GC mass spectrometer (Perkin Elmer, USA) fitted with an RTX-5MS column (30 m x 0.25 mm internal diameter) (Thames Restek, UK). The samples were injected into the column at 60°C and then the temperature was increased at 8°C/min to 300°C.
2.2.9.5 MALDI-QIT-TOF mass spectrometry and tandem mass spectrometry

Permethylated glycan samples were resuspended in 5 μL of 80% methanol/water. 1 μl of sample was then spotted on a stainless steel MALDI target and allowed to air-dry. Then, 1 μl of DHB matrix solution was deposited onto the same position and allowed to air-dry. Samples were analysed by using an Axima Resonance MALDI-QIT-TOF mass spectrometer (Shimadzu, UK) in ‘mid 850’ positive ion mode. The instrument was calibrated by using peptide standards in CHCA (5 mg/ml in 50/50 acetonitrile/0.1% TFA. Argon was used as the CID gas.
Chapter 3:

High sensitivity O-glycomic analysis of wild type and C2GnT knockout murine tissues
3 High sensitivity O-glycomic analysis of wild type and C2GnT knockout murine tissues

3.1 Background

Core 2 β1,6-N-acetylglucosaminyltransferase (C2GnT), which exists in 3 isoforms, C2GnT1, C2GnT2 and C2GnT3, is one of the key enzymes in the O-glycan biosynthetic pathway, as discussed in Section 1.2.2.3. These isoenzymes produce core 2 O-glycans and have been correlated with the biosynthesis of core 4 O-glycans and I-branches. Our collaborators, Jamey Marth and colleagues from the University of California San Diego have generated mice with single and multiple deficiencies of C2GnT isoenzyme(s) and have evaluated the biological consequences of the loss of core 2 function. We at Imperial College have performed comprehensive O-glycomic analyses of neutral and sialylated glycans expressed in the colon, small intestine, stomach, kidney, thyroid/trachea and thymus of wild type, C2GnT2 and C2GnT3 single knockouts and the C2GnT1-3 triple knockout mice. This work contributed to most of the optimisation of our O-glycomic methodologies and knowledge of murine O-glycan structures that were then utilised in other work described in this thesis.

3.2 Results

3.2.1 O-Glycomic strategies

The murine tissues analysed are listed in Table 3.1. Tissues were selected based on the expression profiles of C2GnTs supplemented by phenotypic data emerging throughout the study. The glycomic strategy employed was a refinement of earlier methodologies. A variety of extraction procedures were assessed with respect to O-glycan recovery and mass spectral quality. Two approaches were found to be best suited to the range of tissues examined (described in Chapter 2). Most tissues were homogenised using the technique without glycolipid removal (Section 2.2.1.1), but for stomach the homogenisation procedure which included glycolipid removal (Section 2.2.1.2) was used instead as that methodology appears to give the best O-glycan data for stomach tissues. Tissue lysates were then reduced/ carboxymethylated and digested with trypsin to facilitate subsequent release of N- and O-glycans. After N-glycan digestion by PNGase F, the O-glycans were released by reductive elimination. However, for stomach O-glycans the best data were acquired when O-glycans were analysed without prior N-glycan removal. Purified O-glycans were permethylated to increase the sensitivity of detection and to direct the subsequent mass spectrometric fragmentation. Both the 35% and 50% acetonitrile/water fractions from Sep-
Pak C18 purification were subjected to mass spectrometric analysis. Glycans were detected as singly charged sodiated molecular ions \([M+Na]^+\) in the positive ion mode. Consistent and reproducible data were obtained from at least two repeats of experiments. Most O-glycans were detected in the 35% fraction, but data acquisition from both fractions was combined in some instances to have better signals of the high-mass glycans.

Table 3.1. List of tissues analysed in this study derived from the wild type (WT), C2GnT2 KO (KO2), C2GnT3 KO (KO3) and C2GnT triple KO (TKO) mice.

<table>
<thead>
<tr>
<th>Tissues</th>
<th>WT</th>
<th>KO2</th>
<th>KO3</th>
<th>TKO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Colon (Sulphated glycans)</td>
<td>✔</td>
<td>✔</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Stomach</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Small intestine</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Thyroid/trachea*</td>
<td>✔</td>
<td>❌</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Kidneys</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Thymus</td>
<td>✔</td>
<td>❌</td>
<td>✔</td>
<td>❌</td>
</tr>
</tbody>
</table>

*, trachea tissues could not be completely separated from the thyroid gland; ✔, analysed; ❌, not analysed.

Additional O-sulphoglycomic analysis was also performed on colon tissues. Methods were based on the work of K.H. Khoo and colleagues as described in Chapter 2. In this analysis, the workflow was the same as described for the non-sulphated glycans up to the derivatisation stage, at which a modified permethylation technique was used instead, aimed to preserve the sulphate groups as well as to ensure the recovery of the sulphated glycans (see Section 2.2.5). The subsequent C18 Sep-Pak purification (Section 2.2.5) resulted in two fractions, the 25% acetonitrile/water fraction that contains most of the sulphated glycans and the 50% acetonitrile/water fraction that contains most of the non-sulphated glycans. The 25% fraction was then subjected to amine column fractionation (see Section 2.2.6) to separate neutral glycans (non-sulphated), which elute in the unbound fraction, from charged glycans (sulphated) in the 2.5 mM and 10 mM ammonium acetate/acetoniitrile elutions. The 50% acetonitrile fraction from the Sep-Pak purification and both the ammonium acetate/acetoniitrile fractions were analysed in both
positive and negative ion modes of MALDI-TOF/TOF-MS and MS/MS. In the negative ion mode, sulphated O-glycans were detected as singly charged deprotonated molecular ions \([\text{M-H}]^-\), whereas in the positive ion mode, they were detected as singly charged disodiated deprotonated molecular ions \([\text{M+2Na-H}]^+\).

### 3.2.2 Exhaustive tandem mass spectrometric analysis of permethylated products of reductive elimination differentiates glycan isoforms and terminal epitopes

Mass fingerprinting and sequencing were primarily carried out using a MALDI-TOF/TOF mass spectrometer. Additional analyses using ESI-QTOF-MS, MALDI-QIT-TOF-MS and GC-EI-MS were performed on selected samples. The MS/MS analysis enabled differentiation of the structures and sequences of glycans with the same \(m/z\) value and composition. Figure 3.1 shows an example of how MS/MS data were used to characterise glycan structures. The stomach of the wild type and C2GnT2 knockout mice shared a molecular ion at \(m/z\) 1331 in their mass fingerprint (Figure 3.1A and B) and the MS/MS spectra of each of these ions are reproduced in the figure. It is evident that in the wild type mice the ion at \(m/z\) 1331 corresponds to a core 2 O-glycan (see annotated cartoon in Figure 3.1C). However, in the C2GnT2 knockout mice, the fragmentation was clearly different (Figure 3.1D). In this case, fragment ions attributable to the core 2 O-glycan are relatively minor and all major fragment ions are assignable to the elongated core 1 O-glycan shown in the annotation. Definitive fragments for the core 1 O-glycan are \(m/z\) 298, 520 and 1056. This MS/MS strategy for unambiguously differentiating between core 1 and core 2 antennae, and assigning antennae sequences, was applied to all analyses throughout this thesis.

### 3.2.3 Glycomic analysis of the colon from wild type and knockout mice

The O-glycomic analysis of the colon from wild type mice is presented in a partial mass spectrum from \(m/z\) 1000-2000 for comparison with the knockout tissues in Figure 3.2A as well as in a full range mass spectrum with detailed annotations in Figure 3.3A. All major glycan structures detected, including from the knockout tissues, are collated in Table 3.2. The results showed that the wild type colon is dominated by core 2 O-glycans, for example at \(m/z\) 1024 (\(\text{HexNAc}_3\text{Hex}_1\)) and 1763 (\(\text{HexNAc}_3\text{Hex}_2\text{Fuc}_1\text{NeuAc}_1\)) (Table I). Core 1 O-glycans are of lower abundance, for instance at \(m/z\) 1198 (\(\text{HexNAc}_3\text{Hex}_1\text{Fuc}_1\)) and 1950 (\(\text{HexNAc}_3\text{Hex}_2\text{NeuAc}_2\)). I-branches on the Gal of core 1 arm were also detected on both core 1 and core 2 O-glycans, for instance at \(m/z\) 1432 (\(\text{HexNAc}_3\text{Hex}_3\)) and 1881 (\(\text{HexNAc}_4\text{Hex}_4\)), respectively. It was also evident
that the colon was rich with glycans that are terminated with the Sda$^a$ epitope, for example at \( m/z \) 1140 (HexNAc$_2$Hex$_1$NeuAc$_1$) and 2196 (HexNAc$_4$Hex$_2$NeuAc$_2$), with the latter carrying two Sda$^a$ epitopes (Table 3.2).

![Figure 3.1. MALDI-TOF/TOF-MS and MS/MS spectra of the molecular ion \( m/z \) 1331 [M+Na]$^+$ detected in the mass fingerprinting of the stomach from wild type and C2GnT2 knockout mice.](image)

Selected sections from O-glycomic mass fingerprinting of the stomach from wild type (A) and C2GnT2 knockout (B) mice showing several [M+Na]$^+$ peaks with the same \( m/z \). High sensitivity tandem mass spectrometric data differentiates glycans of the same composition (HexNAc$_2$Hex$_2$Fuc$_2$) and mass (\( m/z \) 1331) but with different O-glycan core types. In the wild type stomach (C), the core 2 O-glycan was the only isoform detected whereas in the C2GnT2 knockout mice (D), there was a mixture of core 2 and core 1 O-glycans with the latter as the major component.
Figure 3.2. MALDI-TOF-MS profile of O-glycans from the colon of wild type and knockout mice.

The glycomic profiles of reduced and permethylated O-glycans [M+Na]^+ detected in the colon of (A) wild type, (B) C2GnT2 knockout, (C) C2GnT3 knockout and (D) C2GnT triple knockout mice. Cartoon structures assigned were based on mass spectrometric data and knowledge of O-glycan biosynthetic pathways for the selected range of m/z 1000 to 2000. X, impurities.
Table 3.2. O-glycan structures from the colon of wild type and knockout mice.

This table summarises major structures [M+Na]^+ observed in the MALDI-TOF-MS spectra of colon tissues from wild type, C2GnT2 knockout, C2GnT3 knockout and C2GnT triple knockout mice.

<table>
<thead>
<tr>
<th>m/z</th>
<th>Structures</th>
<th>m/z</th>
<th>Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>534</td>
<td></td>
<td>1402</td>
<td></td>
</tr>
<tr>
<td>708</td>
<td></td>
<td>1432</td>
<td></td>
</tr>
<tr>
<td>779</td>
<td></td>
<td>1473</td>
<td></td>
</tr>
<tr>
<td>895</td>
<td></td>
<td>1501</td>
<td></td>
</tr>
<tr>
<td>925</td>
<td></td>
<td>1518</td>
<td></td>
</tr>
<tr>
<td>953</td>
<td></td>
<td>1559</td>
<td></td>
</tr>
<tr>
<td>983</td>
<td></td>
<td>1589</td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td>1677</td>
<td></td>
</tr>
<tr>
<td>1069</td>
<td></td>
<td>1692</td>
<td></td>
</tr>
<tr>
<td>1140</td>
<td></td>
<td>1763</td>
<td></td>
</tr>
<tr>
<td>1157</td>
<td></td>
<td>1882</td>
<td></td>
</tr>
<tr>
<td>1187</td>
<td></td>
<td>1950</td>
<td></td>
</tr>
<tr>
<td>1198</td>
<td></td>
<td>2125</td>
<td></td>
</tr>
<tr>
<td>1226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1256</td>
<td></td>
<td>1286</td>
<td></td>
</tr>
<tr>
<td>1316</td>
<td></td>
<td>1331</td>
<td></td>
</tr>
<tr>
<td>1344</td>
<td></td>
<td>1372</td>
<td></td>
</tr>
<tr>
<td>1372</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

WT, wild type mice; C2GnT2 KO, C2GnT2 knockout mice; C2GnT3 KO, C2GnT3 knockout mice; C2GnT triple KO, C2GnT triple knockout mice; ND, not detected.

C2GnT2 is highly expressed in the colon relative to other tissues (Stone et al., 2009). As expected, the deficiency of this isoenzyme caused immense structural changes (As shown in partial mass spectrum in Figure 3.2B, full range mass spectra in Figure 3.3B and collated structures in Table 3.2). Analysis of the colon from C2GnT2 knockout mice showed that most core 2 O-glycan structures have been replaced with linear core 1 structures. For example, the
glycan at \( m/z \) 1157 (HexNAc\(_2\)Hex\(_2\)Fuc) in the wild type mice was a core 2 O-glycan terminated with the H antigen (Table 3.2). In the C2GnT2 knockout colon, it was revealed by MS/MS analysis that the same glycan composition was a linear core 1 O-glycan still terminated with the H antigen (Table 3.2). Interestingly, there were also changes in the type of terminal epitopes consequent upon loss of the core 2 branch. For example, there was an increase in fucosylation density on individual glycan antennae leading to the formation of the atypical glycan epitope Le\(^x\) on the core 1 O-glycan at \( m/z \) 1331 (HexNAc\(_2\)Hex\(_2\)Fuc\(_2\)). The isobaric glycan in the wild type has a core 2 structure terminated with two H antigens (Table 3.2). In addition, there was also a reduction of glycans carrying the I-branch in the C2GnT2 knockout colon compared to the wild type, for example at \( m/z \) 1432 and 1881 (Table 3.2).

C2GnT3 is expressed at relatively low abundance in the colon (Stone et al., 2009). As shown by the mass spectrometric analysis (Figure 3.2C, Figure 3.3C and Table 3.2), O-glycan structures from the colon of C2GnT3 knockout mice were similar to the colon of wild type mice. For example, the glycan at \( m/z \) 1157 in the wild type and C2GnT3 knockout tissues were both core 2 type (Table 3.2). Another example, at \( m/z \) 1432, there was a mixture of core 1 and core 2 isoforms with the same composition in the wild type colon and the mixture remained similar in the C2GnT3 knockout.

The triple knockout mice lack the ability to synthesise all three known isoforms of C2GnT. This multiple deficiency caused a complete absence of core 2 O-glycans in the colon (Figure 3.2D, Figure 3.3D and Table 3.2). In addition, there was also a complete disappearance of I-branches on both core 1 and core 2 O-glycans. For example, in the wild type colon, glycans at \( m/z \) 1432 consisted of core 2 and branched core 1 structures, whereas in the triple knockout colon, an elongated linear core 1 structure was detected (Table 3.2).
Figure 3.3. The full range MALDI-TOF-MS spectra of permethylated O-glycans from the colon of wild type and knockout mice.

The glycomic profiles of reduced and permethylated O-glycans [M+Na]+ detected in the colon of (A) wild type, (B) C2GnT2 knockout, (C) C2GnT3 knockout and (D) C2GnT triple knockout mice. Cartoon structures assigned were based on mass spectrometric data and knowledge of O-glycan biosynthetic pathways. The black arrow showing a signal with -14 mass units in each profile corresponds to underpermethylation of that particular glycan. All unannotated signals with similar intervals from annotated glycan peaks correspond to their respective underpermethylated form. X, impurities.

3.2.4 O-sulphoglycomic analysis of the colon from wild type and knockout mice

As previous work has indicated significant levels of sulphated O-glycans in colonic mucins (Inoue & Yosizawa, 1966; Thomsson et al., 2010), we considered it important to determine if similar reductions of core 2 O-glycans could be observed in the profile of sulphated
O-glycans derived from C2GnT2 knockout mice in comparison to the wild type. To this end we investigated whether the newly proposed methodology for sulphoglycomics (Yu et al., 2009) was suited to the analysis of mucinous tissues. This was, indeed, found to be the case. The O-sulphoglycomic analysis was done according to the adapted strategy as described in Section 3.2.1. The best O-sulphoglycomic profiles in terms of high intensity of sulphated O-glycans and low contamination by their non-sulphated counterparts were observed from the 2.5 mM ammonium acetate fractions in the negative ion mode [M-H]-. A significant number of sulphated O-glycans was also detected in the 10mM ammonium acetate fractions but with slightly higher contamination. MS/MS analyses, however, were acquired from both ammonium acetate/acetonitrile fractions and gave better fragmentation in the positive ion mode (as [M+2Na-H]+) compared to the negative ion mode. Examples of the MALDI-TOF-MS profiles of sulphated O-glycans from the colon of wild type and C2GnT2 knockout mice are shown in Figure 3.4A and B, respectively.

The results showed that the wild type colon, similar to the non-sulphated profile (Figure 3.2A), is dominated by core 2 O-glycans, for example at m/z 1270 [(SO₃)HexNAc₃Hex₂] and m/z 1444 [(SO₃)HexNAc₃Hex₂Fuc₁] (Figure 3.4A). Core 1 O-glycans are of lower abundance, for instance at m/z 1025 and 1199, the earlier as a mixture with an isobaric core 2 O-glycan structure. I-branches on the Gal of core 1 that were quite abundant in the non-sulphated wild type colon profile however were not detected here. In the C2GnT2 knockout mice, there was a complete loss of sulphated core 2 O-glycans concomitant with the increase in elongated sulphated core 1 O-glycans. All sulphated O-glycan compositions detected in the wild type, except for the glycan at m/z 1760, were also observed in the knockout profile entirely as core 1 isoforms. We detected only one sulphate per O-glycan molecule (monosulphated) either on the reducing or the non-reducing end in both wild type and C2GnT knockout colon. Due to the complexity of the MS/MS data and the limited fragmentation observed, not all sulphate moieties managed to be determined their exact positions. However, unexpectedly we have discovered an unusual sulphation on the reducing GalNAc of the glycans at m/z 1025 and 1270 ([M-H]) in the wild type and C2GnT2 knockout colon, respectively (see MS/MS annotations in Figure 3.5A and B, respectively). Especially relevant are the signals at m/z 708 in Figure 3.5A and m/z 953 in Figure 3.5B because they correspond to elimination of the core 1 branch without including any sulphate group. To our knowledge and also by searching the CFG database, such modification has not been reported. Sulfotransferase (ST) isoforms that act on the non-reducing terminal GalNAc have been described in mice, namely GalNAc-4-ST1 and -2 (Boregowda et al., 2005), however none
has been associated with the reducing GalNAc. Therefore, this unique modification is intriguing and should be further investigated, particularly using the MS\textsuperscript{n} analysis, in future studies.

**Figure 3.4. MALDI-TOF-MS profile of sulphated O-glycans from the colon of wild type and C2GnT2 knockout mice.**

The glycomic profiles of reduced and permethylated sulphated O-glycans [M-H]\textsuperscript{-} detected in the colon of (A) wild type and (B) C2GnT2 knockout mice for the selected range of m/z 1000 to 2000. MS/MS data were acquired in the positive ion mode [M+2Na-H]\textsuperscript{+}. The black arrow showing a signal with -14 mass units corresponds to underpermethylation of that particular glycan. All unannotated signals with similar intervals from annotated glycan peaks correspond to their respective underpermethylated form. X, impurities/unknown peaks; S, sulphate.

### 3.2.5 Glycomic analysis of the stomach from wild type and knockout mice

O-Glycomic analysis of the stomach from wild type mice revealed that this tissue, similar to the colon, was dominated with core 2 O-glycans (Figure 3.6A; structures are summarised in Table 3.3). Because the wild type murine stomach was also extensively characterised and presented in the next chapter of this thesis (see Figure 4.1), the full range mass spectrum of this tissue is not included here due to the high similarity between them. One of the unique features of the wild type stomach is the detection of high-mass O-glycan structures up to m/z 3926 (Figure
4.1). The most abundant terminal epitope in the stomach was found to be the H antigen, for example at \( m/z \) 1157 and 1331 (Figure 3.6A and Table 3.3). In addition, core 1 and core 2 O-glycans in this tissue were found to be highly branched on the core 1 arm, similar to what was seen in the colon, for example glycans at \( m/z \) 1431 and 1881. Interestingly, the glycan at \( m/z \) 1473 (HexNAc\(_4\)Hex\(_2\)) which is a core 2 O-glycan carrying HexNAc terminated antennae, was detected at a significant level. This glycan has been postulated to carry the rare and functionally important GlcNAc\(_\alpha1\)-4Gal epitope which was recently identified in the wild type mouse stomach (Kawakubo et al., 2004; Lee et al., 2008). However, enzymatic digestion with \( \beta\)-N-acetylhexosaminidase/HEXase I caused a significant reduction of this glycan showing that the terminal HexNAc residues observed here are mainly \( \beta\)-linked GlcNAc.

![Figure 3.5. The MALDI-TOF/TOF-MS/MS spectra of the sulphated O-glycans showing the unusual reducing GalNAc sulphation.](image)

The fragmentation of (A) the glycan at \( m/z \) 1071 [M+2Na-H]\(^+\) corresponding to the molecular ion \( m/z \) 1025 in the negative ion mode mass spectrum of the wild type colon and (B) the glycan at \( m/z \) 1316 [M+2Na-H]\(^+\) corresponding to the molecular ion at \( m/z \) 1270 in the negative ion mode mass spectrum of the C2GnT2 knockout colon. Due to the complexity of the fragmentation pathway, not all fragments were able to be assigned.
Figure 3.6. MALDI-TOF-MS spectra of O-glycans from the stomach of wild type and knockout mice.

The glycomic profiles of reduced and permethylated O-glycans [M+Na]+ detected in the stomach of (A) wild type, (B) C2GnT2 knockout, (C) C2GnT3 knockout and (D) C2GnT triple knockout mice. Cartoon structures assigned were based on mass spectrometric data and knowledge of O-glycan biosynthetic pathways for the selected range of m/z 1000 to 2000.
Table 3.3. O-glycan structures from the stomach of wild type and knockout mice.
This table summarises major structures [M+Na]+ observed in the MALDI-TOF-MS spectra of stomach tissues from wild type, C2GnT2 knockout, C2GnT3 knockout and C2GnT triple knockout mice.

<table>
<thead>
<tr>
<th>m/z</th>
<th>WT</th>
<th>C2GnT2 KO</th>
<th>C2GnT3 KO</th>
<th>C2GnT Triple KO</th>
</tr>
</thead>
<tbody>
<tr>
<td>534</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>708</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>738</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>779</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>895</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>912</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>925</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>953</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1086</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>1140</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>1157</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1187</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1198</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1228</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1256</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1286</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1296</td>
<td>ND</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1331</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1361</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1374</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1402</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1432</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1473</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1505</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1518</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>1535</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1578</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1589</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>1606</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1647</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1677</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1692</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>1709</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>1780</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1810</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>1852</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1882</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1922</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
</tbody>
</table>
In the C2GnT2 knockout mice, as expected, due to the relatively high expression level of C2GnT2 in the stomach (Stone et al., 2009), there was a great reduction of core 2 O-glycans.
compared to the wild type mice (Figure 3.6B, Figure 3.7A and Table 3.3). Glycans at \( m/z \) 1473, 1677 and 1922 are examples of the remaining core 2 structures, all carrying one or two terminal HexNAc residues (Figure 3.6B and Table 3.3). The C2GnT2 deficient mice synthesised abundant extended core 1 O-glycans in the stomach, for instance at \( m/z \) 1187, 1331 and 1361. Moreover, a significant reduction of branched O-glycans was also observed in the stomach for example at \( m/z \) 1606 and 1780. Similar to the colon of C2GnT2 knockout mice, there was an increase in fucosylation density on individual glycan antennae leading to the formation of Lewis\(^x\), which was not detected in the wild type, for instance on the glycan at \( m/z \) 1331. Another interesting finding from this tissue was an emergence of \( O \)-linked mannosyl glycans at \( m/z \) 738 and 912, which were not detected in the wild type stomach.

C2GnT3 is also not the main C2GnT isoenzyme in the stomach, similar to the colon (Stone et al., 2009). There were no substantial changes in the O-glycomic profile of the stomach from C2GnT3 knockout mice (Figure 3.6C, Figure 3.7B and Table 3.3). However, unexpectedly, there was an increase of the glycan at \( m/z \) 1473 (HexNAc\(_3\)Hex\(_2\)) which is a core 2 O-glycan carrying HexNAc terminated antennae as seen in the wild type and C2GnT2 knockout stomach.

In the stomach of C2GnT triple knockout mice, there was a complete loss of core 2 O-glycans (Figure 3.6D, Figure 3.7C and Table 3.3). A significant reduction of non-fucosylated O-glycans compared to the wild type and both single knockouts was also observed, for example at \( m/z \) 1024, 1432, 1473, 1677 and 1922 in the stomach (Figure 3.6D and Table 3.3). There was also an emergence of highly fucosylated and elongated O-linked mannose structures, for instance at \( m/z \) 1187, 1535, 1709, 1810 and 1985, which were not detected at significant levels in the stomach of wild type and C2GnT3 knockout mice and were very minor in the C2GnT2 knockout stomach (Figure 3.6D and Table 3.3). The existence of O-linked mannose is supported by the detection of 2-linked mannitol by GC-EI-MS linkage analysis (data not shown).
Figure 3.7. The full range MALDI-TOF-MS spectra of permethylated O-glycans from the stomach of knockout mice.

The glycomic profiles of reduced and permethylated O-glycans [M+Na]+ detected in the colon of (A) C2GnT2 knockout, (B) C2GnT3 knockout and (C) C2GnT triple knockout mice. The black arrow showing a signal with -14 mass units in each profile corresponds to underpermethylation of that particular glycan. All
unannotated signals with similar intervals from annotated glycan peaks correspond to their respective underpermethylated form. X, impurities

3.2.6 Glycomic analysis of the small intestine from wild type and knockout mice

The O-glycomic analysis of small intestine from wild type mice showed that, similar to the stomach and colon, it was dominated by core 2 O-glycans, for example at \( \text{m/z} \) 1402 and 1647 (Figure 3.8A, Table 3.4 and Figure 3.9A). The most abundant terminal epitopes were the H antigen (for example on glycans at \( \text{m/z} \) 1157 and 1331) and Sd\(^a\) (for example at \( \text{m/z} \) 1140 and 1589).

In contrast to the stomach and colon, core 2 O-glycans were still the major structures in the small intestine of C2GnT2 knockout mice (Figure 3.8B, Table 3.4 and Figure 3.9B). There was a moderate emergence of core 1 structures whose glycan compositions were not detected in the wild type such as \( \text{m/z} \) 1501 and 1950, both carrying the Sd\(^a\) epitope. An increase in core 1 structures sialylated with NeuAc and/or NeuGc such as at \( \text{m/z} \) 1256, 1286 and 1316 were also observed. In addition there were increases in the abundance of molecular ions whose MS/MS spectra showed that they were mixtures of core 1 and core 2 structures, for example at \( \text{m/z} \) 1157 and 1589.

C2GnT3 is highly expressed in the small intestine relative to other tissues (Stone et al., 2009). The C2GnT3 deficiency in the small intestine caused a relatively moderate reduction of core 2 structures with or without an increase of their core 1 isoforms compared to the wild type mice, for example at \( \text{m/z} \) 1331 and 1589 (Figure 3.8C, Table 3.4 and Figure 3.9C). In contrast to the wild type small intestine, there was a global reduction of glycans carrying the H antigen (for example at \( \text{m/z} \) 1157 and 1402), leaving the Sd\(^a\) epitope as the most abundant terminal epitope, for example at \( \text{m/z} \) 1140 and 1589. Similar to the C2GnT2 deficient small intestine, there were increases in sialylated core 1 structures compared to the wild type, for instance at \( \text{m/z} \) 1256, 1286 and 1316.
Figure 3.8. MALDI-TOF-MS profile of O-glycans from the small intestine of wild type and knockout mice.

The glycomic profiles of reduced and permethylated O-glycans [M+Na]+ detected in the small intestine of (A) wild type, (B) C2GnT2 knockout, (C) C2GnT3 knockout and (D) C2GnT triple knockout mice. Cartoon structures assigned were based on mass spectrometric data and knowledge of O-glycan biosynthetic pathways for the selected range of m/z 1000 to 2000.
Table 3.4. O-glycan structures from the small intestine of wild type and knockout mice.

This table summarises major structures [M+Na]+ observed in the MALDI-TOF-MS spectra of small intestine tissues from wild type, C2GnT2 knockout, C2GnT3 knockout and C2GnT triple knockout mice.

<table>
<thead>
<tr>
<th>m/z</th>
<th>Structures</th>
<th>m/z</th>
<th>Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>534</td>
<td></td>
<td>1432</td>
<td></td>
</tr>
<tr>
<td>708</td>
<td></td>
<td>1473</td>
<td></td>
</tr>
<tr>
<td>779</td>
<td></td>
<td>1501</td>
<td></td>
</tr>
<tr>
<td>895</td>
<td></td>
<td>1518</td>
<td></td>
</tr>
<tr>
<td>925</td>
<td></td>
<td>1576</td>
<td></td>
</tr>
<tr>
<td>953</td>
<td></td>
<td>1589</td>
<td></td>
</tr>
<tr>
<td>983</td>
<td></td>
<td>1763</td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td>1780</td>
<td></td>
</tr>
<tr>
<td>1069</td>
<td></td>
<td>1834</td>
<td></td>
</tr>
<tr>
<td>1140</td>
<td></td>
<td>1851</td>
<td></td>
</tr>
<tr>
<td>1157</td>
<td></td>
<td>1950</td>
<td></td>
</tr>
<tr>
<td>1187</td>
<td></td>
<td>2026</td>
<td></td>
</tr>
<tr>
<td>1228</td>
<td></td>
<td>2039</td>
<td></td>
</tr>
<tr>
<td>1256</td>
<td></td>
<td>2196</td>
<td></td>
</tr>
<tr>
<td>1286</td>
<td></td>
<td>2230</td>
<td></td>
</tr>
<tr>
<td>1296</td>
<td></td>
<td>2645</td>
<td></td>
</tr>
</tbody>
</table>

WT, wild type mice; C2GnT2 KO, C2GnT2 knockout mice; C2GnT3 KO, C2GnT3 knockout mice; C2GnT triple KO, C2GnT triple knockout mice; ND, not detected.

The small intestine of the triple knockout mice again demonstrated a complete disappearance of core 2 structures (Figure 3.8D, Table 3.4 and Figure 3.9D). Similar to the C2GnT3 deficient small intestine, the Sdα epitope was the most abundant type of terminal structure, for instance at m/z 1140 and 1589. Interestingly, at m/z 1344, it was shown by the
MS/MS analysis that there was a mixture of sialylated core 1 O-glycans and an elongated O-linked mannose structure terminated with the Sda epitope. O-linked mannosyl glycans were not detected at any significant levels in the wild type and single knockouts of small intestine.
Figure 3.9. The full range MALDI-TOF-MS spectra of permethylated O-glycans from the small intestine of wild type and knockout mice.

The glycomic profiles of reduced and permethylated O-glycans [M+Na]+ detected in the small intestine of (A) wild type, (B) C2GnT2 knockout, (C) C2GnT3 knockout and (D) C2GnT triple knockout mice. The black
arrow showing a signal with -14 mass units in each profile corresponds to underpermethylation of that particular glycan. All unannotated signals with similar intervals from annotated glycan peaks correspond to their respective underpermethylated form. X, impurities; N, reduced N-glycans.

3.2.7 Glycomic analysis of the thyroid/trachea, kidneys and thymus from wild type and knockout mice

O-Glycomic analysis of the thyroid/trachea and kidneys from wild type mice showed that, similar to other tissues, core 2 structures were the most abundant type of O-glycan (Figure 3.10A and Figure 3.11A), whereas in the wild type thymus, core 1 structures were the most abundant (Figure 3.12A). In the thyroid/trachea, a relatively low number of glycan structures were detected and mostly were terminated with sialylation either on its own or as part of the Sda epitope, for example at \( m/z \) 1344 and 1589, respectively (Figure 3.10A). In the wild type kidneys, however, the core 2 O-glycans were moderately fucosylated leading to the formation of H and Lewis\(^{\text{xy}}\) terminal epitopes, for example on glycans at \( m/z \) 1157 and 1780 (Figure 3.11A). In the thymus, similar to the thyroid/trachea, few O-glycan structures were detected (Figure 3.12A). The O-glycans were also highly sialylated with NeuAc or NeuGc, for example at \( m/z \) 1256 and 1316.

Remarkably the kidneys from C2GnT2 knockout mice exhibited a significant reduction in the abundance of some of the core 2 O-glycans with no concomitant compensation by extended core 1 sequences (Figure 3.11B; note loss of \( m/z \) 1402 and higher compared with the wild type). Thus, only a single glycan structure of the core 1 type (\( m/z \) 1256) was detected in the mass range of \( m/z \) 1000 – 2000, accompanied by residual core 2 O-glycans. The thymus and thyroid/trachea of C2GnT2 knockout mice were not analysed due to the absence of any interesting phenotypic findings from these tissues, in addition to the relatively very low expression of C2GnT2 in the wild type thymus and thyroid/trachea (Stone et al., 2009).

Even though C2GnT3 is believed to play important biological roles in the human thymus (Schwientek et al., 2000), our O-glycomic screening has revealed that there were no significant changes, if any, with respect to glycan structures and abundances between the thymus of wild type and the C2GnT3 knockout mice (Figure 3.12B). Similar effects were also observed in the kidneys of C2GnT3 knockout mice (Figure 3.11C). In the thyroid/trachea, however, there was a minor disappearance of core 2 structures, and if detected, they were at a much reduced abundance relative to the wild type, for example glycan at \( m/z \) 1157, 1589 and 1763, with the latter two being decorated with the Sda epitope (Figure 3.10B).
In the C2GnT triple knockout mice, the thyroid/trachea and the kidneys also exhibit a complete absence of core 2 O-glycans (Figure 3.10C and Figure 3.11D), similar to that seen in other tissues. In the thyroid/trachea of C2GnT triple knockout there was also a complete absence of glycans bearing the Sd\(^a\) epitope (Figure 3.10C). Meanwhile in the kidneys, the glycomic analysis revealed a remarkable loss of O-glycans with only two structures having simple core 1 type sequences being detected (Figure 3.11D). The thymus of C2GnT triple knockout mice was not analysed.

Figure 3.10. MALDI-TOF-MS profile of O-glycans from the thyroid/trachea of wild type and knockout mice.

The glycomic profiles of reduced and permethylated O-glycans [M+Na]\(^+\) detected in the thyroid/trachea of (A) wild type, (B) C2GnT3 knockout and (C) C2GnT triple knockout mice. Cartoon structures assigned
were based on mass spectrometric data and knowledge of O-glycan biosynthetic pathways for the selected range of \( m/z \) 1000 to 2000.

Figure 3.11. MALDI-TOF-MS profile of O-glycans from the kidneys of wild type and knockout mice.
The glycomic profiles of reduced and permethylated O-glycans [M+Na]^+ detected in the kidneys of (A) wild type, (B) C2GnT2 knockout, (C) C2GnT3 knockout and (D) C2GnT triple knockout mice. Cartoon structures assigned were based on mass spectrometric data and knowledge of O-glycan biosynthetic pathways for the selected range of m/z 1000 to 2000.

Figure 3.12. MALDI-TOF-MS profile of O-glycans from the thymus of wild type and knockout mice. The glycomic profiles of reduced and permethylated O-glycans [M+Na]^+ detected in the thymus of (A) wild type and (B) C2GnT3 knockout mice. Cartoon structures assigned were based on mass spectrometric data and knowledge of O-glycan biosynthetic pathways for the selected range of m/z 1000 to 2000.

3.2.8 Determination of branching sites in the stomach and colon

Stomach and colon tissues exhibited very varied O-glycan structures accompanied by dramatic changes in the knockout tissues. In addition, the MS/MS data provided evidence for a number of potentially biologically important glycan epitopes in these tissues. To put some of these structural assignments on a firmer footing, we were fortunate to have access to a novel MALDI-QIT-TOF mass spectrometer (Ojima et al., 2005) which has been developed for high accuracy, ultra-high sensitivity MS^n analyses. Thus, we performed MS^3 analysis on selected glycan MS/MS fragments that contain branched N-acetyllactosamine (LacNAc) chains in order to define the I-branching sites.
As an example, Figure 3.13 shows MALDI-QIT-TOF-MS/MS data for the glycan at m/z 1780 (HexNAc$_3$Hex$_3$Fuc$_2$) of wild type stomach (A), together with MS$^3$ spectrum from collisional activation of the fragment at m/z 1143 (B). MS/MS of the glycan at m/z 1780 has proven that the core 1 Gal of this glycan is branched (see annotation on Figure 3.13A). In fact, this is the only location of the I-branch that we have detected in both stomach and colon in this study even when the core 1 antenna is elongated with an additional GlcNAc residue, for example m/z 2617 in the stomach tissues (Table 3.3), which carries one more GlcNAc residue on both the core 1 antenna and the I-branch than m/z 1780. MS$^3$ on the fragment at m/z 1143 further confirmed the branched Gal, as well as providing additional verification of the known I-branch linking position (Figure 3.13B). The cross-ring fragment of m/z 720 has provided the evidence of a branching on either the positions 4 or 6. Additionally, GC-EI-MS linkage analysis has shown the existence of a 3,6-linked Gal residue (data not shown).

![Figure 3.13. Determination and verification of the I-branching site and position.](image)

(A) MALDI-QIT-TOF-MS/MS spectrum of m/z 1780 [M+Na]$^+$ from wild type stomach shows evidence of branching on Gal of the core 1 arm (fragments m/z 506 and 1143) and the terminal epitope blood group H antigen on each of the branches (fragments m/z 660 and 1370). (B) MALDI-QIT-TOF-MS$^3$ spectrum of the MS/MS fragment at m/z 1143 from (A) has provided additional evidence for the branched Gal residue of
the core 1 arm and verification of the branch attachment on position 6 (cross ring fragment at m/z 720). For simplicity, not all fragments are annotated.

### 3.2.9 Characterisation of RM2-like structure in the colon

We have detected an unusual terminal epitope in the O-glycan observed at m/z 1950 (HexNAc3Hex2NeuAc2) in the wild type and knockout tissues of colon (Figure 3.2 and Table 3.2). Initial MS/MS fragmentation data from the MALDI-TOF/TOF suggested a sialylated Sd\(^a\) epitope (HexNAc2Hex1NeuAc2) whose composition is identical to a rare epitope originally found on a glycolipid and called the RM2 antigen (Ito et al., 2001). For further confirmation of this structure, we have performed additional MS/MS analysis on m/z 1950 (Figure 3.14A) as well as MS\(^3\) analysis on two major MS/MS fragments, namely m/z 1108 and 1453 using MALDI-QIT-TOF (Figure 3.14B and C). Glycosidic and cross-ring fragmentation data of the MS\(^3\) experiment explicitly showed that one of the NeuAc residues was attached to the internal GlcNAc residue rather than the GalNAcitol residue, which is another possible position for sialylation. The latter was detected in the small intestine tissues and C2GnT triple knockout colon tissues. Overall, these MS\(^3\) data supported the RM2 structural assignment of the initial MS/MS analysis.

![Figure 3.14. Structural confirmation of the RM2-like epitope found in the colon tissues.](image)
(A) MALDI-QIT-TOF-MS/MS spectrum of glycan at \( m/z \) 1950 [M+Na]\(^+\) from wild type stomach shows evidence for a sialylated Sda terminal epitope. This assignment is further supported by the subsequent MS\(^3\) analysis on the MS/MS fragments of (B) \( m/z \) 1108 and (C) \( m/z \) 1453. Glycosidic and cross-ring fragments at \( m/z \) 673, 472, 520, 819 and 865 from both MS\(^3\) spectra verified the existence of the Sda epitope with additional sialylation on the internal GlcNAc therefore confirming the structural assignment of the MS/MS analysis. For simplicity, not all fragments are annotated.

3.3 Discussion

O-Glycomic analyses were performed on mice deficient in C2GnT2, C2GnT3 and all three C2GnT isoenzymes (C2GnT triple knockouts). This work aimed to further clarify the functional properties of each of the isoenzymes in various tissues and to associate structural changes with observed phenotypes of the knockout mice. This study also aimed to optimise the O-glycomic methodology as well as providing detail O-glycan fingerprinting of selected wild type murine tissues as a reference database for current and future studies. High quality mass spectrometric and tandem mass spectrometric data were obtained by using a MALDI-TOF/TOF mass spectrometer in concert with ESI-QTOF, GC-EI and MALDI-QIT-TOF mass spectrometers.

Generally, C2GnT2 knockout mice showed a significant decrease of core 2 O-glycans especially in the colon and stomach compared to the wild type mice. This was expected as this isoenzyme is highly expressed in these tissues. The glycosylation machinery synthesised more of the extended core 1 O-glycans possibly to compensate for the diminishing core 2 O-glycans. Position 6 of the GalNAcitol of core 1 O-glycans was either left unmodified or sialylated. Previous work has characterised core 3 and core 4 O-glycans in human intestine (Robbe \textit{et al.}, 2004; Holmen Larsson \textit{et al.}, 2009). In contrast these core types are hardly expressed in the murine intestine. Because trace amounts could not be unambiguously determined in the presence of the abundant core 1/2 sequences, addressing possible changes in their abundance as a consequence of loss of C2GnT2 was not feasible. Nevertheless, our collaborators did not detect a significant level of C4GnT enzymatic activity in colon, small intestine and stomach of C2GnT2 knockout mice (Stone \textit{et al.}, 2009), thus suggesting that C2GnT is entirely responsible for the production of core 4 O-glycans.

Branching of the core 1 arm of both core 1 and core 2 O-glycans was found to be very abundant especially in the stomach of wild type mice. C2GnT2 deficiency in mice caused a major
reduction and disappearance of glycan branches, therefore demonstrating that C2GnT2 has a strong capability to branch the LacNAc chains specifically on position 6 of the Gal of the core 1 arm. To our knowledge, this analysis is the first to show the I-branching activity of C2GnT2 in vivo using definitive structural analysis. This finding supports previous work that has shown in vitro that human and mouse C2GnT2 have the I-branching capability quite similar to IGnT, but no structural proof was provided (Yeh et al., 1999; Hashimoto et al., 2007). Based on in vitro assays, Yeh et al. (1999) suggested that human C2GnT2 possess higher distally acting IGnT (dIGnT) activity than centrally acting IGnT (cIGnT) activity, although the difference was merely 5%. However, the current finding has shown that centrally attached glycan branches were greatly affected by the loss of C2GnT2 activity. The centrally attached glycan branches specifically on the Gal residue of the core 1 arm was the only type identified in this study, even though the arm was elongated with two LacNAc units. All these data suggest that, C2GnT2 acts differently in vivo where it prefers to branch O-glycans centrally instead of distally. Similar branching of the Gal residue of the core 1 arm has previously been characterised rigorously on O-glycans derived from human Tamm-Horsfall glycoprotein (THP) during pregnancy (Easton et al., 2000). Therefore, it is tempting to speculate that C2GnT2’s natural substrate is the Gal of GlcNAcβ1-3Galβ1-3GalNAc-serine/threonine, with or without the core 2 arm attached and elongation on the core 1 arm. The fact that we did not detect other types of branched LacNAc on O-glycan is most likely because the IGnT isoenzymes, the only other enzymes known for I-branching activity, are not significantly expressed in murine GI tract (Twu et al., 2003). We are unable to conclude whether this finding is tissue-specific due to the lack of branched O-glycan structures detected in other tissues analysed.

It was also evident from our data that branched core 1 O-glycans were not completely absent in the stomach and colon of the C2GnT2 single knockout, but were totally lost in the C2GnT triple knockout mice. Therefore, it can be speculated that C2GnT1 and/or C2GnT3 possess a weak I-branching activity and, instead of IGnT, compensated for C2GnT2 I-branching activity in C2GnT2 knockout mice. These results are however in contrast with previous immunohistochemical findings on IGnT deficient mice which concluded that IGnT is the major enzyme that synthesises I-branches in the murine stomach (Chen et al., 2005). This discrepancy might be due to the method the previous study applied to detect the reduction of IGnT activity. It is conceivable that the monoclonal antibody against I antigen that was used could be specific only to I-branches that are attached towards the non-reducing end of a poly-LacNAc chain (distally) which are supposed to be produced by IGnTs instead of C2GnTs.
Our collaborators have demonstrated that C2GnT2 deficient mice have impaired mucosal barrier and increased susceptible to dextran sodium sulphate (DSS)-induced colitis compared to the wild type mice (Stone et al., 2009). In 2007, An et al. reported that mice with glycosylation changes in their GI tract are more susceptible to colitis (An et al., 2007). The mechanism of DSS-induced colitis is not clear but destruction of the mucin content is one of the proposed mechanisms (Melgar et al., 2005). It is anticipated that a similar situation is being observed here. The loss of glycan branches (core 2 branches and/or I-branches) of both sulphated and non-sulphated O-glycans that was observed in the C2GnT2 knockout mice could possibly disrupt the normal glycan structures which protect the underlying proteins in mucin layers. The interrupted conformation of mucins in the GI tract therefore facilitates DSS’s destructive action. These changes might be recognised as foreign by the immune system, which is possibly the reason for the increased susceptibility to inflammation which causes colitis. In addition, a direct relationship between immune cells and C2GnT2 has been demonstrated by the up regulation of this isoenzyme by T helper 2 cytokines in human airway mucins (Beum et al., 2005). On the other hand, Johansson et al. suggested that the disruption of colonic mucins by DSS causes the mucin layer to become more permeable to bacterial intrusion followed by colonisation on the epithelial lining and thus triggering inflammatory reaction (Johansson et al., 2010). Furthermore, the loss of sulphation on colonic mucins also has been indicated to increase the susceptibility towards colitis in humans and mice (Corfield et al., 1993; Corfield et al., 1996; Tobisawa et al., 2010). Therefore, it can be speculated that sulphated branched O-glycans of colonic mucins have a protective role in colitis.

C2GnT3 was originally shown by Northern blot analysis to be highly expressed in human thymus and has been implicated to play important roles in T-cell development and lymphocyte trafficking (Schwientek et al., 2000). From our structural O-glycomic analysis of the C2GnT3 knockout mice, generally there were no substantial changes on the abundance of core 2 O-glycans in any tissues that have been analysed, albeit a minor reduction was observed in the small intestine and thyroid/trachea. The lack of any significant changes to O-glycan structures and abundances in the thymus of C2GnT3 knockout mice confirms quantitative polymerase chain reaction (qPCR) findings by our collaborators that indicated low expression of that isoenzyme in murine thymus (Stone et al., 2009), as opposed to human thymus. This indicates that either the other C2GnT isoenzymes can substantially compensate for the loss of C2GnT3 activity, or C2GnT3 is not normally responsible for synthesising the vast majority of core 2 O-glycans in the murine tissues examined.
As noted above, the loss of C2GnT3 activity caused a slight but reproducible reduction of core 2 O-glycans in the thyroid, the thyroxine producing gland. Our collaborators have shown that in C2GnT3 deficient mice there is a decrease in circulating levels of alkaline phosphatase and thyroxine, but the thyroid stimulating hormone was not affected (Stone et al., 2009). This observation is associated with hypothyroidism but the mechanism is still not fully understood (Fliers et al., 2006). The major reduction of core 2 O-glycans carrying the Sda epitope in the C2GnT3 deficient thyroid/trachea is intriguing, supported by the complete loss of Sda terminated glycans in the C2GnT triple deficient thyroid/trachea, which could illustrate that in the thyroid/trachea, the Sda epitope can be synthesised efficiently only on the core 2 arm. We postulate that the population of core 2 O-glycans in the thyroid/trachea whose biosynthesis is dominantly controlled by C2GnT3 are somehow necessary in the production of thyroxine. Reduced thyroxine levels will subsequently contribute to the hypothyroidism and the reduction of alkaline phosphatase. Whether the presence of the Sda epitope is functionally important in the thyroid remains to be established. The Sda epitope has been found on a limited repertoire of interesting glycoproteins in mammals, including human Tamm-Horsfall protein (Serafini-Cessi et al., 1986), the murine zona pellucida (Dell et al., 2003), bovine pregnancy-associated glycoproteins (Klisch et al., 2008) and glycoproteins on murine cytotoxic T-lymphocytes (Smith & Lowe, 1994), although no specific function has so far been attributed to this epitope.

Glycoproteins from the thyroid which carry these core 2 O-glycans would be an important target for further investigations.

*Clostridium perfringens* and *Helicobacter pylori* are common bacteria that colonise the GI tract. Recently, it has been revealed that *C. perfringens* contains an endo-β-galactosidase that is able to release the disaccharide epitope GlcNAcα1-4Gal from glycoconjugates as part of its pathogenicity (Ashida et al., 2001). It can be interpreted that *C. perfringens* uses this epitope as a binding site to initiate infection. On the other hand, the terminal α1,4-linked GlcNAc residue has been shown to affect the growth of *H. pylori* in the stomach (Kawakubo et al., 2004). Core 2 O-glycans capped with this epitope were found to be capable of serving as a natural antibiotic against *H. pylori* infection (Lee et al., 2008). Expression cloning of human α1,4-N-acetylgalactosaminyltransferase has shown that this epitope is efficiently transferred to core 2 O-glycans forming the structure GlcNAcα1-4Galβ1-4GlcNAcβ1-6(GlcNAcα1-4Galβ1-3)GalNAc (Nakayama et al., 1999). We have detected a putative GlcNAcα1-4Gal epitope on the similar core 2 O-glycan that is observed at m/z 1473 in stomach tissues (Figure 3.6 and Table 3.3). However, the terminal HexNAc residues in murine stomach have been established to be
dominated by β-GlcNAc, using a β-HexNAc specific exoglycosidase. Previously, the α-GlcNAc residue has been shown to be expressed specifically on class III mucin in the deeper layer of gastric mucosa (Nakamura et al., 1998), hence it is very low in abundance and could possibly constitute the residual undigested terminal HexNAc in our samples. Intriguingly, this glycan at m/z 1473 was well preserved as a core 2 isoform in the C2GnT2 and C2GnT3 single knockout mice, and in fact was augmented in the latter tissue, hence illustrating its potential functional importance. Detailed characterisation of differences in the glycosylation profile especially in the GI tract is fundamental to better understand host-pathogen interplay (Dell et al., 1999; van Kooyk & Rabinovich, 2008), which is also the main topic of the next chapter of this thesis.

From the analysis on C2GnT triple knockout mice, we observed a complete absence of core 2 O-glycans in all tissues. This is supported by the findings from enzyme assay experiments done by our collaborators that indicated very little C2GnT activity in these mice that remarkably were viable and fertile (Stone et al., 2009). Therefore, we propose that there is no other C2GnT isoenzyme in mouse. In addition, we have also noted that due to the major loss of the core 2 antennae and the I-branches on core 1 antennae, several unique glycans that were not originally being synthesised in the wild type murine tissues have emerged, for instance the triply fucosylated glycan at m/z 1505 in the stomach which is predicted to be a novel structure (Figure 3.6D and Table 3.3). Our tandem mass spectrometric, linkage analysis and enzymatic digestion data have shown that LacNAc chain type 2 dominated murine stomach (data shown and discussed in Chapter 4), therefore the immense amount of fucosylated Gal and difucosylated LacNAc terminals are likely to be H type 2 and Lewisy antigens, respectively.

O-linked mannosyl glycans were first identified in 1969 in yeast and were found to be very abundant in the cell wall (Sentandreu, 1969). The modification was only discovered in the mammalian system about ten years later on a brain proteoglycan (Finne et al., 1979). Subsequently, it has been recognised to constitute about 30% of brain O-glycans (Chai et al., 1999) and also to be a crucial modification on α-dystroglycan especially in the muscle tissues (Endo, 1999; Yoshida-Moriguchi et al., 2010), as has been discussed in Section 1.2.2.1 in Chapter 1. Unexpectedly, O-linked mannosyl glycans were up-regulated in the stomach of the triple knockout mouse, being barely detected in the wild type and single knockouts. Cells possibly amplified the addition of complex antennae to O-linked mannose in order to replace the diminished core 2 O-glycans. Reduced competition for UDP-GlcNAc, which is the donor substrate for both C2GnT and the O-mannose elongating enzyme POMGnT1 (Takahashi et al.,
m/z 1709 and m/z 1985, respectively in Figure 3.6D). In addition, a glycan whose composition is consistent with a unique elongated O-linked mannose sequence terminated with the Sda epitope was detected in the small intestine of C2GnT triple knockout mice (m/z 1344 in Figure 3.8D). To our knowledge these modifications of O-mannosyl glycans have not been reported elsewhere. It is anticipated that the current findings will open up new avenues for research of O-linked mannose.

In contrast to the GI tract, there seems to be no induction of any types of O-glycan accompanying the loss of core 2 O-glycans in the kidneys of single and triple deficient mice. In the kidneys of C2GnT2 knockout but not the C2GnT3 knockout mice, we observed a substantial reduction in some of the core 2 sequences, most notably those with extended core 1 and core 2 antennae (see Figure 3.11, glycans at m/z 1402 and higher), although between these two isoenzymes, the latter was shown by our collaborators to have slightly higher expression level in the kidneys. The structural changes were unexpected because Ellies et al. have shown previously that the kidneys of the C2GnT1 knockout mouse exhibit over 90% reduction in core 2 activity compared with the wild type (Ellies et al., 1998) suggesting that the C2GnT1 isoform is responsible for most of the core 2 synthesis in this organ. We consider it likely that C2GnT2 plays an important role in biosynthesis of a portion of the O-glycan repertoire in the kidney and that C2GnT1 is not able to compensate for its absence, thereby leading to the disappearance of certain core 2 O-glycan sequences. Furthermore, only two O-glycan structures of core 1 type were detected in the kidneys of C2GnT triple knockout mice. It is intriguing that the glycosylation system in the kidneys was unable to compensate for the loss of core 2 O-glycans by synthesising other types of O-glycans, as has been seen in the stomach and colon. The structural compensations observed in the latter two tissues could be resembling the mechanism of homeostatic regulatory via lectin-glycan interactions leading to changes in gene expression for the purpose of maintaining cell surface glycan density (Dam & Brewer, 2010). Dam and Brewer
also suggested that the density of glycan epitopes on the surface is at least as important as the affinity of single epitopes for the function of lectin receptors (Dam & Brewer, 2010).

Sulphation is a common modification of both N- and O-glycans however, in vertebrates, glycan sulphation is generally restricted to Gal, GlcNAc and GalNAc at internal or terminal positions (Stanley & Cummings, 2009). Sulphated oligosaccharides have posed particular analytical challenges due to their acidity and lability (Zaia, 2004). Nevertheless, it is crucial to develop a robust technique to structurally characterised these glycans since they consist of a number of functionally important glycoepitopes, for instance sulfo-SLe\(^{\alpha}\) and sulfo-Le\(^{\alpha}\), both are ligands for mammalian selectins (see Section 1.2.3 of Chapter 1). In addition to the correlation to increase the susceptibility towards colitis as discussed above, reduction in glycan sulphation has also been associated with adenocarcinoma development (Vavasseur et al., 1994). Over the years, various strategies have been proposed in order to optimise mass spectrometric characterisation of sulphated oligosaccharides, in particular sulphated N- and O-glycans. For instance, Dell and colleagues have proposed the short Hakomori permethylation method for FAB-MS analysis of sulphated samples (Dell et al., 1994). Sequencing of underivatised sulphated glycans isolated from mucins of porcine stomach and large intestine and human colonic MUC2 have been done by using LC-ESI-MS/MS (Thomsson et al., 2000; Holmen Larsson et al., 2009). The same group also has proposed that LC-ESI-MS analysis done at higher pH can improved the detection of sulphated glycans (Thomsson et al., 2010). Recently, a novel sodium hydroxide methylation strategy has been presented by the Khoo group (see Section 1.4.8.3 in Chapter 1) who successfully isolated and characterised sulphated O-glycans from murine secondary lymph nodes using MALDI-TOF/TOF-MS/MS (Yu et al., 2009). This method has been adapted for the analysis of murine colons in this study.

O-sulphoglycomic analysis was only done on the wild type and C2GnT2 knockout colons due the limited time available for investigating this new methodology. The choice of colon was driven by the findings of previous work that have indicated sulphated O-glycans in this tissue from humans and animals (Corfield et al., 1996; Holmen Larsson et al., 2009; Thomsson et al., 2010). However the main objective of this preliminary experiment was to adapt the proposed sulphoglycomic strategy by the Khoo group to complex mucinous tissues. We observed a complete loss of core 2 branches in the C2GnT2 knockout colon with concomitant increase of elongated core 1. Only linear LacNAc chains and monosulphated structures were detected in both wild type and knockout tissues. These preliminary results indicated that sulphated glycans are
affected by the loss of C2GnT2 similar to the non-sulphated glycans. Now that the sulphoglycomic methodologies have been shown to be robust, future glycomic studies should include sulphoglycomic analysis to enable more holistic assessment of the structural changes in C2GnT deficient mice. Analysing the sulphated glycans has proven to be more challenging than the non-sulphated glycans. It was noted that sulphated O-glycans in these tissues are almost equally divided between the 2.5 mM and 10 mM ammonium acetate fractions, and therefore contributed to the rather low intensity of sulphated glycan peaks in the mass spectra of each fraction. Combining these fractions would not be feasible as the latter fraction also contains non-sulphated glycans while the former are reasonably clean. Therefore, eluting with an intermediate concentration of ammonium acetate could be a good resolution. Furthermore, MS/MS sequencing was only beneficial in the positive ion mode which requires the conversion of each of the \( m/z \) of the detected negatively charged molecular ions in the MS analysis to positively charged molecular ions that usually carry more than one sodium ion. MS/MS in the negative mode entails high energy CID to gain useful fragment ions, as has been demonstrated previously (Yu et al., 2009), but requires instrumentation that is not widely available.

Another interesting finding from the murine GI tract is the detection of a sialylated Sda terminal sequence (HexNAc₂Hex₁NeuAc₂) on the glycan at \( m/z \) 1950 in the colon of wild type and knockout mice (Table 3.2). This structure has been confirmed with tandem mass spectrometry and is found to share its terminal sequence with the RM2 antigen of glycolipids; therefore it is designated as RM2-like antigen. The RM2 antigen was originally reported as a novel ganglioside in 2001 after being detected by the RM2 antibody (Ito et al., 2001), and is now well recognised as a prostate cancer marker (Saito et al., 2005). To our knowledge, this work is the first to characterise the RM2 antigen in glycoproteins. In addition, we also have detected a putative fucosylated version of this glycan at \( m/z \) 2125 in the colon of wild type and C2GnT3 knockout mice (Table 3.2). In addition, we have performed immunoblotting experiments of gel-separated wild type colon homogenates using the RM2 antibody, kindly donated by Professor Sen-itirō Hakomori (data not shown). However, the blotted samples failed to be significantly recognised by the RM2 antibody. The complex nature of unpurified colonic lysate in addition to the relatively very low abundance of the RM2-like epitopes might contribute the lack of binding, hence this experiment needs to be done on purified mucins. The RM2-like structure, therefore, remains to be rigorously characterised and its possible functional similarities with the RM2 antigen remain to be studied.
Understanding the functional consequences of the changes to glycan terminal epitopes and branching observed in the knockout tissues, especially the triple knockouts, will require considerable further work, although clues have already been provided by the published phenotypic studies of our collaborators (Stone et al., 2009). Often, the terminal epitopes are the biologically important part of any particular glycan. Research has shown numerous examples of glycan epitopes serving as messengers between cells of similar or different types including cancer cells and microorganisms (Ashida et al., 2001; Rydell et al., 2009; Cazet et al., 2010). The type of chain or branch where these antigens reside is not of any less significance in many cases (Mitoma et al., 2003; Muramatsu et al., 2008). Therefore, it would be interesting to challenge these knockout mice with more biological stress and microorganism infection and evaluate the effects of the structural aberrations.

As a summary, we have performed detailed O-glycomic analysis of neutral and sialylated O-glycans of selected tissues from mice with deficiency in C2GnT2, C2GnT3 and C2GnT1-3. This work also comprises among the very few rigorous O-glycan structural characterisations even in the wild type mice. We have demonstrated the critical functions of C2GnT isoenzymes in producing core 2 O-glycans and I-branches. Putative pathological consequences of the loss of these structures in mice include the increase in susceptibility to colitis and hypothyroidism. In addition, we have serendipitously identified RM2-like antigen from colonic glycoproteins. We also have performed preliminary O-sulphoglycomic analysis of colon and demonstrated similar structural changes in the knockout mice as have been observed for the non-sulphated glycans. It is increasingly clear that C2GnT isoenzymes have similar and also distinct functional properties; each may regulate the synthesis of different subset of structures that have different biological roles. Nevertheless, the fact that C2GnT single- and triple-deficiencies in mice demonstrated relatively mild phenotypic impact may reflect the apparent extensive compensation provided by the induction of elongated core 1 and, unexpectedly, the O-mannosyl glycans carrying the biomedically relevant epitopes.
Chapter 4:

O-glycosylation profiling of gastric mucosa from FuT2-null mice in relation to *Helicobacter pylori* adhesion
4 O-glycosylation profiling of gastric mucosa from FuT2-null mice in relation to *Helicobacter pylori* adhesion

4.1 Background

Glycoconjugates expressed on gastric mucosa play a crucial role in host–pathogen interactions, as has been described in Chapter 1. In this study, O-glycomic analyses were carried out on the stomach and mucosal scrapings (isolated gastric mucosal layer) of wild type and FuT2-null mice, which represent the secretor and non-secretor models, respectively. The purpose of analysing the mucosal scrapings is because it constitutes the mucosal layer of the stomach surface where the vital initial adhesion by *H. pylori* occurs. Lectin and antibody staining and *H. pylori* adhesion studies were also performed on gastric mucosa of both mouse models by our collaborators, Celso A. Reis and colleagues at the University of Porto, Portugal.

4.2 Results

4.2.1 Mass spectrometric strategy

All samples that were analysed in this study are listed in Table 4.1. Stomach and mucosal scrapings samples were prepared for mass spectrometric analysis similar to that of stomach tissues described in Chapter 3, which used the homogenisation with glycolipid removal technique and without N-glycan digestion prior to O-glycan release. Immunoprecipitated samples prepared by our collaborators were first in-gel-tryptic digested prior to O-glycan release and analysis. O-Glycomic profiles of samples from both wild-type and Fut2-null mice were acquired by MALDI-TOF-MS, and components of interest were sequenced by collision induced tandem mass spectrometry using both MALDI and ESI methods, as explained in the previous chapter.

<table>
<thead>
<tr>
<th>Mouse genotype</th>
<th>Sample type</th>
</tr>
</thead>
<tbody>
<tr>
<td>FuT2 knockout</td>
<td>Stomach (whole)</td>
</tr>
<tr>
<td></td>
<td>Stomach mucosal scraping</td>
</tr>
<tr>
<td></td>
<td>Stomach lysates immunoprecipitated using anti-Lea&lt;sup&gt;a&lt;/sup&gt; antibody (SPM279)</td>
</tr>
<tr>
<td></td>
<td>Stomach lysates immunoprecipitated using anti-Muc5AC antibody (45M1)</td>
</tr>
<tr>
<td>Wild type</td>
<td>Stomach (whole)</td>
</tr>
<tr>
<td></td>
<td>Stomach mucosal scraping</td>
</tr>
<tr>
<td></td>
<td>Stomach lysates immunoprecipitated using anti-Lea&lt;sup&gt;a&lt;/sup&gt; antibody (SPM279)</td>
</tr>
<tr>
<td></td>
<td>Stomach lysates immunoprecipitated using anti-Muc5AC antibody (45M1)</td>
</tr>
</tbody>
</table>
4.2.2 Glycomic profiling of wild type stomach and mucosal scrapings

Representative MALDI-TOF-MS data of stomach and mucosal scrapings from wild type mice are shown in Figure 4.1 and Figure 4.2, respectively. Glycan structures were assigned based on their mass value, fragment ion data, and knowledge of the biosynthetic pathways of mucin-type O-glycosylation. Generally, both samples gave highly similar profiles in terms of O-glycan composition and sequences. However, as expected due to the larger sample size, more glycan peaks and structures were detected in the stomach. For example O-glycans at m/z 895, 912 and 1579 managed to be sequenced in the stomach. Furthermore, the stomach showed O-glycans with compositions up to 18 monosaccharides with m/z of 3926 (Figure 4.1). Mucosal scrapings, on the other hand, allowed glycan detection only up to m/z 3027. Due to the limited amount of samples, high mass glycans especially those above m/z 3000 were only partially sequenced and remain to be fully characterised. In both samples, the most abundant types of O-glycan are core 2 structures, many of which are branched on their core 1 arm, as noted in the stomach and colon tissues analysed in Chapter 3. The terminal epitopes are dominated by blood group H antigens (Fucα1,2Galβ1,3/4GlcNAc), for instance at m/z 1157 and 1331. No monofucosylated Lewis structures were detected. However, some of the larger glycans, for example m/z 2578, carry difucosylated antenna which could either be Le\text{b} or Le\text{a}. Some of the high mass glycans at m/z 3027 and above are multifucosylated, thus are also likely to carry more than one fucose on a single antenna. Only a few sialylated glycans were detected; m/z 895 and 1579 in the stomach, as well as m/z 1316 and 1374 in both stomach and mucosal scrapings.
Figure 4.1. The MALDI-TOF-MS O-glycomic profile of reduced and permethylated O-glycans [M+Na]^+ detected in the stomach of wild type mice.

Cartoon structures assigned were based on mass spectrometric data and knowledge of O-glycan biosynthetic pathways. X, impurities/contaminant peaks.
4.2.3 Glycomic profiling of FuT2-null stomach and mucosal scrapings

As expected, the MALDI-TOF-MS profiles of stomach and mucosal scrapings from FuT2-null mice showed a major loss of glycans carrying the H epitope (Figure 4.3 and Figure 4.4). For example, glycans at $m/z$ 1331, 1535 and 1780 were detected at high abundance in the wild type but were missing in the knockout tissues. However, it was possible to observe minor fucosylated terminal galactose, as demonstrated by $m/z$ 708, which on MS/MS analysis gave the same fragment ions (the major ions are $m/z$ 298, 433 and 520, see Figure 4.5) as the isobaric glycan found in the wild type, indicating that a residual $\alpha$1,2-fucosyltransferase activity remains in the stomach and mucosal scrapings of the FuT2-null mice. In contrast to the wild type samples, the largest O-glycan structure which could be sequenced was at $m/z$ 2617 for both stomach and mucosal scrapings of the knockout mice. Due to the massive loss of fucosylated glycans, the overall number of glycan structures detected was far less than that of wild type samples. Both FuT2-null stomach and mucosal scrapings provided almost similar numbers of structures,
however a few structures are unique in either tissue. For example, the glycan at $m/z$ 1579 was only detected in the mucosal scrapings whereas glycans at $m/z$ 953 and 1514 were only found in the stomach. Surprisingly, there were losses in sialylated glycans in both tissues. Only a single structure at $m/z$ 1579 in the mucosal scrapings was revealed to carry a sialylated Gal or sialylated Gal$\alpha1,3$Gal terminal.

### 4.2.4 Differentiation of fucosylated histo-blood group antigens.

Both stomach and mucosal scrapings samples from wild type and FuT2-null mice gave a molecular ion at $m/z$ 1157 whose composition is compatible with putative singly fucosylated Lewis or H antigens. In order to establish what terminal epitope contributes to this molecular ion, MS/MS experiments were first carried out using MALDI-TOF/TOF-MS/MS. However, due to the very low amount of this component in the FuT2-null mice, no definitive fragments were acquired (data not shown). Then, analysis was repeated on the corresponding doubly charged ion at $m/z$ 590 using ESI-QTOF-MS/MS. These experiments successfully revealed that the dominant wild type glycans carry the H epitope on either the core 1 or the core 2 arm (Figure 4.6A). In the latter case, the fragment ions are consistent with the H antigen being of the type 2 sequence in which the Gal is attached to position 4 of GlcNAc. An especially definitive signal is the fragment at $m/z$ 529 which correspond to the concomitant elimination and cleavage of two hexose residues, one of which is fucosylated. Elimination is only possible when the moiety being eliminated is linked to position 3 of a HexNAc. The absence of a fragment ion at $m/z$ 511 demonstrates that only one hexose moiety was 3-linked. However, we cannot rule out minor amounts of the type 1 sequence. Also, because of the dominant H-containing structures it is not possible to determine whether any Lewis structure is contributing to $m/z$ 1157 in the wild type.
Figure 4.3. The MALDI-TOF-MS O-glycomic profile of reduced and permethylated O-glycans [M+Na]$^+$ detected in the stomach of FuT2-null mice.
Figure 4.4. The MALDI-TOF-MS O-glycomic profile of reduced and permethylated O-glycans [M+Na]^+ detected in the mucosal scrapings of FuT2-null mice.
Figure 4.5. The MALDI-TOF/TOF-MS/MS spectrum of the molecular ion $m/z$ 708 [M+Na]$^+$ from the stomach of FuT2-null mice.

The fragmentation of the glycan at $m/z$ 708 provides evidence of a fucosylated core 1 structure with the fucose being on the Gal. *, impurities/unknown fragments.

In contrast to the wild type, the Fut2-null mice showed strong evidence for the Le$^x$ structure (Figure 4.6B), despite the relevant glycan ($m/z$ 1157 in the MALDI-TOF-MS data) being only a small fraction of a percent of the total glycan mixture. Because of the very low abundance of this component, contaminant signals are prominent below $m/z$ 500 in the ESI-QTOF-MS/MS spectrum. Fortunately, however, the region of the MS/MS spectrum which contains structurally useful fragment ions (above $m/z$ 600) is relatively free from contaminating ions. Especially relevant are the signals at $m/z$ 951 and $m/z$ 715 because they correspond to elimination of fucose and separate eliminations of fucose plus hexose, respectively. As shown on the cartoon annotations, these signals can only be derived from a Le$^x$ structure. Elimination is only possible when the moiety being eliminated is linked to the position 3 of a HexNAc, therefore Le$^a$ will not eliminate fucose, and instead should eliminate galactose. Importantly there was no detectable signal at $m/z$ 685 which would be expected for Le$^a$ (i.e. elimination of Hex from Le$^a$ concomitant with elimination of the core 1 galactose from the GalNAcitol). Nevertheless, some of the H-structure is also present because $m/z$ 747 can only be derived from a glycan which carries the fucose attached to a penultimate galactose because this ion is produced by elimination of fucose plus hexose via a single cleavage event (see cartoon annotations).

In order to put the identification of Le$^x$ structure and not Le$^a$ in the FuT2-null samples on a firmer footing, we have performed MS/MS and MS$^3$ analysis on the singly charged molecular ion at $m/z$ 1157 using MALDI-QIT-TOF-MS as described in the previous chapter. The MS/MS provided fragment ions similar to those previously acquired using ESI-QTOF-MS, thereby supporting the assignment of the Le$^x$ antigen as the most abundant terminal epitope (Figure
4.7A). However, the spectrum quality was better than the previous experiment therefore allowing the detection of low molecular weight fragment ions below \( m/z \) 520 which were undetected in the ESI-QTOF-MS/MS. Structurally useful low molecular weight fragments are \( m/z \) 486 and 433. The first fragment ion corresponds to a non-reducing terminal of unmodified LacNAc and the latter corresponds to non-reducing terminal fucose and Gal eliminated via a single cleavage. Both these fragments support the existence of the residual H antigen. To confirm the linkage position of the monosaccharide constituents of the Lewis epitope, MS\(^3\) analyses were performed on the MS\(^2\) fragment ion of \( m/z \) 660 Figure 4.7B. The fragments confirm that the fucose is mainly attached to the GlcNAc residue and not Gal, supported by the absence of the fragment at \( m/z \) 433. However, due to the very low abundance of this component, cross-ring fragment ions observed were unable to provide additional evidence that can definitely confirm or refute the minor existence of \( \text{Le}^a \) antigen.

![Figure 4.6](image)

**Figure 4.6.** The ESI-QTOF-MS/MS spectrum of glycans at \( m/z \) 590 [M+2Na]\(^{2+}\) from the stomach of (A) wild type and (B) FuT2-null mice.

In the wild type, there is only evidence of fucose being on the terminal Gal whereas in the FuT2-null stomach, the main signals are consistent with fucose being on the GlcNAc. Fragments observed are of [M+Na]\(^+\) ions. *, impurities/ unknown fragments.
Both stomach and mucosal scrapings from wild type mice also gave a molecular ion at \( m/z \) 2578, with composition of which is compatible with two H antigens and a putative difucosylated Lewis antigen. The MS/MS data for this component from MALDI-TOF/TOF gave evidence for two H antigen terminated antennae and a Le\(^3\) terminated antenna. The latter structure is supported by fragment ions showing that one antenna is difucosylated (\( m/z \) 834 and 1766) and by a convincing signal for the elimination of fucose from the molecular ion (\( m/z \) 2372), confirming that a portion of the fucose in the glycan is attached to position 3 of GlcNAc (Figure 4.8).

Figure 4.7. The MALDI-QIT-TOF analysis of the glycans at \( m/z \) 1157 \([M+Na]^+\) from the stomach of FuT2-null mice.

(A) MALDI-QIT-TOF-MS/MS analysis of the molecular ion at \( m/z \) 1157 and (B) the subsequent MS\(^3\) analysis of the fragment ion at \( m/z \) 660 (HexNAc\(_2\)Hex\(_1\)Fuc\(_1\)). The fragmentation data from both analyses supports previous assignment using ESI-QTOF-MS/MS indicating major signals for fucose being on the GlcNAc residue and minor signals for fucose being on the terminal Gal of core 1 or core 2 arm. *, impurities/unknown fragments.
Chapter 4

4.2.5 Determination of N-acetyllactosamine unit type

The class of terminal epitope is determined by the type of N-acetyllactosamine (LacNAc) chain upon which it resides. Therefore, to further support the mass spectrometric assignment of type 2 antigens in wild type and FuT2-null mice, exhaustive MS3 analysis was performed on selected glycan MS/MS fragments that contain LacNAc units, plus additional enzymatic digestion and GC-MS linkage analysis in order to define backbone linkages.

The LacNAc chain type 1 consists of the repeating unit \([\text{Gal}\beta_{1-3}\text{GlcNAc}\beta_{1-3}]\), whereas type 2 consists of \([\text{Gal}\beta_{1-4}\text{GlcNAc}\beta_{1-3}]\). As an example of the determination of LacNAc chain type, Figure 4.9 shows MALDI-QIT-TOF-MS/MS data for the glycan at \(m/z\ 1780\) (HexNAC\(_2\)Hex\(_3\)Fuc\(_2\)) of wild type stomach (A), together with MS\(^3\) spectra from collisional activation of the fragments at \(m/z\ 660\) (B). The glycan at \(m/z\ 1780\) was chosen as an exemplar because it produced an intense 660 fragment ion which is derived from glycosidic cleavages of two separate fucosylated LacNAc units, as shown in Figure 4.9A. The fragment ion \(m/z\ 660\) was in fact an excellent precursor ion to produce quality data for the subsequent MS\(^3\) experiment. The fragment ion at \(m/z\ 834\) in the MS/MS spectrum shows that a minor portion of the LacNAc unit is difucosylated. In the MS\(^3\) spectrum (Figure 4.9B), the cross-ring fragments at \(m/z\ 503\) and 586 unequivocally have proven that the Gal residue is linked to the position 4 of GlcNAc. Therefore this LacNAc unit is confirmed to be type 2. Moreover, the fucose residue was shown to be attached to Gal making this trisaccharide a blood group H type 2 antigen.

Figure 4.8. The MALDI-TOF/TOF-MS/MS spectrum of glycan at \(m/z\ 2578\) [M+Na]\(^+\) with the putative Le\(^x\) antigen from the stomach of wild type mice.

This spectrum illustrates the instrument’s capability to exhaustively characterise a minor high-mass glycan from a complex sample. *, impurities/unknown fragments.

Figure 4.9. The MALDI-TOF/TOF-MS/MS spectrum of glycan at \(m/z\ 1780\) [M+Na]\(^+\) with the putative Le\(^x\) antigen from the stomach of wild type mice.
Figure 4.9. Determination and verification of N-acetyllactosamine unit type.

(A) MALDI-QIT-TOF-MS/MS spectrum of m/z 1780 [M+Na]+ from wild type stomach shows evidence of branching on Gal of the core 1 arm (fragments m/z 506 and 1143) and the terminal epitope blood group H antigen on each of the branches (fragments m/z 660 and 1370). Evidence for a minor difucosylated LacNAc unit is shown at m/z 834. (B) MALDI-QIT-TOF-MS3 spectrum of the fragment m/z 660 from (A) shows evidence for 4-linked GlcNAc (cross-ring fragments at m/z 503 and 586) indicating LacNAc unit type 2 as well as a fucose residue linked on the Gal residue (cross-ring fragment at m/z 313) establishing the epitope as H type 2. *, impurities/unknown fragments.

In addition, enzymatic digestion by using a specific β1,4-galactosidase displayed a major reduction of glycans that are terminated with unmodified Gal in the wild type stomach compared to the undigested tissue (Figure 4.10), confirming a significant composition of Gal attached to position 4 of GlcNAc in this tissue. For example glycans at m/z 983, 1677, 1881 and 1923 in the enzyme digested glycomic profile were reduced in abundance by comparing their respective adjacent peaks, relative to the wild type profile. This finding is further supported by GC-EI-MS linkage analysis which has shown the existence of 4-linked GlcNAc in the stomach, whereas 3-linked GlcNAc was not detected (data not shown). Taken together our MS3, enzymatic digestion and GC-MS linkage data support the absence of type 1 LacNAc sequences in the stomach. This conclusion can be extended to type 1 histo-blood group antigens.
Figure 4.10. The O-glycomic profile of permethylated glycans from the stomach of wild type mice before and after β1,4-galactosidase digestion.

MALDI-TOF-MS O-glycomic profile of (A) undigested stomach of wild type mice and (B) β1,4-galactosidase digested stomach of wild type mice. Peaks with reduced abundances relative to the undigested tissue are highlighted in yellow.

4.2.6 Glycomic analysis of immunopurified stomach lysates

In order to increase the possibility of detecting putative O-glycans carrying the Leα antigen in the mouse, stomach lysates were concentrated by immunoprecipitation using antibodies against Leα and Muc5AC. The latter is the most probable carrier of Leα in human stomach (De Bolós et al., 1995). The samples were then separated using gel electrophoresis and identified using Western blotting against the same antibodies. Our collaborators supplied several excised gel bands from this experiment for mass spectrometric analysis. We included several bands of murine IgA standard from electrophoresis gel under reducing condition as control for the in-gel tryptic digestion and the subsequent O-glycomic analysis. Unfortunately, no glycan structures were detected in any of the immunoprecipitated samples (data not shown). However, the experimental control using murine IgA electrophoresis bands gave positive results (Figure 4.11), proving that the extraction of glycopeptides from gel and the O-glycomic analysis were in fact reliable.
Chapter 4

Figure 4.11. The MALDI-TOF-MS O-glycomic profile of reduced and permethylated O-glycans [M+Na]$^+$ detected in the experimental control murine IgA bands.

This indicates the reliability of the in-gel tryptic digestion and the subsequent O-glycomic experiments. X, impurities/contaminant peaks.

4.2.7 Discussion

This study aimed to perform O-glycosylation profiling of secretor and non-secretor mouse models and to characterise murine histo-blood group antigens in relation to the adhesion of *H. pylori*. Specific objectives of this study were: 1) to evaluate the glycosylation changes of the stomach and mucosal scrapings of FuT2-null mice, 2) to differentiate Lewis antigen type 1 and type 2 structures and determine the most abundant type in the gastric mucosa and 3) to associate the changes in glycosylation with immunohistochemistry findings and the binding capability of various strains of *H. pylori*. To this end O-glycomic analyses were performed on stomach and stomach mucosal scrapings of wild type and FuT2-null mice. This work was done in collaboration with Celso A. Reis group from the University of Porto, Portugal, who have produced the samples and also performed phenotypic experiments. Mass spectrometric analysis for this type of tissue was optimised while analysing wild type and C2GnT knockout mice in the previous chapter.

O-Glycomic profiles of wild type stomach and mucosal scrapings were found to be dominated with glycans terminated with the H antigen. For all monofucosylated glycans, the MS/MS data are fully consistent with the fucose being on the terminal Gal in the vast majority of the glycans. This finding corroborates with previous analysis on cervical mucins using LC-ESI-MS that also suggested the H antigen as the most abundant terminal epitope; however only glycan compositions, not sequences, were presented in the earlier study (Domino et al., 2009). Nevertheless, we cannot rule out the possibility of very minor components having fucose on GlcNAc but, if present, they are below the sensitivity limit for MS/MS detection on our
instruments. A similar theme pertains for the more heavily fucosylated glycans. If multiple antennae are available, the fucoses prefer to be on separate antennae, for example the MS/MS data for the glycan at m/z 1780 (Figure 4.9) indicate that the vast majority of the fucoses are attached to Gal of each antenna. Some of the larger glycans, however, have more fucoses than can be attributed to H antigen and therefore carry difucosylated antennae. For example, the MS/MS data for the glycan at m/z 2578 (Figure 4.8) confirms that one antenna is difucosylated and there is also a convincing signal for the elimination of fucose from the molecular ion, confirming that a portion of the fucose in the glycan is attached to position 3 of GlcNAc, indicating the structure Le\(^\alpha\). However, we cannot rule out Le\(^\beta\) as a minor component but on the other hand we have no evidence for its presence. In summary, the MS/MS data have not yet yielded evidence for Le antigens type 1. Importantly the quality of the MS/MS data is generally excellent and it is expected to detect minor isomers if they are present at the level of 5-10% of their isobaric relatives, which is a very small fraction of the overall glycan mixture.

Glycosylation profiles of the stomach and mucosal scrapings from FuT2-null mice were compared to wild type mice. As expected, FuT2-null stomach and mucosal scrapings suffered a major loss of H antigens which were the most abundant terminal epitope in the wild type. Except for the disappearance of most of the fucosylated components, other O-glycan structures were almost identical to the wild type. This finding is supported by hematoxylin and eosin staining of gastric mucosa indicating no histological differences between wild type and FuT2-null mice (Magalhaes et al., 2009). Generally there was very low sialylation in these tissues. O-Glycomic analysis showed that sialylation in the knockout tissues was further reduced, with the absence of most of the sialylated glycans detected in the wild type tissues. This is unexpected since the loss of the dominant H antigen terminated glycans should allow better detection of these minor sialylated glycans. Furthermore, the deficiency of α1,2-fucosyltransferase is expected to result in less competition for sialyltransferases to modify the terminal Gal residues. In contrast to the glycomics data, a minor increase in staining with Maackia amurensis lectin-I (MAL-I), which recognises Siaα2,3Galβ1,4GlcNAc or Galβ1,4GlcNAc structure, was observed in the FuT2-null gastric mucosa (Magalhaes et al., 2009). However, the binding of the Siaα2,6Gal/GalNAc-specific Sambucus nigra (SNA) lectin was similar in both types of mice indicating no differences in α2,6-sialylation. A possible explanation for the discrepancy between the MAL-1 and glycomics results is that there could be improved accessibility to the minor sialylated glycans by the MAL-1 lectin in the absence of heavily fucosylated glycans. In any event it is clear that the absence of terminal Fuc does not result in enhanced sialylation.
The immense loss of fucosylated glycans in FuT2-null mice is supported by the complete loss of binding of the α1,2-fucose-specific lectin ulex europaeus (UAE-I) to FuT2-null mice (Magalhaes et al., 2009). However, based on the MS data on the knockout tissues, there were two residual glycan molecular ions, \( m/z \) 708 and 1157, whose compositions are consistent with monofucosylated structures. Surprisingly, minor fucosylation on terminal Gal could be detected in the FuT2-null mice as demonstrated by the molecular ion at \( m/z \) 708 and the fragment ions at \( m/z \) 747 and 433 (Figure 4.5). Furthermore, this sequence is the only biosynthetically possible sequence for the O-glycan composition \( \text{Fuc}_1\text{Hex}_1\text{HexNAc}_1 \). This remaining α1,2-fucosylation could be derived from the activity of FuT1, the only other α1,2-fucosyltransferase known in mouse. Murine FuT1 RNA was shown to be expressed in stomach and this enzyme can utilise both type 1 and type 2 LacNAc as fucosylation substrate (Domino et al., 1997). However, the previous work also suggested that FuT1 is not the main α1,2-fucosyltransferase in murine stomach.

The absence of significant levels of H-containing glycans in the knockout meant that the presence of putative Lewis structures could be rigorously confirmed by MS/MS analysis. This was accomplished by using ESI-QTOF and MALDI-QIT-TOF techniques. The molecular ion \( m/z \) 1157 in both FuT2-null stomach and mucosal scrapings is very minor indeed. Nevertheless high quality ESI-QTOF-MS/MS data were acquired by selecting a small window centred on the doubly charged molecular ion (\( m/z \) 590). The data showed evidence for a Lewis sequence being at least as abundant as the H-sequence. Especially important diagnostic signals are \( m/z \) 951 and 715 because they correspond to the elimination of fucose, and separate eliminations of fucose plus hexose, respectively. Therefore, it can be concluded that \( \text{Le}^x \) is readily detectable using MS/MS, despite the relevant glycan (\( m/z \) 1157) being only a tiny fraction of a percent of the total glycan mixture, yet no ions characteristic of \( \text{Le}^a \) were observed in the MS/MS data. This result was further confirmed by MS/MS and MS\(^3\) analyses using MALDI-QIT-TOF.

The detection of a \( \text{Le}^x \) containing glycan in the O-glycomic spectra of FuT2-null mice, but not in the wild type, is likely to be an experimental consequence of the disappearance of interfering signals in the former, rather than being due to the augmentation of α1,3-fucosylation. Therefore, it is anticipated that \( \text{Le}^x \) was also present in the wild type tissues, but below the detection levels of current structural strategies. This is supported by the similar immunoreactivity between gastric mucosa of wild type and FuT2-null mice against anti-\( \text{Le}^x \) antibody SH1 (Magalhaes et al., 2009). Surprisingly, antibody staining has also indicated the existence of
Le antigens type 1, Le\textsuperscript{a} and Le\textsuperscript{b}, in the wild type mice. This is not expected since mice lack an orthologue of the human FUT3 gene and no α1,4-fucosyltransferase has so far been described in mice (Nairn et al., 2008). Gastric mucosa staining using Ca3F4 and 7LE (anti-Le\textsuperscript{a}) antibodies indicate that there was an increase in Le\textsuperscript{a} expression in Fut2-null mice compared to the wild type, suggesting that the lack of α1,2-fucosyltransferase activity has led to the accumulation of Le\textsuperscript{a} structures. However, rigorous mass spectrometric analysis has provided no structural evidence for that epitope in either mouse. There is a possibility of antibody cross-reactivity, although the 7LE antibody was shown to display excellent specificity in carbohydrate microarray experiments in a previous study (Manimala et al., 2007).

Complementing their anti-Le\textsuperscript{a} results, the Reis team observed that two anti-Le\textsuperscript{b} antibodies, BG6 and 2.25LE, bound to the mucous cells of wild type gastric mucosa but not to the Fut2-null mice. This contrasted with our MS experiments which identified only its isobaric type 2 structure, Le\textsuperscript{y}, in the wild type mice. One possible interpretation for the anti-Le\textsuperscript{b} staining is antibody cross-reactivity. It has been demonstrated previously that BG6 cross-reacted significantly with LacNAc unit type 1, Le\textsuperscript{y} and H antigen type1 (Monteiro et al., 1998; Manimala et al., 2007) whereas the 2.25LE antibody displayed weak cross-reactivity with Le\textsuperscript{a} (Good et al., 1992; Manimala et al., 2007) although the latter would not explain the loss of binding to Fut2-null mice tissue sections. However, there was also a reduction in the anti-Le\textsuperscript{y} antibody AH6 staining in Fut2-null gastric mucosa, and together with the loss of anti-Le\textsuperscript{b} antibodies staining, this is consistent with mass spectrometric data as there were no detectable difucosylated glycan structures in Fut2-null tissues. These findings support the classical model of the Le antigen biosynthetic pathway involving FuT2 enzyme, in which Le\textsuperscript{b} and Le\textsuperscript{y} are synthesised from H antigens (Ma et al., 2006). In the context of the current work where only type 2 histo-blood group antigens were characterised, it can be summarised that FuT2 is critical in the synthesis of H antigens and Le\textsuperscript{y} in murine gastric.

We complemented the structural evidence for fucosylated type 2 antigens by determining the type of LacNAc chains in wild type stomach. Rigorous MS\textsuperscript{3} experiments accompanied by enzymatic digestion and GC-MS linkage analysis have collectively provided evidence that substantial amounts of terminal Gal are 4-linked to GlcNAc. Therefore, it can be concluded that this tissue is dominated with type 2 LacNAc chains, thus supporting the existence of type 2, but not type 1, Le antigens in the murine stomach. Although the same strategies were not applied to Fut2-null mice due to sample constraints, MS/MS and MS\textsuperscript{3} experiments on these tissues (for
example MS/MS on m/z 1157 and MS³ on m/z 660 in Figure 4.7) also suggested type 2 LacNAc chains.

It was evident that there were no significant structural differences between the whole stomach and mucosal scrapings in both wild type and knockout mice, other than slight discrepancies in the relative abundances of limited glycan structures observed in the wild type tissues. Therefore, it could be postulated that the mucosal barrier is the most highly glycosylated component of the stomach and occupied most of the O-glycan population in the whole organ.

The purpose of immunoprecipitating stomach lysates using antibodies against Le⁹ and Muc5AC was to concentrate the protein carrier of Le⁹ terminated glycans. This optimistically should aid mass spectrometric characterisation of the speculated minute amount of Le⁹ antigens in the stomach of wild type and FuT2-null mice. However, no glycan structures were detected in any immunoprecipitated samples although experimental controls, the IgA bands, gave positive results. The negative findings are probably due to the non-specific antibody binding during the immunoprecipitation and/or immunoblotting thus giving false staining on the blotted membrane. A similar strategy was applied previously and successfully isolated and analysed O-glycosylation of human MUC1 from a number of cancer cell lines (Backstrom et al., 2008). Therefore it is anticipated that improvements on the application of this technique should yield better Le⁹ antigen-concentrated samples for future studies.

In summary, our mass spectrometric characterisation of histo-blood group antigens has established that only type 2 Lewis structures are detectable in murine gastric mucins, consistent with murine genetic incapability to express Lewis type 1 structures (Gersten et al., 1995). The discrepancies observed between the immunohistochemical staining data and mass spectrometric structural characterisation regarding the detection of Le type 1 and type 2 epitopes in wild type tissues, as aforementioned, could be due to antibody cross-reactivity. In addition to that, the stomach is dominated with branched O-glycans (core 2 branch and I-branch), which are often terminally monofucosylated. The core 2 branch and I-branch are 6-linked on GalNAc and Gal, respectively, hence forming a more flexible conformation. It is tempting to speculate that O-glycans having monofucosylated branches (for example glycans at m/z 1331, 1780, 2026 and 2404 in the wild type tissues, see Figure 4.1 and Figure 4.2) could have similar spatial conformation as a difucosylated antenna. Therefore, those glycan structures could be perceived as having Le³⁵³ by antibodies. Hence, caution should be taken when drawing conclusions based only
on binding assays. Mass spectrometry, on the other hand, is a more reliable and definitive method for evaluating the expression of carbohydrate structures from complex samples.

If antibody cross-reactivity is not the correct explanation for the aforementioned discrepancies, the absence of type 1 Lewis antigens in the O-glycomic profiles could perhaps be explained by a higher sensitivity of immunohistochemistry compared with mass spectrometry. As indicated by anti-Lewis antigen antibodies, these structures are only expressed on specific stomach sections (Magalhaes et al., 2009). Le\textsuperscript{a} antigens were indicated at surface mucous cells whereas Le\textsuperscript{b} and Le\textsuperscript{d} were found at superficial mucous cells of the foveolar epithelium. Le\textsuperscript{x} expression, on the other hand, was slightly more widespread at the surface mucous cells and neck cells. These findings are supported by the section-selective binding of BabA and/or SabA positive \textit{H. pylori} strains (see below). Therefore, mass spectrometric analysis of the whole stomach and mucosa scrapings could easily swamp low abundance structures expressed on a limited number of cells. Mass spectrometric analysis of specific stomach sections should be performed in future studies, provided advances in technology yield the required sensitivity, and this optimistically would be a good alternative to immunoprecipitation of stomach lysates to concentrate specific epitopes.

\textit{H. pylori} has been suggested to bind fucosylated histo-blood group antigens Le\textsuperscript{b} and H type 1 in humans mediated by BabA adhesin (Ilver et al., 1998; Bovenkamp et al., 2003). This adhesin is encoded by the gene babA2 (Gerhard et al., 1999). Another important adhesin is SaBA which is encoded by JHP662 gene and binds to sialylated glycans including SL\textsuperscript{a}e and SL\textsuperscript{x}e (Mahdavi et al., 2002). In the current study, a panel of BabA and/or SabA positive and negative model strains of \textit{H. pylori} have been evaluated for their binding capacity on gastric mucosa of wild type and FuT2-null mice (Magalhaes et al., 2009). \textit{H. pylori} strain J99 expresses both BabA and SabA active adhesins with known bacterial binding to antigens including Le\textsuperscript{b}, H type 1, SL\textsuperscript{x}e, SL\textsuperscript{a}e and sialyl-lactosamine (Aspholm-Hurtig et al., 2004; Aspholm et al., 2006). Our collaborators have shown that this strain bound to the surface mucous cells and glands of both wild type and FuT2-null mice. However, the binding to the surface mucous cells was decreased in the FuT2-null mice. In order to determine the involvement of α1,2-fucosylated glycans to the observed decrease, the adhesion of 17875/Le\textsuperscript{b} strain was evaluated (Magalhaes et al., 2009). This strain expresses both BabA and SabA adhesins but lacks binding capability to sialylated glycans (Aspholm et al., 2006). Binding of this strain was restricted to the surface mucous cells of wild type mice and completely absent in the null mice, indicating the importance of fucosylated
glycans in BabA-mediated adherence. In addition to that, a BabA mutant strain 17875babA1A2 which only binds to sialylated glycans has been tested. As expected, this strain bound similarly in both wild type and FuT2-null mice to the faveolar and surface mucous cells, thus illustrating the fucosylation-independence binding in BabA-negative strains. Interestingly, this strain also adhered more to the glandular region of FuT2-null mice than of the wild type mice.

In addition, the binding capacity of a larger panel of \textit{H. pylori} strains from clinical isolates with different active adhesins were tested on wild type and FuT2-null mice and the results in general supported earlier findings of \textit{H. pylori} model strains. Generally, strains with active BabA adhesin have reduced binding to FuT2-null mice, with the presence or absence of active SabA, whereas SabA active strains and strains without any adhesin showed similar binding to both mice. The FuT2-null mouse model resembles the human non-secretor phenotype which has been demonstrated to have similar reduced susceptibility to \textit{H. pylori} infection compared to the wild type (secretor phenotype) (Ikehara \textit{et al.}, 2001; Azevedo \textit{et al.}, 2008). Collectively, based on all the adhesion studies and our mass spectrometric structural data, it can be concluded that $\alpha 1,2$-fucosylated glycans, particularly those decorated with H type 2 and/or Le$^y$, play important roles in the adhesion of \textit{H. pylori} BabA-positive strains and therefore can critically influence the bacterial pathogenicity. Furthermore, a recent finding has shown that both BabA and SabA adhesins play substantially bigger roles in inducing cell death than the proinflammatory factor CagA (Lindén \textit{et al.}, 2009). It is also worth highlighting that greater than 80\% of known \textit{H. pylori} strains were shown to express type 2 Lewis antigens aimed to mimic the gastric epithelial surface in order to evade the host immune system (Wang \textit{et al.}, 2000), thus suggesting the significance of the type 2 carbohydrate structures at \textit{H. pylori}'s colonisation sites.

As mentioned in Chapter 1, persistent colonisation by \textit{H. pylori} results in inflammation and has been shown to promote synthesis of sialylated glycoconjugates purposely to guide leukocyte migration (Mahdavi \textit{et al.}, 2002). Previously, \textit{H. pylori} mutant strain babaA1A2 (BabA negative, SabA positive) was demonstrated by Mahdavi \textit{et al.} to lack binding to healthy gastric mucosa but not to \textit{H. pylori}-infected (inflamed) tissues (Mahdavi \textit{et al.}, 2002). It is important to highlight here that \textit{H. pylori} adhesion seems to promote substantial changes to the glycosylation profile of gastric mucosa. To date, rigorous structural characterisation of \textit{H. pylori} infected tissues in comparison with the healthy ones has not been accomplished. Therefore, it is not known whether \textit{H. pylori} infection induced glycosylation changes other than sialylation, for instance downregulation of fucosylated glycans, and if these changes are also influenced by the secretor
status. As an example, Holmen et al. previously have demonstrated transient induction of FuT2 concomitant with the increase of oligosaccharides containing H antigen during the infectious cycle of parasite *Nippostrongylus brasiliensis* (Holmen et al., 2002). Furthermore, as a long term and common inhabitant of human gastric tissue since birth, it is also not impossible for *H. pylori* to be involved in gut development, as has been suggested by Hooper and colleagues for other microbes (Hooper, 2004). These workers found that during the weaning stage, human intestinal microflora changed from being dominated by *Escherichia coli* and *Streptococci* to *Bacteroides* and *Clostridium* species, accompanied by marked functional and morphological maturation of the gut. It is similarly important for future work to also establish the type of sialylated glycans that are utilised by *H. pylori* for adhesion. SLe\(^\beta\), sialyl-Tn, sialylated terminal Gal, sialylated GlcNAc and Sda\(^a\) are among the possible sialic acid containing structures in mice and, in contrast to humans, could include either NeuAc or NeuGe.

In summary, we have demonstrated a major loss of fucosylated O-glycans in the FuT2-null murine gastric mucosa. Extensive glycan sequencing has proven that type 2 histo-blood group antigens, especially the H antigen and at a lower extent the Le\(^\gamma\), dominate the mucins of murine gastric. Also our data and that of others have clearly demonstrated that *H. pylori* binds to fucosylated glycan receptors on gastric mucosa, and later also to the sialylated counterparts, in order to initiate gastric colonisation on the epithelial layer. The adherence to the cell surface constitutes the crucial step for a successful *H. pylori* infection, as deficiency in α1,2-fucosylated glycans severely affected the BabA-positive strains.
Chapter 5:

Investigation of cellular regulation of protein O-glycosylation by Src
Chapter 5

5 Investigation of cellular regulation of protein O-glycosylation by Src

5.1 Background

Previous findings and recent revelations by our collaborators (Gill et al., 2010) have led to the postulation that activated Src regulates mucin-type protein O-glycosylation through redistribution of polypeptide-N-acetylgalactosaminyltransferases (ppGalNAcTs) from the Golgi to the ER. The preliminary O-glycomic analyses described in Chapter 1 (Figure 1.13) have demonstrated a reduction of core 2 O-glycans relative to core 1 in Src activated cells which could be the consequence of higher O-glycosylation density, but were done only on a single murine fibroblast (SYF) cell line. Therefore, the work in this chapter aimed to determine whether the observed glycomic alterations were a cell-specific response or were indicative of a more general phenomenon by extending the semi-quantitative analysis to two different cell lines. Thus O-glycomic analysis was performed on two samples of NIH3T3 Src murine fibroblast cell lines, namely 3T3 Src wild type (3T3 WT) and 3T3 Src-vSrc (3T3 vSrc). Analyses were also performed on four samples of NBT-II rat epithelial cell lines, namely NBT-II wild type (NBT-II WT), NBT-II with active exogenous ppGalNAcT-2 in the Golgi (NBT-II WT Golgi), NBT-II with active exogenous ppGalNAcT-2 in the ER (NBT-II WT ER) and NBT-II with inactive exogenous ppGalNAcT-2 in the ER (NBT-II H2 ER). The differences between the cell lines and samples are described below. This work was done in collaboration with Frederic Bard and colleagues from the National University of Singapore.

5.2 Results

5.2.1 Mass spectrometric strategy

All samples that were analysed in this study are listed in Table 5.1. Cells were disrupted via sonication in lysis buffer followed by reduction/carboxymethylation and tryptic digestion as described in Chapter 2. In the first set of experiments for NIH3T3 cells, samples were subjected to O-glycan release without prior N-glycan removal. In the second experiment on NIH3T3 cells and all experiments involving NBT-II cells, samples were first treated with PNGase F. Additionally, in the second experiment for NBT-II cells, samples were also subjected to enzymatic and chemical treatments prior to mass spectrometric analysis. Mass spectrometric profiling and sequencing of permethylated O-glycans were carried out using MALDI-TOF/TOF-MS and MS/MS, respectively, in the positive ion mode.
### Table 5.1. List of cell line samples analysed and their descriptions.

<table>
<thead>
<tr>
<th>Cell lines</th>
<th>Sample names</th>
<th>Sample descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH3T3 Src mouse fibroblast</td>
<td>3T3 WT (1) &amp; (2)</td>
<td>NIH3T3 wild type with normal level of Src</td>
</tr>
<tr>
<td></td>
<td>3T3 vSrc (1) &amp; (2)</td>
<td>NIH3T3 transformed with viral Src with about 2.5 fold elevated Src activity than WT</td>
</tr>
<tr>
<td>NBT-II rat epithelial cells</td>
<td>NBT-II WT (1)</td>
<td>NBT-II wild type</td>
</tr>
<tr>
<td></td>
<td>NBT-II WT Golgi (1) &amp; (2)</td>
<td>NBT-II stably expressing full-length ppGalNAcT-2 with a C-terminal mCherry* tag. Exogenous active ppGalNAcT-2 present in the Golgi</td>
</tr>
<tr>
<td></td>
<td>NBT-II WT ER (1) &amp; (2)</td>
<td>NBT-II stably expressing ppGalNAcT-2 fused to the N-terminus of MHC class II with a C-terminal mCherry tag. Exogenous active ppGalNAcT-2 present in the ER.</td>
</tr>
<tr>
<td></td>
<td>NBT-II H2 ER (1) &amp; (2)</td>
<td>NBT-II stably expressing a mutant ppGalNAcT-2 fused to the N-terminus of MHC class II with a C-terminal mCherry tag. The mutation H226D of ppGalNAcT-2 inactivates the enzyme due to inability to bind an essential Mn^{2+} co-factor ion. Exogenous inactive ppGalNAcT-2 present in the ER.</td>
</tr>
</tbody>
</table>

*mCherry is a gene encoding a red fluorescent protein utilised as a protein tag to aid live cell imaging, detection and isolation.

### 5.2.2 O-Glycomic analysis of NIH3T3 cells without N-glycan digestion

Direct removal of O-glycans without prior digestion of N-glycans, as observed while optimising the methodology for the work in the previous chapters, usually gives the best O-glycan data most likely due to shorter processing steps thus allowing better sample recovery. Therefore, it was sensible for the first set of NIH3T3 cells to be analysed according to this strategy. MALDI-TOF mass spectra of permethylated O-glycans as singly charged sodiated molecular ions from 3T3 WT (1) and 3T3 vSrc (1) are shown in Figure 5.1 and Figure 5.2, respectively. All detected O-glycan structures and their assignments are collated in Table 5.2, including those from the second set of experiments (see next section). Data for both samples showed a mixture of N- and O-glycans because a portion of the N-glycans was hydrolysed from Asn under the alkaline conditions of the reductive elimination. The spectra are dominated by reduced (labelled N in figures) and non-reduced (labelled nN) N-glycans in addition to chemical artefacts (labelled X). This unique pattern of N-glycans allows their discrimination from the O-glycans even without MS/MS analysis. In addition, the N-glycan peaks can be regarded as
internal controls indicating the experimental measures were in fact working fine and similar in both samples.

![MALDI-TOF-MS spectra of O-glycans from 3T3 WT (1) cells without N-glycan removal](image)

The glycomic profile of reduced and permethylated O-glycans [M+Na]+ detected in 3T3 WT cells. N, reduced N-glycans; nN, non-reduced N-glycans; X, impurities.

Generally, both samples exhibited a very low number and abundance of core 1 and core 2 O-glycans and therefore would not be an ideal set of samples for the comparison of the O-glycan core ratio. Only three O-glycan structures were detected in 3T3 WT (1) (m/z 534, 1187 and 1548) (Table 5.2). Similar molecular ions were observed in 3T3 vSrc (1) except for m/z 534, which was not detected. It was speculated that the high abundance of N-glycans might have interfered with the reductive elimination step resulting in low O-glycan release, thereby cancelling out the aforementioned advantage of having N-glycan peaks as internal controls for assessing the quality
of O-glycan data. Therefore experiments were repeated on another set of the samples with O-glycan release being carried out after N-glycan digestion.

Figure 5.2. MALDI-TOF-MS spectra of O-glycans from 3T3 vSrc (1) cells without N-glycan removal. The glycomic profile of reduced and permethylated O-glycans [M+Na]^+ detected in 3T3 vSrc cells. N, reduced N-glycans; nN, non-reduced N-glycans; X, impurities.
5.2.3 O-Glycomic analysis of NIH3T3 cells with N-glycan digestion

MALDI-TOF mass spectra of permethylated O-glycans as singly charged sodiated molecular ions from 3T3 WT (2) and 3T3 vSrc (2) with N-glycan removal prior to reductive elimination are shown in Figure 5.3 and Figure 5.4, respectively. The spectra are relatively clear from N-glycans and other contaminants compared to the first set of data. However, the number of detected O-glycan structures has not been much improved, as listed in Table 5.2. Therefore, taking together the findings of both sets of experiments, it can be concluded that 3T3 cells are naturally low in O-glycosylation. However, interestingly, it appears that this second set of 3T3 cell data exhibited additional sialylated structures that were absent in the first set, i.e. at m/z 895 and 1256 (Table 5.2).

In 3T3 WT (2), O-glycans of both core1 (m/z 534 and 895) and core 2 (m/z 1187 and 1548) were observed. The Src-induced cells, 3T3 vSrc (2), unexpectedly exhibited a very similar profile to the wild type cells but with a slight increase in sialylated structures. This was demonstrated by the higher abundance of the glycan at m/z 895 and the emergence of a disialylated core 1 O-glycan at m/z 1256. It is important to note that in both 3T3 WT (2) and 3T3 vSrc (2), core 1 O-glycan structures are mostly sialylated whereas core 2 structures are predominantly capped with Galα1,3Gal alone or both Galα1,3Gal and sialic acid (on separate

<table>
<thead>
<tr>
<th>m/z</th>
<th>Structures</th>
<th>Assignments</th>
<th>First set</th>
<th>Second set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>WT (1)</td>
<td>vSrc (1)</td>
</tr>
<tr>
<td>534</td>
<td></td>
<td>Core 1</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td>895</td>
<td></td>
<td>Sialylated core 1</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>1187</td>
<td></td>
<td>Core 2 capped with Galα1,3Gal</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>1256</td>
<td></td>
<td>Disialylated core 1</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>1548</td>
<td></td>
<td>Core 2 capped with Galα1,3Gal and sialic acid</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

☑️, detected; ✗, not detected.

Table 5.2. Summary table of O-glycan structures observed in both sets of experiments of NIH3T3 cell line.
Due to the very limited number of structures observed and the lack of significant differences in term of peak abundances, we were unable to make any firm conclusions concerning changes in the core 2/core 1 ratio.

**Figure 5.3. MALDI-TOF-MS spectra of O-glycans from 3T3 WT (2) cells with N-glycan removal**

The glycomic profile of reduced and permethylated O-glycans [M+Na]+ detected in 3T3 WT cells. X, impurities.
Figure 5.4. MALDI-TOF-MS spectra of O-glycans from 3T3 vSrc (2) cells with N-glycan removal

The glycomic profile of reduced and permethylated O-glycans [M+Na]+ detected in 3T3 vSrc cells. X, impurities.

5.2.4 O-Glycomic analysis of NBT-II cell lines

Since there were no useful differences observed between the O-glycomic profiles of normal and induced Src in the NIH3T3 cells, our collaborators designed experiments aimed at investigating the effect of the relocalisation of ppGalNAcT from the Golgi apparatus to the ER in NBT-II cells. The cells they provided for O-glycomic analysis are described in Table 5.1. The O-glycomic profiles of NBT-II WT (1), NBT-II WT Golgi (1), NBT-II WT ER (1) and NBT-II H2 ER (1) are shown in Figure 5.5, Figure 5.6, Figure 5.7 and Figure 5.8, respectively. These cells displayed a complex repertoire of core 1 and core 2 O-glycans which are variously capped with fucose, Galα1,3Gal and sialic acid. Generally, all samples exhibited similar glycan compositions
and sequences except for the molecular ion at \( m/z \) 1505, whose composition is compatible with putative core 2 O-glycans carrying fucosylated Gal and Le\( ^{x/y} \) epitopes, which was absent in NBT-II WT.

**Figure 5.5. MALDI-TOF-MS spectra of O-glycans from NBT-II WT (1) cells**

The glycomic profile of reduced and permethylated O-glycans [M+Na]\(^+\) detected in NBT-II WT cells. X, impurities; U, underpermethylations.
Figure 5.6. MALDI-TOF-MS spectra of O-glycans from NBT-II WT Golgi (1) cells

Figure 5.7. MALDI-TOF-MS spectra of O-glycans from NBT-II WT ER (1) cells

The glycomic profile of reduced and permethylated O-glycans $[\text{M+Na}^+]$ detected in NBT-II WT ER cells. X, impurities; U, underpermethylations.
Figure 5.8. MALDI-TOF-MS spectra of O-glycans from NBT-II H2 ER (1) cells

The glycomic profile of reduced and permethylated O-glycans [M+Na]⁺ detected in NBT-II H2 ER cells. X, impurities; U, underpermethylations.

Major peaks in the m/z range of 700 to 1730 from all spectra are collated in a single panel to assist semi-quantitative analysis (Figure 5.9). NBT-II WT (1) has the highest abundance of core 1 O-glycans (taking into account the abundance of m/z 708 and 912) compared to the other 3 cell lines. All these cells exhibited a similar level of the basic core 2 structure (m/z 983) but varied in the levels of elongated core 2 structures. By comparing the abundances of the sialylated glycans at m/z 1548 to their non-sialylated counterparts at m/z 1187 in all samples, it appears that sialylation in NBT-II WT (1) and WT ER (1) is higher compared to WT Golgi (1) and H2 ER (1). The differences in terminal capping of these glycans, in addition to not being close enough in m/z
value, has complicated semi-quantitative analysis of these glycomic profiles thus preventing rigorous comparison of the core 2/core 1 ratio. Therefore, a second set of NBT-II cell line samples were subjected to chemical and enzymatic sugar degradations prior to mass spectrometric analysis aimed to trim down the structures to simple core 1 and core 2 structures, except for NBT-II WT which was no longer available.

Figure 5.9. Combined O-glycomic spectrum of NBT-II cell lines showing only major peaks.

5.2.5 Trimming NBT-II cell glycomes to basic core 1 and core 2 structures

In the second set of experiments, NBT-II WT Golgi, WT ER and H2 ER released O-glycans were treated with Sialidase A (neuraminidase that cleaves non-reducing terminal sialic acids of all linkages), α-galactosidase (cleaves all non-reducing terminal α-galactose residues) and α1,2-fucosidase (cleaves non-reducing terminal α1,2-fucose residues) concurrently prior to permethylation. The results (data not shown) suggested incomplete trimming of O-glycans, indicated by the existence of residual α1,2-fucosylated components and possibly other types of fucose linkages in the mass spectra. Therefore, these permethylated glycans were subjected to TFA hydrolysis at room temperature to remove all the remaining fucose residues. Figure 5.10 shows O-glycomic profiles of glycosidases and TFA treated NBT-II WT Golgi (2), NBT-II WT ER (2) and NBT-II H2 ER (2). It is evident that only two simple O-glycan structures were detected at m/z 534 (core 1) and m/z 983 (core 2), thus allowing the comparison of core 1 and core 2 abundances. Relative to the core 1 structure, NBT-II H2 ER (2) showed the lowest abundance of the core 2 structure (20%), followed by NBT-II WT Golgi (2) (40%). Unexpectedly, NBT-II WT ER (2) showed the highest core 2 O-glycan relative abundance (60%), although the percentage is just slightly above that of NBT-II WT Golgi (2).
Figure 5.10. MALDI-TOF-MS spectra of enzyme- and TFA-treated NBT-II cell lines. O-glycomes of (A) NBT-II WT Golgi (2), (B) NBT-II WT ER (2) and (C) NBT-II H2 ER (2) were trimmed to basic core 1 and core 2 structures to aid semi-quantitative analysis. X, impurities; N, reduced N-glycans.

It was of concern that TFA hydrolysis could have caused minor degradation of core 2 structures. Therefore, two wild type kidney tissues were used to test the stability of core 2 sequences towards TFA hydrolysis at room temperature. This organ was chosen because it is rich in fucosylated core 2 O-glycans (see Figure 3.11 in Chapter 3). O-Glycomic analysis of kidney tissues were performed according to the same strategies applied in the analysis of kidney tissues in Chapter 3. After permethylation, each sample was split into two, one for direct mass spectrometric analysis whilst the other was given prior treatment with TFA hydrolysis. Representative O-glycomic spectra of wild type kidney with and without TFA hydrolysis are
shown in Figure 5.11. The loss of fucosylated glycans leads to an increase in abundance of their respective non-fucosylated counterparts. For instance, after the removal of fucose, the glycan at $m/z$ 1157 is now detected at $m/z$ 983, contributing to the increase in intensity of that glycan peak (Figure 5.11B) compared to the control (Figure 5.11A). The loss of detection of originally non-fucosylated glycans after the treatment, for instance at $m/z$ 1187 and 1473, might indicate minor hydrolysis of non-fucose residues by TFA. However, by assessing the highly similar ratio observed between the abundances of the basic core 1 structure at $m/z$ 534 and the basic core 2 structure at $m/z$ 779 before and after the hydrolysis, it can be concluded that, even if TFA hydrolysis at room temperature cleaves residues other than fucose, the degradation was similar between core 1 and core 2 structures. Hence, the data analysis of enzyme and TFA-treated NBT-II cell samples above is valid.

Figure 5.11. MALDI-TOF-MS spectra of TFA-treated and non-treated wild type kidneys.
O-Glycomic profiles of (A) control and (B) TFA-treated permethylated O-glycan from wild type kidneys for investigation on the stability of O-glycan structures in TFA hydrolysis. X, impurities; U, underpermethylations.
5.3 Discussion

O-Glycomic profiling of two cell lines was performed in order to investigate and confirm previous findings on the postulated role of Src in the regulation of protein O-glycosylation. In a preliminary analysis done by a colleague, a marked decrease in core 2 O-glycans was demonstrated by SYF cells when active Src was introduced. In this more comprehensive thesis work, NIH3T3 cells were prepared to contain different levels of activated Src as well as NBT-II cells having different localisation of exogenous ppGalNAcT-2 between the ER and Golgi. This was done to allow assessment of the effects of viral induction of Src and ppGalNAcT relocalisation from the Golgi to the ER, respectively, on cellular O-glycomes. Particularly this study was aimed to verify whether the level of activated Src or the localisation of ppGalNAcT influences the ratio between core 1 and core 2. High quality mass spectrometric and tandem mass spectrometric data were obtained by using a MALDI-TOF/TOF mass spectrometer aided by enzymatic and chemical digestions.

The first set of 3T3 cells were analysed according to the strategy that usually gives the best O-glycan data, that is via direct reductive elimination without prior digestion of N-glycans. As well as O-glycans, this strategy also resulted in the release of some N-glycans which are observed in both reduced and non-reduced forms. Their presence can be useful in assessing the quality of the data but they can be a problem if they dominate the spectra. This protocol also seems to discriminate against small sialylated O-glycans compared to the second set of 3T3 cells that were subjected to N-glycan release prior to reductive elimination. For instance, a monosialylated ($m/z$ 895) and a disialylated core 1 O-glycan ($m/z$ 1256) were completely absent in 3T3 WT (1) and 3T3 vSrc (1) (Table 5.2). Unlike the monosialylated structure that was present in both 3T3 WT (2) and 3T3 vSrc (2), the disialylated structure was only detected and indeed gave an intense signal in 3T3 vSrc (2), the cells with induced Src. This might indicate Src’s influence on the level of sialylation. This is further supported by our finding on the NBT-II cell line that cells with overexpressed ppGalNAcT-2 in the ER demonstrated higher sialylation than cells that overexpressed the enzyme in the Golgi. Since an increase in sialylation is one of the hallmarks of cancerous cells (Brockhausen, 1999) and has been implicated to be important in metastatic process (Bresalier et al., 1996), this finding corroborates previous correlations of Src with the increase in metastatic potential of tumour cells (Boyer et al., 2002). However, the targets of Src activity that may be crucial for this process are still unknown. Based on the recent findings by our collaborators, it can be hypothesised that activated Src phosphorylates and activates
component(s) of COP-I vesicles thereby inducing the redistribution of Golgi enzymes (possibly including ppGalNAcTs relocalisation to the ER) leading to the changes in protein glycosylation, particularly alterations that benefit the progression of cancer cells. Putting together the facts that i) GEF1, a GTPase exchange factor that is involved in the formation of COP-I vesicles, and other GEFs such as Lfc and Ect2 are substrates for protein kinases (Meiri et al., 2009; Moseley & Nurse, 2009; Fields & Justilien, 2010) and ii) GEFs including Lfc and Ect2 are oncogenes (Cerione & Zheng, 1996), it is tempting to speculate that GEF1 could be the substrate for Src kinase.

As discussed earlier (Section 1.2.6 in Chapter 1), previous O-glycomic analysis by a colleague (unpublished work) has demonstrated that core 2 O-glycans, found in significant amount relative to core 1 O-glycans in cells deficient in Src (SYF), were substantially reduced in cells with induced Src (SYFsrc) (Figure 1.13). However, from the current O-glycomic analyses of 3T3 cell lines with or without prior N-glycan removal, there were no significant differences regarding the core 2/core 1 ratio observed between the wild type cells and Src-induced cells. Furthermore, without internal standards, quantitation by peak heights is unreliable in these cells since the core 1 and core 2 components are differently capped (mainly sialic acid for the former and Galα1,3Gal for the latter) and insufficiently close in m/z value to have similar ionisation properties for any valid quantitative conclusions to be drawn. Due to the limited availability of this cell line, we were unable to perform enzymatic/chemical degradation as has been carried out for the NBT-II cells. In addition, it was also evident that the 3T3 cells have a rather low abundance of O-glycans, thus they are not an ideal cell line to accomplish the aims of this work.

O-Glycomic analysis of NBT-II cells was expected to illustrate the effects of ppGalNAcT redistribution from the Golgi to ER, which has been speculated to be part of the cellular events regulated by activated Src. The localisation of ppGalNAcTs in the ER is hypothesised to increase the total O-glycosylation initiation by exposing protein substrates to ppGalNAcTs earlier and for a longer time. From our analysis, these cells appeared to be much richer in O-glycans than the NIH3T3 cells. Similar to the 3T3 cells, they exhibited a complex repertoire of core 1 and core 2 O-glycans which are variously capped with fucose, Galα1,3Gal and sialic acid. However, there were insufficient types of core 1 O-glycans to serve as internal standards to their core 2 counterparts, thus preventing direct assessment of the relative abundances of the two cores. Therefore, these cells were subjected to enzymatic and chemical degradations aimed to trim their O-glycomes to basic core 1 and core 2 structures. By treating the cells with a cocktail of
glycosidases, the sequences were partially trimmed indicating incomplete defucosylation. Further hydrolysis with TFA eliminated the remaining fucosylated components without significantly degrading other non-fucose residues. The results unexpectedly revealed that NBT-II WT ER has the highest core 2/core 1 ratio in contrast to previous findings on SYF cell lines demonstrating an increase in core 1 O-glycans parallel to Src activation. Therefore current findings are not in line with the hypothesis that relocalisation of ppGalNAcTs to the ER should produce more core 1 O-glycans due to the higher number of O-glycosylation sites occupied resulting in steric inhibition of core 2 branching of GalNAc. In addition, the dissimilarities between NBT-II H2 ER (cells with overexpressed non-functional ppGalNAcT-2) and NBT-II WT (Figure 5.9) raises the possibility that the observed O-glycomic changes in the modified cells (WT ER and WT Golgi) might be partly influenced by the disruption in the expression and/or distribution of endogenous glycosyltransferases as a result of the expression of the exogenous ppGalNAcT-2.

There are a few possible explanations of the disagreement between the NBT-II and SYF cell lines regarding the core 2/core 1 ratio in active Src conditions. One reason could be that the observed glycosylation alterations in SYF cell lines might have been a cell-type specific response to Src activation. The NBT-II cell line has not been studied for cellular organisation and ppGalNAcT distribution upon stimulation by growth hormone or Src induction, as has been demonstrated previously on SYF, WI38 and HeLa cell lines by our collaborator (Gill et al., 2010). The NBT-II cell line is derived from murine epithelia whereas SYF cells are derived from murine fibroblasts, WI38 are human fibroblasts and HeLa are human cervical cancer cells, illustrating the possibility of variations in cell morphology and physiology. Previously, an increase in Src activity in the NBT-II cell line has been shown to contribute to the enhanced metastatic potential but not to cell proliferation (Boyer et al., 2002). In contrast, Src has also been demonstrated to induce cancer cell growth and division but not metastatic potential on a number of human colon adenocarcinoma cell lines (Irby et al., 1997; Staley et al., 1997). Thus, it can be envisaged that Src acts differently in different types of cells. Accordingly, Boyer et al. have also concluded that the biological effects of Src are subjected to the cellular context, particularly to the type of cell proliferation (Boyer et al., 2002).

While the SYF cell line was shown to be very rich in α2,3/6-sialylated terminal epitopes (see Figure 1.11), NBT-II cells are dominated by the Galα1,3Gal terminal epitope. Sialylation can occur at very early stages of the O-glycan biosynthetic pathways, i.e. on the Tn and T antigens, thus completely inhibiting elongation and/or core 2 synthesis. Galα1,3Gal, on the other hand,
requires at least the T antigen and it is not known to block the synthesis of core 2 branch. Furthermore, recently it has been suggested that core 2 biosynthesis is co-regulated by the substrate peptide sequence and existing glycosylation patterns of glycoproteins (Brockhausen et al., 2009). A combination of substrate interactions with the enzyme and steric accessibility was shown to be important factors determining C2GnT1 activity (Brockhausen et al., 2009). In addition, ppGalNAcT-1 (Brockhausen et al., 1996) and C1GalT (Granovsky et al., 1994; Gerken et al., 2002) was also shown to be influenced by similar properties. Hence, the nature of peptides being synthesised and the expression of glycosyltransferases in each cell type highly influences the number of O-glycosylation sites and the O-glycome, in particular core 1 and core 2 profiles. All these observations might explain the distinctive responses of SYF and NBT-II cells towards ppGalNAcT relocalisation to the ER.

As mentioned in Chapter 1 (Section 1.2.2.2), ppGalNAcTs can bind to the Tn antigen on their protein substrates via the lectin domain, and this interaction may inhibit or enhance their glycosylating activity on the catalytic domain (Brockhausen et al., 2009). The ppGalNAcT isoform used by our collaborators in this study, ppGalNAcT-2, is possibly not the best model for the ppGalNAcT family. Even though ER-targeted ppGalNAcT-2 has been demonstrated to increase the glycosylation of a reporter protein when the enzyme is active (Gill et al., 2010), ppGalNAcT-2 has also been implicated to be inhibited by neighbouring glycosylations (Hanisch et al., 2001). The latter result is suggesting that ppGalNAcT-2 tends to keep the distance between its glycosylation sites regardless of how easily accessible they are on unfolded or folding proteins, thus the speculated steric inhibition of core 2 branching might not be applicable in this case. In contrast, other ppGalNAcTs, for instance ppGalNAcT-7 and -9, exhibit preferential addition of successive GalNAc residues next to those already present (Ten Hagen et al., 1999; Ten Hagen et al., 2001). It is evident that O-glycosylation of multisite substrates can proceed in a hierarchical manner depending on the available ppGalNAcTs in situ (Ten Hagen et al., 2003). Therefore, overexpressing ppGalNAcT-2 alone in the ER is a rather poor representation of Src mediated regulation of protein O-glycosylation.

Our collaborators have observed higher staining of Helix pomatia lectin (HPL) in cells with activated Src compared to Src-deficient cells indicating an increase in O-glycosylation initiation (Gill et al., 2010). However, HPL only recognises GalNAc residues which have been exposed (Hammarstrom et al., 1977). Our collaborators also demonstrated an increase of O-glycosylated proteins in SYFs src compared to the wild type SYF cell line and a specific increase
of O-glycosylation on a Muc1 prototypical protein when cotransfected with active Src in human fibroblast HEK293T cells, both via metabolic labelling with GalNAz (N-azidoacetyl galactosamine) and detection on Western blotting (Gill et al., 2010). GalNAz however needs to compete with endogenous GalNAc to be incorporated in O-glycosylation initiation, thus reducing the efficiency of this technique to about 30% according to Dube et al. (Dube et al., 2006), but this is likely to vary according to cell type.

Therefore, more definitive strategies are needed for instance by using reporter mucin-like glycoproteins with known sequences that will allow rigorous determination of their O-glycosylation sites. The Muc1 prototypical protein mentioned above can be used for this purpose by expressing it in cells under study. The protein would need to be digested optimally in terms of glycopeptide length and the number of glycosylation sites per digest. The glycopeptides could then be treated with glycosidases to trim the glycan structures to as simple as possible, preferably until T or Tn antigens without affecting the stability of the peptide backbones. This should be followed by glycopeptide separation using online or offline LC prior to mass spectrometric sequencing. Mass spectrometric techniques have been proven to be valuable for O-glycopeptide analysis even as early as the 1970’s when EI-MS was the main tool. For instance, the determination of O-glycosylation sites and/or site-specific glycan analysis has been accomplished on anti-freeze glycoproteins (Morris et al., 1978). Subsequently FAB-MS was shown to be a more powerful technique for example in the characterisation of interleukin-2 and erythropoietin (Robb et al., 1984; Sasaki et al., 1988). Mass spectrometric analysis of this kind of sample has become a relatively routine since then. However, this pioneering work was done on glycopeptides with only a limited number of O-glycans attached, preceded by simple fractionation using chromatography. On mucin or mucin-type glycoproteins, on the other hand, it is usually more difficult for proteolytic enzymes to digest to glycopeptides especially the tandem repeat regions due the lack of digestion sequences. Even if digestion is possible, determination of the mass or composition of the glycopeptides does not directly give indications on the number of glycosylation sites. This is because, unlike N-glycosylation, O-glycosylation has no consensus sequences that could allow determination of the sites of glycosylation. For example, a glycopeptide having two sites occupied each with a core 1 O-glycan will have the same molecular weight as the glycoform which has only one site occupied with a core 2 O-glycan with the same overall composition as the two core 1 O-glycans. Moreover CID MS/MS analysis on the glycopeptides causes the sugar molecules to fall off thus resulting in ambiguities regarding their sites of attachment.
Future work should take advantage of the more recent technology for glycoproteomic analysis. Analysis of glycopeptides using modern mass spectrometry techniques, such as LC-ESI-MS and MALDI-TOF-MS, has been shown to be able to identify post-translationally modified amino acids on increasingly complex glycopeptides, such as MUC1 and fetuin, accompanied by rigorous isolation strategies (Carr et al., 1993; Yang & Orlando, 1996; Müller et al., 1997; Müller et al., 1999). Continuous optimisation of these and related techniques are providing better and more reliable tools for site mapping of O-glycosylation on more complicated glycoproteins, for instance site-specific analysis on human and murine ZP3 glycoproteins using nanoLC-ESI-QTOF-MS (Chalabi et al., 2005). They have identified a specific O-glycosylation domain on ZP3 consisting of two closely spaced O-glycan sites that are conserved between humans and mice. Multiple stage tandem mass spectrometry using ETD ionisation has been shown to directly elucidate glycan structures and chemically/enzymatically tagged glycosylation sites without LC separation (Wang et al., 2010). Although very promising, this technology remains unproven for characterising highly glycosylated glycopeptides with the complex structures usually seen in mucins. In addition, tagging methods have rarely been shown to work well with mucinous samples (Wada et al., 2010). Therefore much work still needs to be done to ensure that glycoproteomic methodologies are suited to the suggested future work for investigating Src function.

In summary, although the current study has not conclusively described the effects of Src induction and the associated ppGalNAcT relocalisation to the ER on cellular O-glycomes, particularly regarding the core 1 and core 2 ratio that appears to be in contrast with the previous analysis on SYF cells, valuable information has been acquired which is guiding further studies of the role of Src in protein O-glycosylation. By taking into account the current and previous data, it is likely that different types of cells respond distinctively to Src overexpression and to the controversial Src-mediated ppGalNAcT redistribution. In addition, we have demonstrated a possible connection between activated Src and increased in sialylation. However, it remains to be seen whether the sialylation changes are linked to the predicted relocalisation of ppGalNAcTs to the ER. New knowledge that has been assembled here hopefully will lead to a more comprehensive elucidation of the regulation mechanism of O-glycosylation, thus boosting understanding of O-glycan biological roles, particularly in cancer metastasis.
Chapter 6:

Concluding remarks
6 Concluding remarks

This thesis encompasses a significant amount of experimental work and discussion that could open up new avenues for glycobiological research, particularly on O-glycoconjugates. As discussed in the first chapter, with O-glycans, mucins, glycan-protein interactions, mouse models and mass spectrometry as the main themes, it is increasingly clear that comprehensive characterisation of the O-glycome is a prerequisite in order to better understand the functional properties of O-glycans. Whilst the genome encodes the synthesis of functional proteins via nucleic acid sequences, the glycome carries explicit functional properties via highly varied glycan sequences and modifications that can be decoded usually by GBPs. It is evident now that glycome alterations can result in the disruption of cellular interactions and may lead to pathological conditions. Mouse models are an exceptionally powerful tool in demonstrating the function of glycans and are also widely utilised across various life sciences, but throughout the work of this thesis, it was surprising to learn that much is still ambiguous concerning their glycosylation machineries. Three separate but related research projects documented in this thesis, by using mouse models and cultured cell lines, have provided the platform for the optimisation as well as application of O-glycomic methodology with the ultimate goal to better appreciate glycan structure-function correlation. The first project (Chapter 3) addressed the effects of the disruption in O-glycan biosynthetic pathways from a system point of view, whereas the second project (Chapter 4) focused on the interaction between fucosylated O-glycans with *H. pylori* infection. The third project (Chapter 5), on the other hand, has discussed the dynamics of cellular O-glycomes from the perspective of biosynthetic regulation.

In comparison with genomics and proteomics, the progression of glycomics has faced unique challenges in the pursuit of developing analytical and biochemical tools and biological interpretations to investigate glycan structure-function relationships, mainly because of the diversity in terms of glycan chemical structure and information density. The situation is arguably even worse for the analysis of O-glycoconjugates due to the existence of different types and cores and the lack of consensus glycosylation sequences. As a result, structural analysis of O-glycans has been much hampered, as can be illustrated by the relatively low amount of data available for characterised O-glycans, as opposed to studies providing compositional data only, compared to that for N-glycans. Throughout the work of this thesis, a number of discoveries and contributions in terms of O-glycomic methodology, O-glycan sequencing and structure-function relationships have been made. By using a combination of MALDI-TOF/TOF, ESI-QTOF, GC-EI MS/MS
sequencing capabilities, supplemented by MALDI-QIT MS² technology, we have performed detailed O-glycomic analyses to unravel a number of biological challenges raised by our collaborators. As exemplified in Chapters 3 and 4, the manipulation of glycosyltransferases leading to O-glycome aberrations has paved the way for exhaustive O-glycan structural characterisation in correlation with phenotypic studies. This thesis has illustrated possible correlations between core 2 and branched core 1 O-glycans in several murine pathological conditions including hypothyroidism and colitis (Chapter 3). The improvement in methodology and instrumentation was best exemplified by the characterisation of complex and high-mass O-glycans, rigorous sequencing of terminal epitopes as well as the discovery of the RM2-like antigen and O-mannosyl glycans in the GI tract. The apparent extensive compensation seen in the C2GnT knockout tissues provided by the induction of elongated core 1 O-glycans and O-mannosyl glycans carrying terminal epitopes with higher density of fucosylation is consistent with the recently proposed homeostatic regulatory mechanism for the maintenance of cell surface glycan density (Dam & Brewer, 2010).

The mucins of the GI tract are rich in various terminal epitopes that have been linked to symbiotic or pathogenic interactions with diverse species of microbes (Chapter 4). The elimination of the dominating H-antigen capped glycans has led to the sequencing of the lower abundance Lewis antigens in the stomach, which has proven to be helpful for better understanding of *H. pylori* pathogenicity. This work also constitutes one of the first detailed descriptions of type 2 histo-blood group antigens as the dominant epitopes in murine GI tissues. The expression and sorting of glycosyltransferases require proper regulation as their actions can highly influence the messages that cells are presenting on their surfaces. As exemplified in Chapter 5, extracellular signals that are passed via signalling molecules, such as Src family kinases, can alter the localisation of glycosyltransferases, thereby promoting changes in O-glycan biosynthesis. A complete elucidation of the cellular regulation of O-glycosylation is highly anticipated to bring our understanding of functional glycomics to a whole new level. During the course of this thesis work, other groups have also done some excellent work in improving O-glycan sequencing. Perhaps the most notable is the work of the Karlsson group that has rigorously characterised O-glycans from purified mucinous glycoproteins with *m/z* up to 3000 using LC-ESI-MS² (Issa et al., 2010).

Profiling glycosylation from a whole tissue or organ is a reasonable first step towards understanding glycan functions. However, as glycoconjugates work individually, it is crucial to
be able to push forward the acquired data to identify and characterise the molecules that are biologically important. Although very promising and encouraging, the current technology is still limited with respect to this aspect of structural analysis particularly on biological samples that are often complex and heterogeneous. As discussed in Chapter 5, various purification and mass spectrometric techniques have been deployed in order to investigate specific glycoconjugates and their glycosylation even since the 1970’s, however the samples that were addressed were relatively low in complexity. A very good example of the current status of O-glycopeptide analysis worldwide is the recent Human Proteome Organisation (HUPO) Human Disease Glycomics/Proteome Initiative (HGPI) which coordinated the evaluation of methodologies that are widely used for defining the O-glycans by using glycopeptides of IgA1 (Wada et al., 2010). It is important to note that IgA1 bears only up to 5 O-glycosylation sites with glycan compositions up to 5 monosaccharides, yet some laboratories failed to identify all structures. In real mammalian tissues, mucins or mucin-type glycoproteins can contain from dozens to several hundreds of O-glycosylation sites (Rose & Voynow, 2006) with glycan compositions, as presented in Chapter 3 and 4, of up to 18 monosaccharides, thus illustrating how far the technology needs to be improved in order to achieve similar high quality data on these molecules. A very promising strategy for O-glycosylation site analysis that is getting increasing attention is by expressing a selected portion of the glycoprotein or mucin under study in a cell line with limited glycosyltransferases to make complex structures. For instance, a single tandem repeat of MUC1 with 20 amino acids and up to 5 potential O-glycosylation sites has been characterised from mutant Chinese hamster ovary cell lines (Sihlbom et al., 2009). However, this strategy is only possible when the biologically important glycoconjugates have been identified, which is not always the case.

Techniques to improve the isolation and purification of specific cellular organelles, such as the Golgi, or specific glycoproteins of interest also need to be further exploited and developed. Immunoaffinity isolation/purification using lectins and antibodies have been a very useful means to fish-out specific glycoconjugates from complex biological samples (Taylor & Drickamer, 2009). Nevertheless this type of methodology still requires improvements in its applicability as well as resolution of specificity issues, as exemplified in Chapter 4. These techniques could be valuable, for instance, in the isolation of glycoproteins with core 2 O-glycans in the murine thyroid that have a possible connection with the synthesis of thyroxine (Chapter 3). It also would be interesting to be able to pull out glycoproteins that are O-mannosylated in the GI tract as this type of modification is extremely rare. Ongoing research on methodology and technology to
analyse glycoproteins and their associated glycans hopefully will help to push the limit of what is currently achievable. For instance the work of the Hanisch group involving in-gel and on-column de-O-glycosylation methods that can preserve the peptide backbones shows promise for defining glycosylation sites in mucins (Hanisch et al., 2009).

As we enter the second decade of the 21st century, the field of glycobiology is expanding and the demand for knowledge is now even more challenging, therefore continuous optimisation of the analytical methods and technology are necessary. Both the HGPI studies of N- and O-glycans convincingly endorsed mass spectrometry as the technique of choice for glycomic profiling, with MALDI and ESI as the most efficient and reproducible techniques instead of FT-ICR and linear trap quadrupole (LTQ) (Wada et al., 2007; Wada et al., 2010). This partly highlighted that the newer and more sophisticated mass spectrometric instrumentations are not yet capable of rivalling the overall performance of the more established techniques in comparative glycomics. Other than improving the sensitivity and reproducibility, methodology optimisation should also consider ways to reduce cost and processing time. This is in accord with the current economic situation as well as the increasing demand for high-throughput analysis.

Histochemical approaches to analyse molecular and cellular dynamics can provide useful guidance to answering questions about glycan biological functions associated with pathology and were well exemplified by the work of our collaborators. However, development of glyco-biomarkers will require a systems view of the biological systems via the integration of all these approaches including bioinformatics with the structural data from glycomic analysis. Comparative glycan analysis of biological samples with and without disease lesions as presented in this thesis could provide clues to help indentify biomedically important molecules and provide better insights into the molecular properties of carbohydrate related diseases. One of the obstacles of comparative glycomics is the lack of well characterised and comparable glycan standards for quantitative analysis. For this purpose, comprehensive cataloguing of glycan structures in nature, for instance the work of the Dell group (North et al., 2010), as well as optimisation of methods to synthesise these structures, for example the work of the Seeberger group (Weishaupt et al., 2010), are essential. In conclusion, as the glycobiology field moves towards a more holistic understanding of glycosylation function, the expansion and cross-linking of the currently available bioinformatic platforms such as the CFG (www.functionalglycomics.org), GlycoWorkbench (http://www.glycoworkbench.org), GlycoSuite (//glycosuitedb.expasy.org), Carbohydrate-Active Enzymes (www.cazy.org), UniProtKB (www.uniprot.org) and Mouse
Genome Informatics (www.informatics.jax.org) are of importance to integrate the diverse datasets generated using different technologies and approaches.
Appendix and references
### Appendix I

<table>
<thead>
<tr>
<th>GROUP</th>
<th>MONOSACCHARIDE</th>
<th>RING STRUCTURE</th>
<th>CARTOON STRUCTURE</th>
<th>RESIDUE MASS (Da)</th>
<th>PERMETHYLATED RESIDUE MASS (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hexose (Hex)</td>
<td>D-Mannose (Man)</td>
<td><img src="image" alt="D-Mannose" /></td>
<td><img src="image" alt="D-Mannose" /></td>
<td>162</td>
<td>204</td>
</tr>
<tr>
<td>Hexose (Hex)</td>
<td>D-Glucose (Glc)</td>
<td><img src="image" alt="D-Glucose" /></td>
<td><img src="image" alt="D-Glucose" /></td>
<td>162</td>
<td>204</td>
</tr>
<tr>
<td>Hexose (Hex)</td>
<td>D-Galactose (Gal)</td>
<td><img src="image" alt="D-Galactose" /></td>
<td><img src="image" alt="D-Galactose" /></td>
<td>162</td>
<td>204</td>
</tr>
<tr>
<td>N-acetylglycosamine (HexNAc)</td>
<td>N-acetyl-D-glucosamine (GlcNAc)</td>
<td><img src="image" alt="N-acetyl-D-glucosamine" /></td>
<td><img src="image" alt="N-acetyl-D-glucosamine" /></td>
<td>203</td>
<td>245</td>
</tr>
<tr>
<td>N-acetylglycosamine (HexNAc)</td>
<td>N-acetyl-D-galactosamine (GalNAc)</td>
<td><img src="image" alt="N-acetyl-D-galactosamine" /></td>
<td><img src="image" alt="N-acetyl-D-galactosamine" /></td>
<td>203</td>
<td>245</td>
</tr>
<tr>
<td>Deoxyhexose (dHex)</td>
<td>L-Fucose (Fuc)</td>
<td><img src="image" alt="L-Fucose" /></td>
<td><img src="image" alt="L-Fucose" /></td>
<td>146</td>
<td>174</td>
</tr>
<tr>
<td>Sialic acid</td>
<td>N-acetyl-D-neuraminic acid (NeuAc)</td>
<td><img src="image" alt="N-acetyl-D-neuraminic acid" /></td>
<td><img src="image" alt="N-acetyl-D-neuraminic acid" /></td>
<td>291</td>
<td>361</td>
</tr>
<tr>
<td>Sialic acid</td>
<td>N-glycolyl-D-neuraminic acid (NeuGc)</td>
<td><img src="image" alt="N-glycolyl-D-neuraminic acid" /></td>
<td><img src="image" alt="N-glycolyl-D-neuraminic acid" /></td>
<td>307</td>
<td>391</td>
</tr>
</tbody>
</table>
References


Appendix and references


and tissue-specific expression of a murine α-(1,3)-fucosyltransferase locus corresponding to the human ELAM-1 ligand fucosyltransferase. *J. Biol. Chem.* 270, 25047-25056.


Kawar ZS, Johnson TK, Natunen S, Lowe JB & Cummings RD. 2008. PSGL-1 from the murine leukocytic cell line WEHI-3 is enriched for core 2-based O-glycans with sialyl Lewis x antigen. *Glycobiology* 18, 441-446.


selectin glycoprotein ligand-1 but not to 6-sulfo-sialyl Lewis x. *J. Biol. Chem.* 283, 11493-11500.


Appendix and references   205


Appendix and references


Spiro RG. 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology 12, 43R-56R.


Ujita M, Misra AK, McAuliffe J, Hindsgaul O & Fukuda M. 2000. Poly-N-acetyllactosamine extension in N-glycans and core 2- and core 4-branched O-glycans is differentially controlled by i-extension enzyme and different members of the beta 1,4-galactosyltransferase gene family. J. Biol. Chem. 275, 15868-15875.


